



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

(11) 공개번호 10-2023-0124682  
(43) 공개일자 2023년08월25일

- |  |   |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/><i>C12N 15/86</i> (2006.01)</p> <p>(52) CPC특허분류<br/><i>C12N 15/86</i> (2013.01)<br/><i>C12N 2750/0022</i> (2013.01)</p> <p>(21) 출원번호 10-2023-7025029</p> <p>(22) 출원일자(국제) 2021년12월22일<br/>심사청구일자 없음</p> <p>(85) 번역문제출일자 2023년07월21일</p> <p>(86) 국제출원번호 PCT/US2021/064887</p> <p>(87) 국제공개번호 WO 2022/140560<br/>국제공개일자 2022년06월30일</p> <p>(30) 우선권주장<br/>63/130,360 2020년12월23일 미국(US)<br/>63/147,064 2021년02월08일 미국(US)</p> | <p>(71) 출원인<br/>플래그쉽 파이어니어링 이노베이션스 브이, 인크.<br/>미국 02142 매사추세츠주 캠브리지 8쓰 플로어 캠프리지 파크웨이 55</p> <p>(72) 발명자<br/>하자르, 로저 조셉<br/>미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 8쓰플로어<br/>텔라그레이브, 사이먼<br/>미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 8쓰플로어<br/>스완슨, 커트 아담<br/>미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 8쓰플로어</p> <p>(74) 대리인<br/>양영준, 이상남</p> |
|--|---|

전체 청구항 수 : 총 9 항

(54) 발명의 명칭 RNA를 봉입하는 아넬로바이러스 캡시드의 시험관 내어셈블리

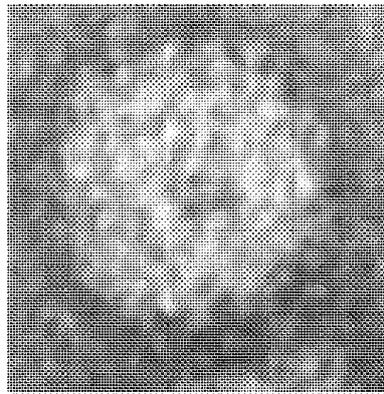
**(57) 요약**

본 발명은 일반적으로 아넬로바이러스를 제조하기 위한 조성물 및 그의 용도에 관한 것이다.

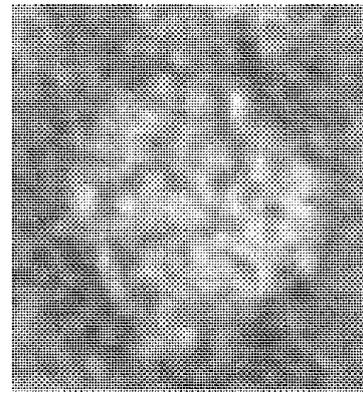
**대표도** - 도1



(A) 고리 2 ORF1



(B) 고리 10 ORF1



(C) CAV VP1

(52) CPC특허분류  
C12N 2750/00043 (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

아넬로박테르로서,

- a) ORF1 분자를 포함하는 단백질성 외부;
- b) RNA를 포함하는 유전 요소를 포함하며,  
유전 요소는 단백질성 외부 내에 봉입되는, 아넬로박테르.

#### 청구항 2

제1항에 있어서, 유전 요소는 적어도 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% RNA로 이루어지는, 아넬로박테르.

#### 청구항 3

제1항 또는 제2항에 있어서, RNA는 하나 이상의 화학적 변형을 포함하는, 아넬로박테르.

#### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 유전 요소는 RNA로 이루어지거나, 본질적으로 이로 이루어지는, 아넬로박테르.

#### 청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 유전 요소는 DNA 영역을 포함하는, 아넬로박테르.

#### 청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, DNA 영역의 적어도 일부는 유전 요소의 RNA의 적어도 일부에 혼성화되는, 아넬로박테르.

#### 청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 유전 요소는 환형인, 아넬로박테르.

#### 청구항 8

아넬로박테르의 제조 방법으로서,

(a) 다음을 포함하는 혼합물을 제공하는 단계:

- (i) RNA를 포함하는 유전 요소, 및
- (ii) ORF1 분자; 그리고

(b) ORF1 분자를 포함하는 단백질성 외부 내에 유전 요소를 봉입하기에 적합한 조건 하에서 혼합물을 항온처리하며, 이에 따라 아넬로박테르를 제조하는 단계를 포함하며;

선택적으로 혼합물은 세포에 포함되지 않는, 방법.

#### 청구항 9

유전 요소를 세포에 운반하는 방법으로서,

제1항 내지 제7항 중 어느 한 항의 아넬로박테르를 세포, 예를 들어, 진핵 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포와 접촉시키는 단계를 포함하는, 방법.

**발명의 설명**

**기술 분야**

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 2020년 12월 23일자 출원된 미국 가출원 제63/130,360호 및 2021년 2월 8일자 출원된 미국 가출원 제 63/147,064호의 이익을 주장한다. 전술된 출원의 내용은 그들 전체가 본원에 참조로 포함된다.

[0003] **서열 목록**

[0004] 본 출원은 ASCII 형식으로 전자 제출된 서열 목록을 포함하며, 그의 전문이 본원에 참조로 포함된다. 2021년 12월 21일자 생성된 상기 ASCII 카피는 명칭이 V2057-7017WO\_SL.txt이며, 286,720 바이트 크기이다.

**배경 기술**

[0005] 환자에게 치료적 이펙터를 운반하기 위한 적합한 벡터를 제조하기 위한 조성물 및 방법을 개발하는 것이 계속 필요하다.

**발명의 내용**

[0006] 본 개시내용은 예를 들어, 유전 물질을 운반하기 위해, 이펙터, 예를 들어, 페이로드를 운반하기 위해, 또는 진핵 세포(예를 들어, 인간 세포 또는 인간 조직)에 대한 치료제 또는 치료적 이펙터를 운반하기 위해 운반 비히클로서 사용될 수 있는 아넬로벡터(예를 들어, 합성 아넬로벡터)를 생성하기 위한 조성물 및 방법을 제공한다. 자연적으로 발생하는 아넬로바이러스(*Anellovirus*)는 DNA 게놈을 가지고 있지만, 본 개시내용은 RNA를 포함하는 유전 요소를 가진 아넬로벡터를 제공한다.

[0007] 아넬로벡터(예를 들어, 본원에 기재된 바와 같은 조성물 또는 방법을 사용하여 생성됨)는 일반적으로 유전 요소를 세포(예를 들어, 포유동물 세포, 예를 들어, 인간 세포) 내로 도입할 수 있는 단백질성 외부(예를 들어, 아넬로바이러스 캡시드 단백질, 예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 아넬로바이러스 ORF1 단백질 또는 폴리펩티드를 포함하는 단백질성 외부)에 캡슐화된 유전 요소(예를 들어, 이펙터, 예를 들어, 외인성 또는 내인성 이펙터, 예를 들어, 치료적 이펙터를 포함하거나 이를 인코딩하는 유전 요소)를 포함한다. 유전 요소는 RNA를 포함할 수 있다. 일부 실시형태에서, 아넬로벡터는 아넬로바이러스 ORF1 핵산(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 알파토르크바이러스(*alphatorquevirus*), 베타토르크바이러스(*Betatorquevirus*) 또는 감마토르크바이러스(*Gammatorquevirus*)의 ORF1 핵산, 예를 들어, 알파토르크바이러스 클레이드 1, 알파토르크바이러스 클레이드 2, 알파토르크바이러스 클레이드 3, 알파토르크바이러스 클레이드 4, 알파토르크바이러스 클레이드 5, 알파토르크바이러스 클레이드 6 또는 알파토르크바이러스 클레이드 7의 ORF1)에 의해 인코딩된 폴리펩티드를 포함하는 단백질성 외부를 포함하는 감염성 비히클 또는 입자이다. 본 개시내용의 아넬로벡터의 유전 요소는 전형적으로 이를 봉입하는 단백질성 외부 또는 이에 부착된 폴리펩티드에 결합하는 단백질 결합 서열을 갖는 환형 및/또는 단일-가닥 RNA 분자(예를 들어, 환형 및 단일 가닥)이며, 이는 단백질성 외부 내의 유전 요소의 봉입 및/또는 단백질성 외부 내에서 다른 핵산에 비한 유전 요소의 농축을 용이하게 할 수 있다. 일부 실시형태에서, 아넬로벡터의 유전 요소는 예를 들어, 본원에 기재된 바와 같은 배칼로바이러스, 핵산 작제물(예를 들어, 박미드 및/또는 공여자 벡터), 곤충 세포 및/또는 동물 세포주를 사용하여 생성된다.

[0008] 일부 예에서, 유전 요소는 예를 들어, 표적 세포에서 발현될 수 있는 이펙터를 포함하거나 이를 인코딩한다(예를 들어, 비-코딩 RNA와 같은 핵산 이펙터를 포함하거나, 폴리펩티드 이펙터, 예를 들어, 단백질을 인코딩함). 일부 실시형태에서, 이펙터는 예를 들어, 본원에 기재된 바와 같은 치료제 또는 치료적 이펙터이다. 일부 실시형태에서, 이펙터는 내인성 이펙터 또는 예를 들어, 야생형 아넬로바이러스 또는 표적 세포에 대하여 외인성인 외인성 이펙터이다. 일부 실시형태에서, 이펙터는 야생형 아넬로바이러스 또는 표적 세포에 대하여 외인성이다. 일부 실시형태에서, 아넬로벡터는 세포와 접촉시키고, 이펙터를 인코딩하는 유전 요소를 세포 내로 도입하여, 이펙터가 세포에 의해 제조되거나 발현되게 함으로써 이펙터를 세포 내로 운반할 수 있다. 특정 예에서, 이펙터는 내인성 이펙터이다(예를 들어, 표적 세포에 대하여 내인성이지만, 예를 들어, 아넬로벡터에 의해 증가된 양으로 제공된다). 다른 예에서, 이펙터는 외인성 이펙터이다. 이펙터는 일부 예에서, 세포의 기능을 조절하거나, 세포 내의 표적 분자의 활성 또는 수준을 조절할 수 있다. 예를 들어, 이펙터는 (예를 들어, PCT/US19/65995의 실시예 3 및 4에 기재된 바와 같이) 세포 내의 표적 단백질의 수준을 감소시킬 수 있다. 또 다른 예에서, 아넬

로백터는 (예를 들어, PCT/US19/65995의 실시예 19 및 28에 기재된 바와 같이) 생체 내에서 이펙터, 예를 들어, 외인성 단백질을 운반하고 발현할 수 있다. 아넬로백터는 예를 들어, 유전 물질을 표적 세포, 조직 또는 대상체에 운반하거나; 이펙터를 표적 세포, 조직 또는 대상체에 운반하거나; 또는 예를 들어, 원하는 세포, 조직 또는 대상체에 대하여 치료제로서 작동할 수 있는 이펙터를 운반함으로써 질병 및 장애를 치료하기 위해 사용될 수 있다.

[0009] 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 예를 들어, 숙주 세포에서 합성 아넬로백터의 유전 요소를 생성하기 위해 사용될 수 있다. 합성 아넬로백터는 야생형 바이러스(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스)에 비해 적어도 하나의 구조적 차이, 예를 들어, 야생형 바이러스에 비해 결실, 삽입, 치환, 변형(예를 들어, 효소적 변형)을 갖는다. 일반적으로, 합성 아넬로백터는 단백질질성 외부 내에 봉입된 외인성 유전 요소를 포함하며, 이는 유전 요소 또는 그 내에 인코딩된 이펙터(예를 들어, 외인성 이펙터 또는 내인성 이펙터)(예를 들어, 폴리펩티드 또는 핵산 이펙터)를 진행(예를 들어, 인간) 세포 내로 운반하기 위해 사용될 수 있다. 실시형태들에서, 아넬로백터는 검출 가능하고/하거나 원하지 않는 면역 또는 염증 반응을 야기하지 않으며, 예를 들어, 염증의 분자 마커(들), 예를 들어, TNF-알파, IL-6, IL-12, IFN, 및 B-세포 반응, 예를 들어, 반응성 또는 중화 항체의 1%, 5%, 10%, 15% 초과 증가를 야기하지 않으며, 예를 들어, 아넬로백터는 표적 세포, 조직 또는 대상체에 대하여 실질적으로 비-면역원성일 수 있다.

[0010] 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 (i) ORF1 분자를 포함하는 단백질성 외부; 및 (ii) RNA를 포함하는 유전 요소를 포함하는 아넬로백터의 유전 요소를 생성하기 위해 사용될 수 있으며, 여기서 유전 요소는 단백질성 외부 내에 봉입되어 있다. 일부 실시형태에서, 유전 요소는 적어도 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% RNA로 이루어진다. 일부 실시형태에서, 유전 요소는 DNA를 포함하지 않는다. 일부 실시형태에서, 유전 요소는 ssDNA를 포함하지 않는다. 대안적으로, 일부 실시형태에서, 유전 요소는 DNA 영역을 포함한다. 일부 실시형태에서, 본원에 기재된 DNA 또는 RNA 분자는 하나 이상의 변형된 뉴클레오티드(예를 들어, 염기 변형, 당 변형 또는 골격 변형)를 포함한다. 일부 실시형태에서, 유전 요소는 단일 가닥이다. 일부 실시형태에서, 유전 요소는 이중 가닥 영역을 포함한다. 일부 실시형태에서 유전 요소는 선형 폴리펩티드이다. 대안적으로, 일부 실시형태에서, 유전 요소는 환형 폴리뉴클레오티드이다. 일부 실시형태에서, 핵산 서열은 예를 들어, 곤충 세포에서의 발현을 위해 코돈-최적화된다. 일부 실시형태에서, 핵산 서열 내의 코돈의 적어도 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%는 예를 들어, 곤충 세포에서의 발현을 위해 코돈-최적화된다. 일부 실시형태에서, 핵산 서열은 예를 들어, 포유동물(예를 들어, 인간) 세포에서의 발현을 위해 코돈-최적화된다. 일부 실시형태에서, 핵산 서열 내의 코돈의 적어도 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%는 예를 들어, 포유동물(예를 들어, 인간) 세포에서의 발현을 위해 코돈-최적화된다. 일부 실시형태에서, 유전 요소는 약 10 내지 20, 20 내지 30, 30 내지 40, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 125, 125 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 900 내지 1000, 1000 내지 1500, 1500 내지 2000, 2000 내지 2500, 2500 내지 3000, 3000 내지 3500, 3500 내지 4000 또는 4000 내지 4500개의 뉴클레오티드 길이이다. 일부 실시형태에서, 유전 요소는 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000 또는 4500개의 뉴클레오티드 길이이다.

[0011] 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 캡시드에 결합하는 단백질 결합 서열 및 치료적 이펙터를 인코딩하는 (아넬로바이러스에 대하여) 이중 서열을 포함하는 유전 요소를 캡슐화하는 캡시드(예를 들어, 아넬로 바이러스 ORF, 예를 들어, ORF1, 폴리펩티드를 포함하는 캡시드)를 포함하는 (예를 들어, 인간 세포에 대한) 감염성 아넬로백터, 비히클, 또는 입자의 유전 요소를 생성하기 위해 사용될 수 있다. 실시형태에서, 아넬로백터는 유전 요소를 포유동물, 예를 들어, 인간 세포 내로 운반할 수 있다.

[0012] 일 양태에서, 본 발명은 시험관 내 어셈블리에 의해 아넬로백터를 제조하는 방법을 특징으로 한다. 일부 실시형태에서, 아넬로백터를 제조하는 방법은 (a) (i) RNA를 포함하는 유전 요소, 및 (ii) ORF1 분자를 포함하는 혼합물을 제공하는 단계; 및 (b) ORF1 분자를 포함하는 단백질성 내에 유전 요소를 봉입하기에 적합한 조건 하에서 혼합물을 향온처리하여, 아넬로백터를 제조하는 단계를 포함하며; 선택적으로 여기서 혼합물은 세포 내에 포함되지 않는다. 일부 실시형태에서, 방법은 (a)를 제공하기 전에, 예를 들어, 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)에서 ORF1 분자를 발현시키는 단계를 추가로 포함한다. 일부 실시형태에서, 발현은 ORF1 분자를 생성하기에 적합한 조건 하에서 ORF1 분자를 인코딩하는 핵산 분자(예를 들어, 배큘로바이러스 발현 벡터)를 포함하는 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)를 향온처리하는 단계를 포함한다. 일부 실시형

태에서, 방법은 (a)를 제공하기 전에, 숙주 세포에 의해 발현된 ORF1 분자를 정제하는 단계를 추가로 포함한다.

- [0013] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로벡터는 본원에 기재된 이펙터와 같은 제제를 표적 세포, 예를 들어, 치료적 또는 예방적으로 치료될 대상체 내의 표적 세포에 도입하기 위한 효과적인 운반 비히클로서 사용될 수 있다.
- [0014] 일 양태에서, 본 발명은 본원에 기재된 바와 같은 아넬로벡터(예를 들어, 합성 아넬로벡터)을 포함하는 약제학적 조성물을 특징으로 한다. 일부 실시형태에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체 또는 부형제를 추가로 포함한다. 실시형태들에서, 약제학적 조성물은 표적 대상체 킬로그램당 약  $10^5$  내지  $10^{14}$ 개(예를 들어, 약  $10^6$  내지  $10^{13}$ ,  $10^7$  내지  $10^{12}$ ,  $10^8$  내지  $10^{11}$  또는  $10^9$  내지  $10^{10}$ 개)의 계놈 당량의 아넬로벡터를 포함하는 단위 용량을 포함한다. 일부 실시형태에서, 제제를 포함하는 약제학적 조성물은 허용 가능한 기간 및 온도에 걸쳐 안정할 것이고/것이거나 요망되는 투여 경로 및/또는 투여 경로가 필요로 할 임의의 디바이스, 예를 들어, 니들(needle) 또는 주사기와 양립 가능할 것이다. 일부 실시형태에서, 약제학적 조성물은 단위 용량 또는 다회 용량으로서의 투여를 위해 제형화된다. 일부 실시형태에서, 약제학적 조성물은 투여 장소에서, 예를 들어, 의료 전문가(healthcare professional)에 의해 제형화된다. 일부 실시형태에서, 약제학적 조성물은 (예를 들어, 부피당 계놈의 수에 의해 정의된 바와 같은) 요망되는 농도의 아넬로벡터 계놈 또는 계놈 당량을 포함한다.
- [0015] 일 양태에서, 본 발명은 대상체에서의 질병 또는 장애의 치료 방법을 특징으로 하며, 당해 방법은 예를 들어, 본원에 기재된 바와 같은, 아넬로벡터, 예를 들어, 합성 아넬로벡터를 대상체에게 투여하는 단계를 포함한다.
- [0016] 일 양태에서, 본 발명은 이펙터 또는 페이로드(예를 들어, 내인성 또는 외인성 이펙터)를 세포, 조직 또는 대상체에 운반하는 방법을 특징으로 하며, 당해 방법은 예를 들어, 본원에 기재된 바와 같은, 아넬로벡터, 예를 들어, 합성 아넬로벡터를 대상체에게 투여하는 단계를 포함하며, 아넬로벡터는 이펙터를 인코딩하는 핵산 서열을 포함한다. 일부 실시형태에서, 페이로드는 핵산이다. 일부 실시형태에서, 페이로드는 폴리펩티드이다.
- [0017] 일 양태에서, 본 발명은 예를 들어, 본원에 기재된 바와 같은 아넬로벡터, 예를 들어, 합성 아넬로벡터를 예를 들어, 생체 내에서 또는 생체 외에서, 세포, 예를 들어, 진핵 세포, 예를 들어, 포유동물 세포와 접촉시키는 단계를 포함하는 세포로의 아넬로벡터의 운반 방법을 특징으로 한다.
- [0018] 전술한 아넬로벡터, 조성물 또는 방법 중 임의의 추가의 특징은 다음의 열거된 실시형태 중 하나 이상을 포함한다.
- [0019] 당업자는 단지 일상적인 실험을 사용하여, 본원에 기재된 본 발명의 특정 실시형태의 다수의 등가물을 인식하거나 확인할 수 있을 것이다. 그러한 등가물은 다음의 열거된 실시형태에 의해 포괄되는 것으로 의도된다.
- [0020] **열거된 실시형태**
- [0021] 1. 아넬로벡터로서,
- [0022] a) ORF1 분자를 포함하는 단백질성 외부;
- [0023] b) RNA를 포함하는 유전 요소를 포함하며,
- [0024] 유전 요소는 단백질성 외부 내에 봉입되는, 아넬로벡터.
- [0025] 2. 실시형태 1에 있어서, 유전 요소는 적어도 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% RNA로 이루어지는, 아넬로벡터.
- [0026] 3. 실시형태 1 또는 2에 있어서, RNA는 하나 이상의 화학적 변형을 포함하는, 아넬로벡터.
- [0027] 4. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 RNA로 이루어지거나 본질적으로 RNA로 이루어지는, 아넬로벡터.
- [0028] 5. 선행하는 실시형태 중 어느 하나에 있어서, 아넬로벡터는 DNA를 포함하지 않는, 아넬로벡터.
- [0029] 6. 선행하는 실시형태 중 어느 하나에 있어서, 아넬로벡터는 ssDNA를 포함하지 않는, 아넬로벡터.
- [0030] 7. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 DNA 영역을 포함하는, 아넬로벡터.
- [0031] 8. 선행하는 실시형태 중 어느 하나에 있어서, DNA 영역의 모든 뉴클레오티드는 화학적으로 변형되는, 아넬로벡터.

- [0032] 9. 실시형태 7 또는 8에 있어서, DNA 영역은 유전 요소의 RNA에 공유적으로 연결되는, 아넬로백터.
- [0033] 10. 선행하는 실시형태 중 어느 하나에 있어서, DNA 영역의 적어도 일부는 유전 요소의 RNA의 적어도 일부에 혼성화되는, 아넬로백터.
- [0034] 11. 선행하는 실시형태 중 어느 하나에 있어서, DNA 영역은 10 내지 20, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 또는 900 내지 1000 개의 뉴클레오티드 길이인, 아넬로백터.
- [0035] 12. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 단일 가닥인, 아넬로백터.
- [0036] 13. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 이중 가닥 영역(예를 들어, RNA와 쌍을 형성하는 RNA의 영역 또는 RNA와 쌍을 형성하는 DNA의 영역)을 포함하는, 아넬로백터.
- [0037] 14. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 선형인, 아넬로백터.
- [0038] 15. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 환형인, 아넬로백터.
- [0039] 16. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 제1 영역, 및 제1 영역과 혼성화될 수 있는 제2 영역을 포함하는, 아넬로백터.
- [0040] 17. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 5' 말단 또는 3' 말단을 포함하지 않는, 아넬로백터.
- [0041] 18. 실시형태 15 내지 17 중 어느 하나에 있어서, 유전 요소는 유리 인산염 및 유리 당 중 하나 또는 둘 모두를 포함하지 않는, 아넬로백터.
- [0042] 19. 실시형태 15 내지 18 중 어느 하나에 있어서, 유전 요소 내의 모든 인산염은 인산염에 포함된 제1 산소 원자에 의해 제1 당, 및 인산염에 포함된 제2 산소 원자에 의해 제2 당에 공유적으로 연결되는, 아넬로백터.
- [0043] 20. 실시형태 15 내지 19 중 어느 하나에 있어서, 유전 요소 내의 모든 당은 당에 포함된 제1 탄소 원자에 의해 제1 인산염, 및 당에 포함된 제2 탄소 원자에 의해 제2 인산염에 공유적으로 연결되는, 아넬로백터.
- [0044] 21. 실시형태 15 내지 20 중 어느 하나에 있어서, 유전 요소는 선형 RNA를 환화시킴으로써 생성되는, 아넬로백터.
- [0045] 22. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 약 10 내지 20, 20 내지 30, 30 내지 40, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 125, 125 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 900 내지 1000, 1000 내지 1500, 1500 내지 2000, 2000 내지 2500, 2500 내지 3000, 3000 내지 3500, 3500 내지 4000 또는 4000 내지 4500개의 뉴클레오티드 길이인, 아넬로백터.
- [0046] 23. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000 또는 4500개의 뉴클레오티드 길이인, 아넬로백터.
- [0047] 24. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 ORF1 분자에 결합하는, 아넬로백터.
- [0048] 25. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 ORF1 분자에 포함된 젤리 롤(jelly roll) 도메인에 결합하는, 아넬로백터.
- [0049] 26. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 ORF1 분자에 포함된 아르기닌-풍부 도메인에 결합하는, 아넬로백터.
- [0050] 27. 선행하는 실시형태 중 어느 하나에 있어서, 아넬로백터는 복수의 유전 요소, 예를 들어, 적어도 2, 3, 4, 5, 10, 20, 30, 40, 50 또는 60개의 유전 요소를 포함하는, 아넬로백터.
- [0051] 28. 실시형태 27에 있어서, 복수의 유전 요소 각각은 동일한 서열을 포함하는, 아넬로백터.
- [0052] 29. 실시형태 27에 있어서, 복수의 유전 요소는 상이한 서열을 포함하는, 아넬로백터.
- [0053] 30. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 외인성 이펙터를 인코딩하는, 아넬로백터.

- [0054] 31. 실시형태 30에 있어서, 외인성 이펙터를 인코딩하는 서열은 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500 또는 3000 개의 뉴클레오티드 길이인, 아넬로벡터.
- [0055] 32. 실시형태 30 또는 31에 있어서, 외인성 이펙터는 치료적 이펙터(예를 들어, 폴리펩티드 또는 핵산 분자)를 포함하는, 아넬로벡터.
- [0056] 33. 실시형태 30 내지 32 중 어느 하나에 있어서, 외인성 이펙터는 인간 단백질, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드를 포함하는, 아넬로벡터.
- [0057] 34. 실시형태 30 내지 33 중 어느 하나에 있어서, 외인성 이펙터는 핵산 분자를 포함하는, 아넬로벡터.
- [0058] 35. 실시형태 30 내지 34 중 어느 하나에 있어서, 외인성 이펙터는 비코딩 핵산 분자, 예를 들어, 기능적 RNA, 예를 들어, mRNA, miRNA 또는 siRNA를 포함하는, 아넬로벡터.
- [0059] 36. 실시형태 30 내지 35 중 어느 하나에 있어서, 유전 요소는 외인성 이펙터(예를 들어, 펩티드 또는 폴리펩티드, 예를 들어, 치료적 펩티드 또는 폴리펩티드)를 인코딩하는 mRNA 분자인, 아넬로벡터.
- [0060] 37. 실시형태 36에 있어서, 비코딩 핵산 분자는 인간 비코딩 핵산 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함하는 핵산 분자인, 아넬로벡터.
- [0061] 38. 실시형태 30 내지 37 중 어느 하나에 있어서, 외인성 이펙터는 사이토졸 폴리펩티드 또는 사이토졸 펩티드(예를 들어, DPP-4 억제제, GLP-1 신호전달의 활성화제, 또는 호중구 엘라스타제의 억제제, 또는 그의 기능적 단편)을 포함하는, 아넬로벡터.
- [0062] 39. 실시형태 38에 있어서, 외인성 이펙터는 조절 세포내 폴리펩티드를 포함하는, 아넬로벡터.
- [0063] 40. 실시형태 30 내지 39 중 어느 하나에 있어서, 외인성 이펙터는 분비 폴리펩티드 또는 펩티드(예를 들어, 사이토카인, 항체 분자, 호르몬, 성장 인자 또는 응고-연관된 인자, 또는 그의 기능적 단편)를 포함하는, 아넬로벡터.
- [0064] 41. 실시형태 30 내지 40 중 어느 하나에 있어서, 외인성 이펙터는 단백질 대체 치료제를 포함하는, 아넬로벡터.
- [0065] 42. 실시형태 30 내지 41 중 어느 하나에 있어서, 외인성 이펙터는 효소를 포함하는, 아넬로벡터.
- [0066] 43. 실시형태 30 내지 42 중 어느 하나에 있어서, 외인성 이펙터는 에리트로포이에틴(EPO) 또는 인간 성장 호르몬(hGH), 또는 그의 기능적 단편을 포함하는, 아넬로벡터.
- [0067] 44. 실시형태 30 내지 43 중 어느 하나에 있어서, 외인성 이펙터는 유전자 편집 시스템의 구성성분(예를 들어, CRISPR 시스템의 구성성분, 예를 들어, Cas9, Cpf1 또는 그의 기능적 단편)을 포함하는, 아넬로벡터.
- [0068] 45. 선행하는 실시형태 중 어느 하나에 있어서, RNA는 예를 들어, 본원에 기재된 바와 같은 화학적으로 변형된 RNA를 포함하는, 아넬로벡터.
- [0069] 46. 선행하는 실시형태 중 어느 하나에 있어서, RNA는 캡을 포함하는, 아넬로벡터.
- [0070] 47. 선행하는 실시형태 중 어느 하나에 있어서, RNA는 폴리-A 꼬리, 예를 들어, 적어도 약 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 또는 100개의 아데노신 길이를 포함하는, 아넬로벡터.
- [0071] 48. 선행하는 실시형태 중 어느 하나에 있어서, RNA에는 폴리-A 꼬리가 결합되어 있으며, 예를 들어, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 또는 100개 이하의 순차적인 아데노신을 포함하는, 아넬로벡터.
- [0072] 49. 선행하는 실시형태 중 어느 하나에 있어서, 단백질성 외부는 ORF1 분자의 약 60개(예를 들어, 약 40, 50, 60, 70 또는 80개)의 카피를 포함하는, 아넬로벡터.
- [0073] 50. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자의 켈리 롤 도메인은 단백질성 외부의 내부를 향하는, 아넬로벡터.
- [0074] 51. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자는 표 N 내지 S 및 37A 내지 37C 중 임의의 것에 열거된 바와 같은 아미노산 서열, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%,

96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 아넬로백터.

- [0075] 52. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자는 예를 들어, 표 N 내지 S 및 37A 내지 37C 중 임의의 것에 열거된 바와 같은 아르기닌-풍부 영역의 아미노산 서열, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 아르기닌-풍부 영역을 포함하는, 아넬로백터.
- [0076] 53. 실시형태 52에 있어서, 아르기닌-풍부 영역은 적어도 약 70%(예를 들어, 적어도 약 70%, 75%, 80%, 85%, 90%, 95% 또는 100%)의 염기성 잔기(예를 들어, 아르기닌 또는 라이신)를 포함하는, 아넬로백터.
- [0077] 54. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자는 예를 들어, 표 N 내지 S 및 37A 내지 37C 중 임의의 것에 열거된 젤리 롤 도메인의 아미노산 서열, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 젤리 롤 도메인을 포함하는, 아넬로백터.
- [0078] 55. 실시형태 54에 있어서, 젤리 롤 도메인은 다음의 특징 중 하나 이상(예를 들어, 1, 2, 3 또는 4개)을 포함하는, 아넬로백터:
  - [0079] (i) 젤리-롤 도메인의 적어도 30%(예를 들어, 적어도 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90% 이상)의 아미노산은 하나 이상의  $\beta$ -시트의 일부이고/이거나;
  - [0080] (ii) 젤리-롤 도메인의 2차 구조는 적어도 4개(예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10, 11 또는 12개)의  $\beta$ -가닥을 포함하고/하거나;
  - [0081] (iii) 젤리-롤 도메인의 3차 구조는 적어도 2개(예를 들어, 적어도 2, 3 또는 4개)의  $\beta$ -시트를 포함하고/하거나;
  - [0082] (iv) 젤리-롤 도메인은 적어도 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1 또는 10:1의  $\beta$ -시트 대  $\alpha$ -나선의 비율 포함함.
- [0083] 56. 실시형태 54에 있어서, 젤리 롤 도메인은 예를 들어, 서로에 대해 역평행 배향으로 배열된 2개의  $\beta$ -시트를 포함하는, 아넬로백터.
- [0084] 57. 실시형태 54에 있어서, 젤리 롤 도메인은 8개의  $\beta$ -가닥을 포함하는, 아넬로백터.
- [0085] 58. 실시형태 52 내지 57 중 어느 하나에 있어서, 젤리 롤 도메인은 예를 들어, 도 3에 나타난 바와 같은 D  $\beta$ -가닥의 서열에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 영역을 포함하는, 아넬로백터.
- [0086] 59. 실시형태 52에 있어서, D  $\beta$ -가닥은 1, 2 또는 3개 이상의 염기성 잔기(예를 들어, 아르기닌 또는 라이신)를 포함하는, 아넬로백터.
- [0087] 60. 실시형태 52 내지 59 중 어느 하나에 있어서, 젤리 롤 도메인은 예를 들어, 도 3에 나타난 바와 같은 G  $\beta$ -가닥의 서열에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 영역을 포함하는, 아넬로백터.
- [0088] 61. 실시형태 52에 있어서, G  $\beta$ -가닥은 적어도 약 1, 2 또는 3개 이상의 염기성 잔기(예를 들어, 아르기닌 또는 라이신)를 포함하는, 아넬로백터.
- [0089] 62. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자는, 예를 들어, 표 N 내지 S 및 37A 내지 37C 중 임의의 것에 열거된 바와 같은 N22 도메인의 아미노산 서열, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 N22 도메인을 포함하는, 아넬로백터.
- [0090] 63. 선행하는 실시형태 중 어느 하나에 있어서, N22 도메인은 아미노산 서열  $YNPX^2DXGX^2N$ 을 포함하며, 여기서  $X^n$ 은 각각 독립적으로 임의의  $n$ 개의 아미노산의 인접 서열인, 아넬로백터.
- [0091] 64. 선행하는 실시형태 중 어느 하나에 있어서, N22 도메인은 아미노산 서열  $YNPX^2DXGX^2N$ 에 축적하는 제1 베타 가닥 및 제2 베타 가닥을 포함하며, 예를 들어, 제1 베타 가닥은 아미노산 서열  $YNPX^2DXGX^2N$ 의 티로신(Y) 잔기를

포함하고/하거나, 제2 베타 가닥은 아미노산 서열  $YNPX^2DXGX^2N$ 의 제2 아스파라긴(N) 잔기(N에서 C로)를 포함하는, 아넬로백터.

- [0092] 65. 선행하는 실시형태 중 어느 하나에 있어서, ORF1 분자는 C-말단 도메인, 예를 들어, 표 N 내지 S 및 37A 내지 37C 중 임의의 것에 열거된 바와 같은 C-말단 도메인의 아미노산 서열, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 C-말단 도메인을 포함하는, 아넬로백터.
- [0093] 66. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소에는 아넬로바이러스 ORF1 단백질(예를 들어, 본원에 기재된 바와 같음)을 인코딩하는 서열이 결합되어 있는, 아넬로백터.
- [0094] 67. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소에는 아넬로바이러스 ORF2 단백질(예를 들어, 본원에 기재된 바와 같음)을 인코딩하는 서열이 결합되어 있는, 아넬로백터.
- [0095] 68. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소에는 아넬로바이러스 ORF3 단백질(예를 들어, 본원에 기재된 바와 같음)을 인코딩하는 서열이 결합되어 있는, 아넬로백터.
- [0096] 69. 선행하는 실시형태 중 어느 하나에 있어서, 아넬로백터는 유전 요소를 세포(예를 들어, 진핵 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포)로 운반하도록 구성되는, 아넬로백터.
- [0097] 70. 실시형태 69에 있어서, 적어도 1000개의 아넬로백터의 집단은 적어도 약 100개의 카피(예를 들어, 적어도 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 5000, 10,000, 50,000, 100,000, 500,000 또는 1,000,000개의 카피)의 유전 요소를 하나 이상의 세포 내로 운반할 수 있는, 아넬로백터.
- [0098] 71. 실시형태 69 또는 70에 있어서, 아넬로백터의 집단(예를 들어, 세포당 적어도 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1000 개놈 당량의 유전 요소)은 유전 요소를 적어도 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% 이상의 세포의 집단 내로 운반할 수 있는, 아넬로백터.
- [0099] 72. 실시형태 69 내지 71 중 어느 하나에 있어서, 아넬로백터의 집단(예를 들어, 세포당 적어도 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1000 개놈 당량의 유전 요소)은 세포당 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 8,000,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$ 개 또는 이상의 카피의 유전 요소를 세포의 집단에 운반할 수 있는, 아넬로백터.
- [0100] 73. 실시형태 69 내지 72 중 어느 하나에 있어서, 아넬로백터의 집단(예를 들어, 세포당 적어도 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1000 개놈 당량의 유전 요소)은 세포당 1 내지 3, 1 내지 4, 1 내지 5, 1 내지 6, 1 내지 7, 1 내지 8, 1 내지 9, 1 내지 10, 5 내지 10, 10 내지 20, 20 내지 50, 50 내지 100, 100 내지 1000, 1000 내지  $10^4$ ,  $1 \times 10^4$  내지  $1 \times 10^5$ ,  $1 \times 10^4$  내지  $1 \times 10^6$ ,  $1 \times 10^4$  내지  $1 \times 10^7$ ,  $1 \times 10^5$  내지  $1 \times 10^6$ ,  $1 \times 10^5$  내지  $1 \times 10^7$  또는  $1 \times 10^6$  내지  $1 \times 10^7$ 개의 카피의 유전 요소를 세포의 집단에 운반할 수 있는, 아넬로백터.
- [0101] 74. 선행하는 실시형태 중 어느 하나에 있어서, 아넬로백터는 요망되는 세포 유형, 조직 또는 기관(예를 들어, 골수, 혈액, 심장, GI, 피부, 망막 내의 광수용기, 상피 내벽 또는 횡장)에 이펙터를 선택적으로 운반하는 또는 그 내에서 더 높은 수준으로 존재하는(예를 들어, 그 내에 우선적으로 축적되는), 아넬로백터.
- [0102] 75. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 RNase에 의한(예를 들어, 단백질성 외부에 의한) 소화로부터 보호되거나 이에 저항성인, 아넬로백터.
- [0103] 76. 선행하는 실시형태 중 어느 하나에 있어서, 단백질성 외부 내에 봉입된 유전 요소는 엔도뉴클레아제 소화, 예를 들어, RNase 소화에 대해 저항성인, 아넬로백터.
- [0104] 77. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 프로모터 요소를 포함하는, 아넬로백터.
- [0105] 78. 선행하는 실시형태 중 어느 하나에 있어서, 유전 요소는 단백질 결합 서열을 포함하는, 아넬로백터.
- [0106] 79. 실시형태 78에 있어서, 단백질 결합 서열은 ORF1 분자에 결합할 수 있는, 아넬로백터.

- [0107] 80. 선행하는 실시형태 중 어느 하나의 복수의 아넬로백터를 포함하는 조성물.
- [0108] 81. 실시형태 80에 있어서, 조성물에 ORF1 분자를 포함하는 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%의 단백질성 외부는 유전 요소의 적어도 하나의 카피를 포함하는, 조성물.
- [0109] 82. 실시형태 80에 있어서, 조성물에 ORF1 분자를 포함하는 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%의 단백질성 외부는 아넬로백터 유전 요소의 적어도 하나의 카피를 포함하는, 조성물.
- [0110] 83. 실시형태 80 내지 82 중 어느 하나에 있어서, 조성물은 적어도  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^4$ ,  $10^5$ ,  $10^6$  또는  $10^7$ 개의 동일한 아넬로백터를 포함하는, 조성물.
- [0111] 84. 실시형태 80 내지 83 중 어느 하나에 있어서, 복수는 적어도  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$  또는  $10^{15}$ 개의 아넬로백터(예를 들어, 아넬로백터의 카피)를 포함하거나; 조성물은 mL당 적어도  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$  또는  $10^{15}$ 개의 아넬로백터 게놈을 포함하는, 조성물.
- [0112] 85. 실시형태 80 내지 84 중 어느 하나에 있어서, 다음의 특징 중 하나 이상(예를 들어, 1, 2, 3, 4, 5 또는 6개)을 갖는, 조성물:
- [0113] a) 조성물은 약제학적 또는 우수 제조관리 기준(GMP: good manufacturing practice) 표준을 충족함;
- [0114] b) 조성물은 우수 제조관리 기준(GMP)에 따라 제조됨;
- [0115] c) 조성물은 사전결정된 기준 값 미만의 병원체 수준을 가짐, 예를 들어, 실질적으로 병원체가 없음;
- [0116] d) 조성물은 사전결정된 기준 값 미만의 오염물질 수준을 가짐, 예를 들어, 실질적으로 오염물질(예를 들어, 변성제, 예를 들어, 요소)이 없음;
- [0117] e) 조성물은 사전결정된 비-감염성 입자의 수준 또는 사전결정된 입자:감염성 유닛의 비(예를 들어, 300:1 미만, 200:1 미만, 100:1 미만 또는 50:1 미만)를 갖거나,
- [0118] f) 조성물은 예를 들어, 본원에 기재된 바와 같이, 낮은 면역원성을 갖거나, 실질적으로 비-면역원성임;
- [0119] 선택적으로 여기서 조성물은 0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 0.9 M, 1 M, 1.1 M, 1.2 M, 1.3 M, 1.5 M, 1.5 M, 1.6 M, 1.7 M, 1.8 M, 1.9 M 또는 2 M 미만의 농도의 요소를 포함함.
- [0120] 86. 실시형태 80 내지 85 중 어느 하나에 있어서, 약제학적 조성물은 사전결정된 기준 값 미만의 오염물질 수준을 갖는, 예를 들어, 오염물질이 실질적으로 없는, 조성물.
- [0121] 87. 실시형태 80 내지 86 중 어느 하나에 있어서, 조성물은 0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 0.9 M, 1 M, 1.1 M, 1.2 M, 1.3 M, 1.5 M, 1.5 M, 1.6 M, 1.7 M, 1.8 M, 1.9 M 또는 2 M 미만의 농도의 요소를 포함하는, 조성물.
- [0122] 88. 실시형태 86 또는 87에 있어서, 오염물질은 마이코플라스마, 내독소, 숙주 세포 핵산(예를 들어, 숙주 세포 DNA 및/또는 숙주 세포 RNA), 동물-유래 프로세스 불순물(예를 들어, 혈청 알부민 또는 트립신), 복제-적격 제제(RCA), 예를 들어, 복제-적격 바이러스 또는 원치 않는 아넬로백터(예를 들어, 요망되는 아넬로백터, 예를 들어, 본원에 기재된 바와 같은 합성 아넬로백터 이외의 아넬로백터), 유리 바이러스 캡시드 단백질, 외래성 물질(adventitious agent) 및/또는 응집물 중 하나 이상을 포함하는, 조성물.
- [0123] 89. 실시형태 80 내지 88 중 어느 하나에 있어서, 조성물은 중량 기준 10% 미만(예를 들어, 약 10%, 5%, 4%, 3%, 2%, 1%, 0.5% 또는 0.1% 미만)의 오염물질을 포함하는, 조성물.
- [0124] 90. 실시형태 80 내지 89 중 어느 하나에 있어서, 적어도 90%의 단백질성 외부는 동일한 유전 요소(예를 들어, 아넬로백터의 유전 요소)를 포함하는, 조성물.
- [0125] 91. 실시형태 80 내지 90 중 어느 하나에 있어서, 적어도 90%의 단백질성 외부는 동일한 ORF1 분자를 포함하는, 조성물.
- [0126] 92. 아넬로백터 또는 전술한 실시형태 중 임의의 것의 조성물, 및 약제학적으로 허용 가능한 담체 또는 부형제를 포함하는 약제학적 조성물.

- [0127] 93. 아넬로백터의 제조 방법으로서, 방법이
- [0128] (a) 다음을 포함하는 혼합물을 제공하는 단계:
- [0129] (i) RNA를 포함하는 유전 요소, 및
- [0130] (ii) ORF1 분자; 그리고
- [0131] (b) ORF1 분자를 포함하는 단백질성 외부 내에 유전 요소를 봉입하기에 적합한 조건 하에서 혼합물을 향온처리하며, 이에 따라 아넬로백터를 제조하는 단계를 포함하며;
- [0132] 선택적으로 혼합물은 세포에 포함되지 않는, 방법.
- [0133] 94. 실시형태 93에 있어서, (a)를 제공하기 전에, 예를 들어, 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)에서 ORF1 분자를 발현하는 단계를 추가로 포함하는, 방법.
- [0134] 95. 실시형태 94에 있어서, 발현은 ORF1 분자를 생성하기에 적합한 조건 하에서 ORF1 분자를 인코딩하는 핵산 분자(예를 들어, 배콜로바이러스 발현 벡터)를 포함하는 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)를 향온처리하는 단계를 포함하는, 방법.
- [0135] 96. 실시형태 94 또는 95에 있어서, (a)를 제공하기 전에 숙주 세포에 의해 발현된 ORF1 분자를 정제하는 단계를 추가로 포함하는, 방법.
- [0136] 97. 아넬로백터를 정제하는 방법으로서, 방법이
- [0137] (a) 다음을 포함하는 아넬로백터(예를 들어, 본원에 기재된 바와 같음)를 제공하는 단계:
- [0138] (i) 유전 요소, 예를 들어, RNA를 포함하는 유전 요소, 및
- [0139] (ii) 단백질성 외부로서, ORF1 분자를 포함하며, 유전 요소를 봉입하는, 단백질성 외부; 그리고
- [0140] (b) 아넬로백터를 정제하는 단계를 포함하는, 방법.
- [0141] 98. 실시형태 96 또는 97에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 친화도 정제, 예를 들어, 헤파린 친화도 정제를 포함하는, 방법.
- [0142] 99. 실시형태 96 내지 98 중 어느 하나에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 크기 배제 크로마토그래피(예를 들어, 트리스 완충액 이동상을 사용함)를 포함하는, 방법.
- [0143] 100. 실시형태 96 내지 99 중 어느 하나에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 친화도 정제(예를 들어, 헤파린 친화도 정제)에 이어서 크기 배제 크로마토그래피를 포함하는, 방법.
- [0144] 101. 실시형태 96 내지 100 중 어느 하나에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 음이온 교환 크로마토그래피(예를 들어, Mustang Q 막 크로마토그래피)를 포함하는, 방법.
- [0145] 102. 실시형태 96 내지 101 중 어느 하나에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 혼합 모드 크로마토그래피(예를 들어, 혼합 모드 수지, 예를 들어, Cato700 수지를 사용함)를 포함하는, 방법.
- [0146] 103. 실시형태 96 내지 102 중 어느 하나에 있어서, 정제(예를 들어, ORF1 분자의 정제 또는 아넬로백터의 정제)는 ORF1 분자의, 적어도 약 20, 30, 40, 50 또는 60개의 카피, 또는 20 내지 30, 30 내지 40, 40 내지 50 또는 50 내지 60개의 카피를 포함하는 하나 이상의 바이러스-유사 입자(VLP)를 포함하는 조성물을 생성하는, 방법.
- [0147] 104. 실시형태 103에 있어서, 적어도 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%의 바이러스-유사 입자는 60 mer이거나 적어도 30, 31, 32, 33, 34 또는 35 nm의 직경의 입자인 단백질성 외부를 포함하는, 방법.
- [0148] 105. 실시형태 103에 있어서, 조성물은 적어도  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$  개의 입자/mL를 포함하거나,  $10^5$  내지  $10^6$ ,  $10^6$  내지  $10^7$ ,  $10^7$  내지  $10^9$ ,  $10^8$  내지  $10^9$ ,  $10^9$  내지  $10^{10}$ , 또는  $10^{10}$  내지  $10^{11}$  개의 입자/mL(예를 들어, 전자 현미경에 의해 측정되는 바와 같음)을 포함하는, 방법.
- [0149] 106. 실시형태 93 내지 105 중 어느 하나에 있어서, (a)를 제공하기 전에, ORF1 분자를 포함하는 단백질성 외부(예를 들어, 바이러스-유사 입자(VLP))의 분해(disassembly)에 적합한 조건 하에서 ORF1 분자를 향온처리하는

단계를 추가로 포함하는, 방법.

- [0150] 107. 실시형태 106에 있어서, ORF1 분자를 포함하는 단백질성 외부의 분해에 적합한 조건은 변성제의 존재 하에 서의 항온처리를 포함하는, 방법.
- [0151] 108. 실시형태 106 또는 107항에 있어서, 변성제는 카오트로픽제(chaotropic agent)(예를 들어, 요소), 또는 세 제(예를 들어, SDS(예를 들어, 0.1% SDS), Tween 또는 Triton)를 포함하는, 방법.
- [0152] 109. 실시형태 106 내지 108 중 어느 하나에 있어서, ORF1 분자를 포함하는 단백질성 외부의 분해에 적합한 조 건은 사전결정된 전도도, 고염 용액(예를 들어, NaCl을 포함하는 용액, 예를 들어, 적어도 약 1 M, 예를 들어, 적어도 약 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2, 3, 4 또는 5 M의 농도), 열(예를 들어, 약 85, 86, 87, 88, 89, 90, 91, 92, 93, 94 또는 95°C 초과 온도) 또는 pH(예를 들어, 산성 pH 또는 염기성 pH)를 포함하는, 방법.
- [0153] 110. 실시형태 106 내지 109 중 어느 하나에 있어서, ORF1 분자를 포함하는 단백질성 외부의 분해에 적합한 조 건은 적어도 0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 0.9 M, 1 M, 1.1 M, 1.2 M, 1.3 M, 1.5 M, 1.5 M, 1.6 M, 1.7 M, 1.8 M, 1.9 M 또는 2 M의 농도의 요소를 포함하는 용액 중에서의 항온처리를 포 함하는, 방법.
- [0154] 111. 실시형태 106 내지 110 중 어느 하나에 있어서, (b)의 항온처리는 분해될 ORF1 분자 또는 그의 카피를 포 함하는, 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 96% 또는 100%의 입자의 집단을 초래하는, 방법.
- [0155] 112. 실시형태 106 내지 111 중 어느 하나에 있어서, 단백질성 외부의 분해에 적합한 조건은 ORF1 분자의, 적어 도 약 20, 30, 40, 50 또는 60개의 카피, 또는 20 내지 30, 30 내지 40, 40 내지 50 또는 50 내지 60개의 카피 를 포함하는 복합체(예를 들어, 단백질성 외부)를 분해하기에 충분한, 방법.
- [0156] 113. 실시형태 106 내지 112 중 어느 하나에 있어서, 단백질성 외부의 분해에 적합한 조건은  $10^8$ 개 미만의 남아 있는 온전한 입자/mL를 초래하는, 방법.
- [0157] 114. 실시형태 106 내지 113 중 어느 하나에 있어서, (a)를 제공하기 전에, 단백질성 외부의 분해에 적합한 조 건으로부터 ORF1 분자를 제거하는 단계(예를 들어, ORF1 분자가 비-변성 조건을 거치게 하는 단계)를 추가로 포 함하는, 방법.
- [0158] 115. 실시형태 114에 있어서, 단백질성 외부의 분해에 적합한 조건으로부터 ORF1 분자를 제거하는 단계는 변성 제의 농도를 감소시키는 단계, 예를 들어, 변성제(예를 들어, 요소)의 농도를 0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 0.9 M, 1 M, 1.1 M, 1.2 M, 1.3 M, 1.5 M, 1.5 M, 1.6 M, 1.7 M, 1.8 M, 1.9 M 또는 2 M 미만으로 감소시키는 단계를 포함하는, 방법.
- [0159] 116. 실시형태 114 또는 115에 있어서, 제거는 적어도 하나의 카피(예를 들어, 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900 또는 1000개의 카피)의 유전 요소(예를 들어, mRNA)를 각각 봉입하는 하나 이상의 아넬로벡터의 형성을 초 래하는, 방법.
- [0160] 117. 실시형태 116에 있어서, 생성된 용액 내의 유전 요소의 적어도 하나의 카피를 봉입하는 아넬로벡터의 수는 적어도  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$  또는  $10^{11}$ 개이거나;  $10^5$  내지  $10^6$ ,  $10^6$  내지  $10^7$ ,  $10^7$  내지  $10^9$ ,  $10^8$  내지  $10^9$ ,  $10^9$  내지  $10^{10}$ , 또는  $10^{10}$  내지  $10^{11}$ 개(예를 들어, 전자 현미경에 의해 측정되는 바와 같음)인, 방법.
- [0161] 118. 실시형태 116에 있어서, 생성된 용액 내의 유전 요소의 적어도 하나의 카피를 봉입하는 아넬로벡터의 수는 적어도  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$  또는  $10^{11}$ 개의 아넬로벡터/mL이거나;  $10^5$  내지  $10^6$ ,  $10^6$  내지  $10^7$ ,  $10^7$  내지  $10^9$ ,  $10^8$  내지  $10^9$ ,  $10^9$  내지  $10^{10}$ , 또는  $10^{10}$  내지  $10^{11}$ 개의 아넬로벡터/mL(예를 들어, 전자 현미경에 의해 측정 되는 바와 같음)인, 방법.
- [0162] 119. 실시형태 114 내지 118 중 어느 하나에 있어서, 제거는 적어도  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$  또는  $10^{11}$ 개의 아넬로벡터/mL; 또는  $10^5$  내지  $10^6$ ,  $10^6$  내지  $10^7$ ,  $10^7$  내지  $10^9$ ,  $10^8$  내지  $10^9$ ,  $10^9$  내지  $10^{10}$ , 또는  $10^{10}$  내지  $10^{11}$ 개의 아넬로벡터/mL(예를 들어, 전자 현미경에 의해 측정되는 바와 같음)을 포함하는 용액을 초래하며, 여기

서 아넬로바이러스 각각은 적어도 하나의 카피(예를 들어, 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900 또는 1000개의 카피)의 유전 요소(예를 들어, mRNA)를 봉입하는, 방법.

- [0163] 120. 실시형태 114 내지 119 중 어느 하나에 있어서, 적어도 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%의 아넬로바이러스는, 60 mer 또는 입자이거나 적어도 30, 31, 32, 33, 34 또는 35 nm의 직경의 입자인 단백질 질 성 외부를 포함하는, 방법.
- [0164] 121. 실시형태 93 내지 120 중 어느 하나에 있어서, 아넬로바이러스의 유전 요소는 엔도뉴클레아제(예를 들어, RNase)에 대해 저항성인, 방법.
- [0165] 122. 실시형태 93 내지 121 중 어느 하나에 있어서, (a)는 (i) 및 (ii)를 혼합하는 단계를 포함하는, 방법.
- [0166] 123. 실시형태 93 내지 122 중 어느 하나에 있어서, 무세포 시스템에서 수행되는, 방법.
- [0167] 124. 아넬로바이러스 조성물의 제조 방법으로서,
- [0168] (a) 선행하는 실시형태 중 어느 하나에 따른 복수의 아넬로바이러스 또는 조성물을 제공하는 단계;
- [0169] (b) 선택적으로, 다음 중 하나 이상에 대하여 복수를 평가하는 단계: 본원에 기재된 오염물질, 광학 밀도 측정치(예를 들어, OD 260), (예를 들어, HPLC에 의한) 입자 개수, 감염성(예를 들어, 형광 및/또는 ELISA에 의해 결정되는 바와 같은, 예를 들어, 입자:감염성 유닛 비); 및
- [0170] c) 예를 들어, (b)의 파라미터 중 하나 이상이 명시된 임계값을 만족한다면, 복수의 아넬로바이러스를 예를 들어, 대상체로의 투여에 적합한 약제학적 조성물로서 제형화하는 단계를 포함하는, 방법.
- [0171] 125. 질병 또는 장애의 치료를 필요로 하는 대상체에서 질병 또는 장애를 치료하는 방법으로서, 방법은 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물을 대상체에게 투여함으로써, 대상체에서 질병 또는 장애(예를 들어, 본원에 기재된 바와 같음)를 치료하는 단계를 포함하는, 방법.
- [0172] 126. 대상체에서 생물학적 기능(예를 들어, 본원에 기재된 바와 같음)을 조절, 예를 들어, 향상 또는 억제하는 방법으로서, 방법은 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물을 대상체에게 투여하는 단계를 포함하는, 방법.
- [0173] 127. 유전 요소를 세포에 운반하는 방법으로서, 방법은 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물을 세포, 예를 들어, 진핵 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포와 접촉시키는 단계를 포함하는, 방법.
- [0174] 128. 대상체에서 질병 또는 장애(예를 들어, 본원에 기재된 바와 같음)를 치료하기 위한 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물의 용도.
- [0175] 129. 대상체에서 질병 또는 장애(예를 들어, 본원에 기재된 바와 같음)를 치료하는 방법에서 사용하기 위한 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물.
- [0176] 130. 대상체에서의 질병 또는 장애(예를 들어, 본원에 기재된 바와 같음)의 치료용 의약의 제조에 사용하기 위한 선행하는 실시형태 중 어느 하나의 아넬로바이러스 또는 조성물.
- [0177] 본 발명의 다른 특징, 목적 및 이점이 설명 및 도면, 및 청구범위로부터 명백해질 것이다.
- [0178] 달리 정의되지 않는 한, 본원에 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 분야의 당업자에 의해 통상적으로 이해되는 바와 동일한 의미를 갖는다. 본원에 언급된 모든 간행물, 특허 출원, 특허 및 다른 참고문헌은 그들 전문이 참조로 포함된다. 또한, 물질, 방법 및 실시예는 단지 예시적인 것일 뿐이며, 제한하려는 것이 아니다.

**도면의 간단한 설명**

- [0179] 도 1은 아넬로바이러스 균주로부터의 재조합 캡시드 단백질이 시험관 내에서 바이러스-유사 입자(VLP)를 형성함을 보여주는 일련의 전자 현미경 사진이다. 상이한 세포주에서 생성된 캡시드 단백질은 음성 염색 전자 현미경에 의해 관찰되는 바와 같이 시험관 내에서 VLP를 형성한다. (A) 대략 35 nm의 관찰된 직경을 갖는 곤충 세포로부터 정제된 고리 2 ORF1 VLP. (B) 대략 35 nm의 관찰된 직경을 갖는 곤충 세포로부터 정제된 고리 10 ORF1 VLP. (C) 대략 20 nm의 관찰된 직경을 갖는 포유동물 세포로부터 정제된 CAV VP1 VLP.

도 2는 부리 및 깃털 질병 바이러스(BFDV)의 구조에서 관찰된 8개의 베타-가닥 젤리 롤 도메인(또는 젤리 롤 폴드)을 나타낸 다이어그램이다. 관례에 따라, 베타 가닥은 B에서 I로 표시된다. 가닥은 B-I-D-G 및 C-H-E-F의 배향을 갖는 4개의 역평행 베타 시트를 형성한다. B-I-D-G 시트는 바이러스 캡시드의 내부를 형성한다.

도 3은 부리 및 깃털 질병 바이러스(BFDV)/E형 간염 캡시드 단백질(HepE)의 젤리 롤 도메인과 비교하여 아넬로바이러스 ORF1 단백질에 대한 젤리 롤 서열을 도시하는 아미노산 서열 정렬이다.

도 4a 내지 도 4e는 eGFP를 인코딩하는 mRNA 분자를 봉입하는 아넬로바이러스 ORF1 단백질-기반 바이러스-유사 입자(VLP)를 생성하는 예시적인 방법을 보여주는 일련의 다이어그램이다. 도 4a: ORF1 단백질을 세포에서 생성하고 본원에 기재된 바와 같이 단리하였다. 도 4b: 그런 다음, ORF1 단백질로부터 형성된 VLP를 2 M 요소 용액에서 분해시켰다. 도 4c: 요소가 mRNA의 부재 하에 제거되었을 때, VLP가 거의 재형성되지 않았다(전자 현미경에 의해 검출된  $10^8$  개 입자/mL 미만의 역가). 도 4d, 도 4e: 요소가 eGFP를 인코딩하는 mRNA의 존재 하에 제거되었을 때, 상당한 양의 VLP( $1 \times 10^9$  내지  $1 \times 10^{10}$  개 입자/mL의 역가)가 전자 현미경(EM)에 의해 검출되었다.

본 발명의 실시형태의 다음의 상세한 설명은 첨부된 도면과 함께 관독하는 경우 더 잘 이해될 것이다. 본 발명의 예시의 목적을 위하여, 본원에 예시된 실시형태가 도면에 나타나 있다. 그러나, 본 발명이 도면에 나타난 실시형태의 정확한 배열 및 수단으로 제한되지 않음을 이해해야 한다.

### 발명을 실시하기 위한 구체적인 내용

- [0180] 정의
- [0181] 본 발명은 특정 실시형태에 대해 그리고 특정 도면을 참조하여 설명될 것이지만 본 발명은 이것에 한정되지 않고, 청구범위에 의해서만 한정된다. 이하에 기재된 용어는 일반적으로 다르게 지시되지 않는 한 그들의 일반적인 의미로 이해될 것이다.
- [0182] 발명을 실시하기 위한 구체적인 내용 및 청구범위에서 "포함하다"라는 용어가 사용되는 경우, 이는 다른 요소를 배제하지 않는다. 본 발명의 목적을 위하여, "로 이루어지다"라는 용어는 "구성된다"라는 용어의 바람직한 실시형태인 것으로 간주된다. 이하에서 그룹이 적어도 특정 개수의 실시형태를 포함하는 것으로 정의되는 경우, 이것은 또한 바람직하게는 이들 실시형태만으로 이루어지는 그룹을 개시하는 것으로 이해될 것이다.
- [0183] 단수 명사를 언급할 때, 부정 관사 또는 정관사, 예를 들어, 하나의("a", "an") 또는 상기("the")가 사용되는 경우, 이것은 그 밖의 어떤 것이 특별히 언급되지 않는 한, 그 명사의 복수형을 포함한다.
- [0184] 용어 "치료, 조절 등을 위한 화합물, 조성물, 생성물 등"은 표기된 치료, 조절 등의 목적에 적합한 화합물, 조성물, 생성물 등 그 자체를 지칭함을 이해할 것이다. 용어 "치료, 조절 등을 위한 화합물, 조성물, 생성물 등"은 추가로 일 실시형태로서, 이러한 화합물, 조성물, 생성물 등이 치료, 조절 등에 사용하기 위한 것임을 개시한다.
- [0185] 용어 "...에 사용하기 위한 화합물, 조성물, 생성물 등", "...를 위한 의약, 약제학적 조성물, 수의학적 조성물, 진단적 조성물 등의 제조에서의 화합물, 조성물, 생성물 등의 용도", 또는 "의약으로서 사용하기 위한 화합물, 조성물, 생성물 등..."은 이러한 화합물, 조성물, 생성물 등이 인간 또는 동물 신체에서 실시될 수 있는 치료적 방법에 사용될 것임을 나타낸다. 그들은 치료 방법 등에 관한 실시형태 및 청구범위의 동등한 개시로서 간주된다. 따라서, 실시형태 또는 청구범위가 "질병을 앓는 것으로 의심되는 인간 또는 동물을 치료하는 데 사용하기 위한 화합물"을 지칭한다면, 이것은 또한 "질병을 앓는 것으로 의심되는 인간 또는 동물을 치료하기 위한 의약의 제조에서의 화합물의 용도" 또는 "질병을 앓는 것으로 의심되는 인간 또는 동물로의 화합물의 투여에 의한 치료 방법"의 개시인 것으로 간주된다. 용어 "치료, 조절 등을 위한 화합물, 조성물, 생성물 등"은 표기된 치료, 조절 등의 목적에 적합한 화합물, 조성물, 생성물 등 그 자체를 지칭함을 이해할 것이다.
- [0186] 기간, 값, 수 등의 하기의 예가 괄호 안에 제공된다면, 이것은 괄호 안에 언급된 예가 실시형태를 구성할 수 있다는 표시로서 이해해야 한다. 예를 들어, "실시형태들에서, 핵산 분자가 표 1의 아넬로바이러스 ORF1-인코딩 뉴클레오티드 서열(예를 들어, 표 1의 핵산 서열의 뉴클레오티드 571 내지 2613)에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다"고 언급된다면, 일부 실시형태는 표 1의 핵산 서열의 뉴클레오티드 571 내지 2613에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함하는 핵산 분자에 관한 것이다.
- [0187] 본원에 사용되는 바와 같이, 용어 "증폭"은 하나 이상의 추가의 카피의 핵산 분자 또는 그의 일부(예를 들어,

유전 요소 또는 유전 요소 영역)를 생성하기 위한, 핵산 분자 또는 그의 일부의 복제를 지칭한다. 일부 실시형태에서, 증폭은 핵산 서열의 부분 복제를 초래한다. 일부 실시형태에서, 증폭은 회전원 복제(rolling circle replication)를 통해 발생한다.

[0188] 본원에 사용되는 바와 같이, 용어 "아넬로백터"는 단백질성 외부에 봉입된 유전 요소, 예를 들어, RNA, 예를 들어, 환형 RNA를 포함하는 비히클을 지칭한다. 일부 실시형태에서, 유전 요소는 단백질성 외부에 의해 RNase를 사용하는 소화로부터 실질적으로 보호된다. 본원에 사용되는 바와 같이, "합성 아넬로백터"는 일반적으로, 자연 발생이 아닌, 예를 들어, 야생형 바이러스(예를 들어, 본원에 기재된 바와 같은 야생형 아넬로바이러스)에 비하여 상이한 서열을 갖는 아넬로백터를 지칭한다. 일부 실시형태에서, 합성 아넬로백터는 엔지니어링되거나 또는 재조합이며, 예를 들어, 야생형 바이러스 게놈(예를 들어, 본원에 기재된 바와 같은 야생형 아넬로바이러스 게놈)에 비하여 차이 또는 변형을 포함하는 유전 요소를 포함한다. 일부 실시형태에서, 단백질성 외부 내의 봉입은 단백질성 외부에 의한 100% 커버리지, 및 100% 미만, 예를 들어, 95%, 90%, 85%, 80%, 70%, 60%, 50% 또는 그 미만의 커버리지를 포괄한다. 유전 요소가 예를 들어, 숙주 세포로의 유입 이전에 단백질성 외부에 보유되거나 RNase를 사용하는 소화로부터 보호되는 한, 예를 들어, (예를 들어, 단백질성 외부를 물, 이온, 펩티드 또는 소분자에 대해 투과 가능하게 만드는) 캡 또는 불연속부가 단백질성 외부에 존재할 수 있다. 일부 실시형태에서, 아넬로백터는 정제되며, 예를 들어, 아넬로백터는 그의 원래의 공급원으로부터 분리되고/분리되거나 다른 구성성분이 실질적으로 없다(50% 초과, 60% 초과, 70% 초과, 80% 초과, 90% 초과). 일부 실시형태에서, 아넬로백터는 (예를 들어, 감염을 통해) 표적 세포 내로 유전 요소를 도입할 수 있다. 일부 실시형태에서, 아넬로백터는 아넬로바이러스 ORF1 분자와 같은 특정 아넬로바이러스 요소를 함유하는 감염성 합성 바이러스 입자이다.

[0189] 본원에 사용되는 바와 같이, 용어 "항체 분자"는 단백질, 예를 들어, 적어도 하나의 면역글로불린 가변 도메인 서열을 포함하는 면역글로불린쇄 또는 그의 단편을 지칭한다. 용어 "항체 분자"는 전장 항체 및 항체 단편(예를 들어, scFv)을 포괄한다. 일부 실시형태에서, 항체 분자는 다중특이적 항체 분자이며, 예를 들어, 항체 분자는 복수의 면역글로불린 가변 도메인 서열을 포함하며, 복수의 것들 중 제1 면역글로불린 가변 도메인 서열은 제1 에피토프에 대한 결합 특이성을 가지며, 복수의 것들 중 제2 면역글로불린 가변 도메인 서열은 제2 에피토프에 대한 결합 특이성을 갖는다. 일부 실시형태에서, 다중특이적 항체 분자는 이중특이적 항체 분자이다. 이중특이적 항체 분자는 일반적으로 제1 에피토프에 대한 결합 특이성을 갖는 제1 면역글로불린 가변 도메인 서열 및 제2 에피토프에 대한 결합 특이성을 갖는 제2 면역글로불린 가변 도메인 서열에 의해 특성화된다.

[0190] 본원에 사용되는 바와 같이, 용어 "백본" 또는 "백본 영역"은 숙주 세포에서 핵산 분자의 복제 및/또는 유지에 관련된(예를 들어, 이에 필요하고/하거나 충분한) 하나 이상의 요소를 포함하는, (예를 들어, 본원에 기재된 바와 같은, 박미드 또는 공여자 백터 내의) 핵산 분자 내의 영역을 지칭한다. 일부 실시형태에서, "배칼로바이러스 백본 영역"과 같은 백본 영역은 예를 들어, 곤충 세포(예를 들어, Sf9 세포)에서 핵산 작제물의 복제에 적합한 하나 이상의 배칼로바이러스 요소(예를 들어, 배칼로바이러스 게놈 또는 그의 기능적 단편)를 포함한다. 일부 실시형태에서, 백본은 선택 가능한 마커를 추가로 포함한다. 일부 실시형태에서, 핵산 분자는 유전 요소 영역 및 백본 영역(예를 들어, 배칼로바이러스 백본 영역 및/또는 박테리아 세포에서의 복제에 적합한 백본 영역)을 포함한다.

[0191] 본원에 사용되는 바와 같이, 용어 "박미드"는 곤충 세포에서의 복제에 적합하고 또한 박테리아 세포에서의 복제에 적합하도록 충분한 배칼로바이러스 백본 요소를 포함하는 핵산 분자를 지칭한다. 일부 실시형태에서, 핵산 분자는 박테리아 세포(예를 들어, 이. 콜라이(*E. coli*) 세포, 예를 들어, DH 10Bac 세포)에서의 복제에 적합하다.

[0192] 본원에 사용되는 바와 같이, "환형" 핵산은 자유 5' 또는 3' 말단이 없는 구조를 형성하는 핵산을 지칭한다. 일부 실시형태에서, 환형 핵산은 공유 또는 비-공유 결합을 통해 폐쇄된다. 예를 들어, 환형 핵산은 선형 핵산의 말단을 예를 들어, 인산염-당 결합 또는 합성 링커 모이어티와 공유적으로 연결함으로써 제조될 수 있다. 다른 실시형태에서, 환형 핵산은 근접하고 자유롭지 않은(실질적으로 엑소뉴클레아제에 접근할 수 없는) 2개의 말단을 포함한다. 예를 들어, 환형 핵산은 선형 핵산의 말단을 직접 또는 핵산 부목을 통해 혼성화시킴으로써 제조될 수 있다.

[0193] 본원에 사용되는 바와 같이, "DNA 영역"은 복수의 DNA 뉴클레오티드를 포함하는 폴리뉴클레오티드 가닥의 일부를 지칭한다. 예를 들어, 일부 실시형태에서, DNA 영역은 RNA 가닥 내로 통합된 복수의 DNA 뉴클레오티드이다. 예를 들어, DNA 영역은 폴리뉴클레오티드 가닥 내에 약 5 내지 10, 10 내지 20, 20 내지 30, 30 내지 40, 40

내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90 또는 90 내지 100개의 DNA 뉴클레오티드를 포함한다.

- [0194] 본원에 사용되는 바와 같이, "인코딩"하는 핵산은 아미노산 서열 또는 폴리뉴클레오티드, 예를 들어, mRNA 또는 기능적 폴리뉴클레오티드(예를 들어, 비-코딩 RNA, 예를 들어, siRNA 또는 miRNA)를 인코딩하는 핵산 서열을 지칭한다.
- [0195] 본원에 사용되는 바와 같이, "외인성" 작용제(예를 들어, 이펙터, 핵산(예를 들어, RNA), 유전자, 페이로드, 단백질)는 상응하는 야생형 바이러스, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스에 의해 포함되지 않거나, 그에 의해 인코딩되지 않는 작용제를 지칭한다. 일부 실시형태에서, 외인성 작용제, 예컨대, 자연 발생 단백질 또는 핵산에 비하여(예를 들어, 삽입, 결실 또는 치환에 의해) 변경된 서열을 갖는 단백질 또는 핵산은 자연적으로 존재하지 않는다. 일부 실시형태에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지 않는다. 일부 실시형태에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지만, 바이러스에 대하여 외인성이다. 일부 실시형태에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지만, 요망되는 수준으로 또는 요망되는 시간에 존재하지 않는다.
- [0196] 또 다른 작용제 또는 요소(예를 들어, 이펙터, 핵산 서열, 아미노산 서열)에 관하여 본원에 사용되는 바와 같은, "이종" 작용제 또는 요소(예를 들어, 이펙터, 핵산 서열, 아미노산 서열)는 천연적으로, 예를 들어, 야생형 바이러스, 예를 들어, 아넬로바이러스에서 함께 관찰되지 않는 작용제 또는 요소를 지칭한다. 일부 실시형태에서, 이종 핵산 서열은 천연 발생 핵산 서열(예를 들어, 아넬로바이러스에서 천연적으로 발생하는 서열)과 동일한 핵산에 존재할 수 있다. 일부 실시형태에서, 이종 작용제 또는 요소는 아넬로바이러스의 다른(예를 들어, 나머지) 요소가 기반으로 하는 아넬로바이러스에 비하여 외인성이다.
- [0197] 본원에 사용되는 바와 같이, 용어 "유전 요소"는 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스를 형성하기 위해 단백질성 외부 내에 봉입되거나 봉입될 수 있는(예를 들어, 이에 의해 RNase 소화로부터 보호됨) 핵산 분자를 지칭한다. 유전 요소는 네이키드 RNA로서 생성될 수 있고, 선택적으로 추가로 단백질성 외부 내로 어셈블될 수 있음이 이해된다. 또한, 아넬로바이러스는 그의 유전 요소를 세포 내로 삽입하여, 유전 요소가 세포 내에 존재하게 하고, 단백질성 외부가 반드시 세포에 유입되는 것은 아니라는 점이 이해된다.
- [0198] 본원에 사용되는 바와 같이, "유전 요소 작제물"은 유전 요소 서열 또는 그의 단편을 포함하는 핵산 작제물(예를 들어, 플라스미드, 박미드, 공여자 벡터, 코스미드 또는 미니서클)을 지칭한다. 일부 실시형태에서, 본원에 기재된 바와 같은 박미드 또는 공여자 벡터는 유전 요소 서열 또는 그의 단편을 포함하는 유전 요소 작제물이다.
- [0199] 본원에 사용되는 바와 같이, 용어 "유전 요소 영역"은 유전 요소의 서열을 포함하는 작제물의 영역을 지칭한다. 일부 실시형태에서, 유전 요소 영역은 단백질성 외부에 의해 봉입되며 이에 따라 아넬로바이러스를 형성할, 야생형 아넬로바이러스 서열 또는 그의 단편에 대하여 충분한 동일성을 갖는 서열(예를 들어, 야생형 아넬로바이러스 서열 또는 그의 단편에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열)을 포함한다. 실시형태들에서, 유전 요소 영역은 예를 들어, 본원에 기재된 바와 같은 단백질 결합 서열(예를 들어, 본원에 기재된 바와 같은 5' UTR, 3' UTR 및/또는 GC-풍부 영역, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열)을 포함한다. 일부 실시형태에서, 유전 요소 영역을 포함하는 작제물은 단백질성 외부에서 봉입되지 않지만, 작제물로부터 생성된 유전 요소는 단백질성 외부에서 봉입될 수 있다. 일부 실시형태에서, 유전 요소 영역을 포함하는 작제물은 벡터 백본(예를 들어, 박미드 백본 또는 공여자 벡터 백본)을 추가로 포함한다. 일부 실시형태에서, 작제물(예를 들어, 박미드)은 하나 이상의 배칼로바이러스 요소(예를 들어, 유전 요소 영역을 포함하는 예를 들어, 배칼로바이러스 계놈)를 포함한다.
- [0200] 본원에 사용되는 바와 같이, 용어 "돌연변이체"는 계놈(예를 들어, 아넬로바이러스 계놈) 또는 그의 단편과 관련하여 사용될 때 상응하는 야생형 아넬로바이러스 서열에 비하여 적어도 하나의 변화를 갖는 서열을 지칭한다. 일부 실시형태에서, 돌연변이체 계놈 또는 그의 단편은 상응하는 야생형 아넬로바이러스 서열에 비하여 적어도 하나의 단일 뉴클레오티드 다형성, 부가, 결실 또는 프레임이동을 포함한다. 일부 실시형태에서, 돌연변이체 계놈 또는 그의 단편은 상응하는 야생형 아넬로바이러스 서열에 비하여 적어도 하나의 아넬로바이러스 ORF(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및/또는 ORF1/2 중 하나 이상)의 결실을 포함한다. 일부 실시형태에서, 돌연변이체 계놈 또는 그의 단편은 상응하는 야생형 아넬로바이러스 서열에 비하여 모든 아넬로바이러스 ORF(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및 ORF1/2 모두)의 결실을 포함한다. 일부 실시형태에서,

돌연변이체 게놈 또는 그의 단편은 상응하는 야생형 아넬로바이러스 서열에 비하여 적어도 하나의 아넬로바이러스 비코딩 영역(예를 들어, 5' UTR, 3' UTR 및/또는 GC-풍부 영역 중 하나 이상)의 결실을 포함한다. 일부 실시 형태에서, 돌연변이체 게놈 또는 그의 단편은 외인성 이펙터를 포함하거나, 이를 인코딩한다.

- [0201] 본원에 사용되는 바와 같이, 용어 "ORF 분자"는 아넬로바이러스 ORF 단백질(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및/또는 ORF1/2 단백질을 포함하는 폴리펩티드) 또는 그의 기능적 단편의 활성 및/또는 구조적 특징을 갖는 폴리펩티드를 지칭한다. 일반적으로 (즉, "ORF 분자"로) 사용되는 경우, 폴리펩티드는 본원에 기재된 아넬로바이러스 ORF 중 임의의 것(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및/또는 ORF1/2 중 임의의 것) 또는 그의 기능적 단편의 활성 및/또는 구조적 특징을 포함할 수 있다. 특정 오픈 리딩 프레임(예를 들어, "ORF1 분자", "ORF2 분자", "ORF2/2 분자", "ORF2/3 분자", "ORF1/1 분자" 또는 "ORF1/2 분자")를 나타내기 위해 수식어와 함께 사용될 때, 이는 일반적으로 폴리펩티드가 상응하는 아넬로바이러스 ORF 단백질 또는 그의 기능적 단편(예를 들어, "ORF1 분자"에 대해 아래에 정의된 바와 같음)의 활성 및/또는 구조적 특징을 포함함을 의미한다. 예를 들어, "ORF2 분자"는 아넬로바이러스 ORF2 단백질 또는 그의 기능적 단편의 활성 및/또는 구조적 특징을 포함한다.
- [0202] 본원에 사용되는 바와 같이, 용어 "ORF1 분자"는 아넬로바이러스 ORF1 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질) 또는 그의 기능적 단편의 활성 및/또는 구조적 특징을 갖는 폴리펩티드를 지칭한다. ORF1 분자는 일부 예에서, 다음 중 하나 이상(예를 들어, 이 중 1, 2, 3 또는 4개)을 포함할 수 있다: 적어도 60%의 염기성 잔기(예를 들어, 적어도 60%의 아르기닌 잔기)를 포함하는 제1 영역, 적어도 약 6개의 베타 가닥(예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10, 11 또는 12개의 베타 가닥)을 포함하는 제2 영역, 아넬로바이러스 N22 도메인(예를 들어, 본원에 기재된 바와 같음, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 N22 도메인)의 구조 또는 활성을 포함하는 제3 영역 및/또는 아넬로바이러스 C-말단 도메인(CTD)(예를 들어, 본원에 기재된 바와 같음, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 CTD)의 구조 또는 활성을 포함하는 제4 영역. 일부 예에서, ORF1 분자는 N-말단에서 C-말단 순서로, 제1, 제2, 제3 및 제4 영역을 포함한다. 일부 예에서, 아넬로바이러스는 N-말단에서 C-말단 순서로, 제1, 제2, 제3 및 제4 영역을 포함하는 ORF1 분자를 포함한다. ORF1 분자는 일부 예에서, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함할 수 있다. ORF1 분자는 일부 예에서, 이중 서열, 예를 들어, 추가 변 영역(HVR), 예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 HVR을 추가로 포함할 수 있다. 본원에 기재된 바와 같은, "아넬로바이러스 ORF1 단백질"은 아넬로바이러스 게놈(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 게놈)에 의해 인코딩된 ORF1 단백질을 지칭한다.
- [0203] 본원에 사용되는 바와 같이, 용어 "ORF2 분자"는 아넬로바이러스 ORF2 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF2 단백질)의 활성 및/또는 구조적 특징을 갖는 폴리펩티드 또는 그의 기능적 단편을 지칭한다. 본원에 사용되는 바와 같이, "아넬로바이러스 ORF2 단백질"은 아넬로바이러스 게놈(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 게놈)에 의해 인코딩된 ORF2 단백질을 지칭한다.
- [0204] 본원에 사용되는 바와 같이, 용어 "단백질성 외부"는 주로(예를 들어, 50% 초과, 60% 초과, 70% 초과, 80% 초과, 90% 초과) 단백질인 외부 구성성분을 지칭한다.
- [0205] 본원에 사용되는 바와 같이, 용어 "조절 핵산"은 발현 생성물을 인코딩하는 DNA 서열의 발현, 예를 들어, 전사 및/또는 번역을 변경시키는 핵산 서열을 지칭한다. 일부 실시 형태에서, 발현 생성물은 RNA 또는 단백질을 포함한다.
- [0206] 본원에 사용되는 바와 같이, 용어 "조절 서열"은 표적 유전자 생성물의 전사를 변경시키는 핵산 서열을 지칭한다. 일부 실시 형태에서, 조절 서열은 프로모터 또는 인핸서이다.
- [0207] 본원에 사용되는 바와 같이, "실질적으로 비-병원성인" 유기체, 입자 또는 구성성분은 예를 들어, 숙주 유기체, 예를 들어, 포유동물, 예를 들어, 인간에서 받아들일 수 없는 질병 또는 병원성 질환을 야기하거나 유도하지 않는 유기체, 입자(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로바이러스) 또는 그의 구성성분을 지칭한다. 일부 실시 형태에서, 대상체로의 아넬로바이러스의 투여는 치료 기준의 일부로서 허용되는 부작용 또는 부작용을 초래할 수 있다.
- [0208] 본원에 사용되는 바와 같이, 용어 "비-병원성"은 숙주 유기체, 예를 들어, 포유동물, 예를 들어, 인간에서 바람직하지 않은 질환(예를 들어, 질병 또는 병원성 질환을 야기하거나 유도하지 않는 유기체 또는 그의 구성성분을 지칭한다.

[0209] 본원에 사용되는 바와 같이, "실질적으로 비-통합성인" 유전 요소는 숙주 세포(예를 들어, 진핵 세포) 또는 유기체(예를 들어, 포유동물, 예를 들어, 인간) 내로 유입되는 유전 요소 중 약 0.01%, 0.05%, 0.1%, 0.5% 또는 1% 미만이 게놈 내로 통합되는 유전 요소, 예를 들어, 본원에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로박터 내의 유전 요소를 지칭한다. 일부 실시형태에서, 유전 요소는 예를 들어, 숙주 세포의 게놈 내로 검출 가능하게 통합되지 않는다. 일부 실시형태에서, 게놈 내로의 유전 요소의 통합은 본원에 기재된 바와 같은 기법, 예를 들어, 핵산 서열결정, PCR 검출 및/또는 핵산 혼성화를 사용하여 검출될 수 있다. 일부 실시형태에서, 통합 빈도는 예를 들어, 문헌 [Wang et al. (2004, *Gene Therapy* 11: 711-721, 그의 전문이 본원에 참조로 포함됨)]에 기재된 바와 같이 유리 벡터로부터 분리된 게놈 DNA의 정량적 겔 정제 검정에 의해 측정된다.

[0210] 본원에 사용되는 바와 같이, "실질적으로 비-면역원성인" 유기체, 입자 또는 구성성분은 예를 들어, 숙주 조직 또는 유기체(예를 들어, 포유동물, 예를 들어, 인간)에서 요망되지 않는 또는 비표적화된 면역 반응을 야기하거나 유도하지 않는 유기체, 입자(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로박터) 또는 그의 구성성분을 지칭한다. 일부 실시형태에서, 실질적으로 비-면역원성인 유기체, 입자 또는 구성성분은 임상적으로 유의한 면역 반응을 생성하지 않는다. 일부 실시형태에서, 실질적으로 비-면역원성인 아넬로박터는 아넬로바이러스 또는 아넬로박터 유전 요소의 아미노산 서열을 포함하는 단백질 또는 그의 핵산 서열에 의해 인코딩된 단백질에 대하여 임상적으로 유의한 면역 반응을 생성하지 않는다. 일부 실시형태에서, 면역 반응(예를 들어, 요망되지 않는 또는 비표적화된 면역 반응)은 예를 들어, 문헌 [Tsuda et al., 1999; *J. Virol. Methods* 77: 199-206](본원에 참조로 포함됨)에 기재된 항-TTV 항체 검출 방법 및/또는 문헌 [Kakkola et al., 2008; *Virology* 382: 182-189](본원에 참조로 포함됨)에 기재된 항-TTV IgG 수준의 결정 방법에 따라 대상체에서 항체(예를 들어, 중화 항체) 존재 또는 수준(예를 들어, 항-아넬로박터 항체의 존재 또는 수준, 예를 들어, 본원에 기재된 바와 같은 아넬로박터에 대한 항체의 존재 또는 수준)을 검정함으로써 검출된다. 아넬로바이러스 또는 그에 기반한 아넬로박터에 대한 항체(예를 들어, 중화 항체)는 또한 항-바이러스 항체를 검출하기 위한 해당 분야의 방법, 예를 들어, 문헌 [Calcedo et al., 2013; *Front. Immunol.* 4(341): 1-7](본원에 참조로 포함됨)에 기재된 바와 같은 예를 들어, 항-AAV 항체의 검출 방법에 의해 검출될 수 있다.

[0211] 본원에 사용되는 바와 같은 "하위서열"은 각각 더 큰 핵산 서열 또는 아미노산 서열에 포함되는 핵산 서열 또는 아미노산 서열을 지칭한다. 일부 예에서, 하위서열은 더 큰 서열의 도메인 또는 기능적 단편을 포함할 수 있다. 일부 예에서, 하위서열은 더 큰 서열로부터 단리되는 경우, 더 큰 서열의 나머지와 함께 존재할 때 하위서열에 의해 형성되는 2차 및/또는 3차 구조와 유사한 2차 및/또는 3차 구조를 형성할 수 있는 더 큰 서열의 단편을 포함할 수 있다. 일부 예에서, 하위서열은 또 다른 서열(예를 들어, 외인성 서열 또는 더 큰 서열의 나머지에 대하여 이중인 서열을 포함하는 하위서열, 예를 들어, 상이한 아넬로바이러스 유래의 상응하는 하위서열)에 의해 대체될 수 있다.

[0212] 본 발명은 일반적으로 아넬로박터, 예를 들어, 합성 아넬로박터 및 그의 용도에 관한 것이다. 본 개시내용은 아넬로박터, 아넬로박터를 포함하는 조성물, 및 아넬로박터의 제조 또는 이용 방법을 제공한다. 아넬로박터는 일반적으로 예를 들어, 치료제를 진핵 세포로 운반하기 위한 운반 비히클로서 유용하다. 일반적으로, 본원에 기재된 아넬로박터는 단백질성 외부 내에 봉입된 RNA 서열(예를 들어, 이펙터, 예를 들어, 외인성 이펙터 또는 내인성 이펙터를 인코딩하는 RNA 서열)을 포함하는 유전 요소를 포함할 것이다. 아넬로박터는 (예를 들어, 본원에 기재된 바와 같은) 아넬로바이러스 서열에 비한 서열(예를 들어, 본원에 기재된 바와 같은 영역 또는 도메인)의 하나 이상의 결실을 포함할 수 있다. 아넬로박터는 예를 들어, 세포를 포함하는 대상체에서 질병 또는 장애를 치료하기 위하여, 유전 요소 또는 그 내에 인코딩된 이펙터(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 폴리펩티드 또는 핵산 이펙터)를 진핵 세포 내로 운반하기 위한 실질적으로 비-면역원성인 비히클로서 사용될 수 있다.

[0213] **목차**

[0214] I. 시험관 내 어셈블리에 의한 아넬로박터 제조용 조성물 및 방법

[0215] A. 아넬로박터의 구성성분 및 어셈블리

[0216] i. 아넬로박터의 어셈블리를 위한 ORF1 분자

[0217] ii. 아넬로박터의 어셈블리를 위한 ORF2 분자

[0218] B. 유전 요소

- [0219] i. RNA를 포함하는 유전 요소
- [0220] a. RNA-전용 유전 요소
- [0221] b. 혼성화된 RNA-ssDNA 유전 요소
- [0222] c. RNA/DNA 컨주게이트
- [0223] C. 유전 요소 작제물
- [0224] D. RNA-기반 유전 요소의 생성
- [0225] E. 단백질 구성성분의 생성
- [0226] i. 배큘로바이러스 발현 시스템
- [0227] ii. 곤충 세포 시스템
- [0228] iii. 포유동물 세포 시스템
- [0229] F. 이펙터
- [0230] G. 시험관 내 어셈블리 방법
- [0231] H. 농축 및 정제
- [0232] II. 아넬로벡터
- [0233] A. 아넬로바이러스
- [0234] B. ORF1 분자
- [0235] C. ORF2 분자
- [0236] D. 유전 요소
- [0237] E. 단백질 결합 서열
- [0238] F. 5' UTR 영역
- [0239] G. GC-풍부 영역
- [0240] H. 이펙터
- [0241] I. 조절 서열
- [0242] J. 기타 서열
- [0243] K. 단백질성 외부
- [0244] III. 이용 방법
- [0245] IV. 투여/운반

[0246] **I. 시험관 내 어셈블리에 의한 아넬로벡터 제조용 조성물 및 방법**

[0247] 본 개시내용은, 일부 양태에서, 아넬로벡터, 예를 들어, 본원에 기재된 바와 같은, RNA를 포함하는 유전 요소를 갖는 아넬로벡터를 생성하기 위해 사용될 수 있는 조성물 및 방법을 제공한다. 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 유전 요소 또는 유전 요소 작제물을 생성하기 위해 사용될 수 있다. 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 시험관 내 어셈블리에 의해 유전 요소 또는 유전 요소 작제물을 생성하기 위해 사용될 수 있다. 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 하나 이상의 아넬로바이러스 ORF 분자(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 또는 ORF1/2 분자, 또는 그의 기능적 단편 또는 스플라이스 변이체)를 생성하기 위해 사용될 수 있다. 일부 실시형태에서, 본원에 기재된 조성물 및 방법은 예를 들어, 숙주 세포(예를 들어, 곤충 세포, 예를 들어, Sf9 세포)에서 단백질성 외부 또는 그의 구성성분(예를 들어, ORF1 분자)을 생성하기 위해 사용될 수 있다.

[0248] **아넬로벡터의 구성성분 및 어셈블리**

[0249] 본원의 조성물 및 방법은 아넬로백터를 생성하기 위해 사용될 수 있다. 본원에 기재된 바와 같이, 아넬로백터는 일반적으로 (예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함하는) 단백질성 외부 내에 봉입된 유전 요소(예를 들어, RNA 분자)를 포함한다. 일부 실시형태에서, 유전 요소는 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 또는 ORF1/2 중 하나 이상)를 인코딩하는 하나 이상의 서열을 포함한다. 본원에 사용된 바와 같이, 아넬로바이러스 ORF 또는 ORF 분자(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2)는 예를 들어, PCT/US2018/037379호 또는 PCT/US19/65995호(이들 각각은 그의 전문이 본원에 참조로 포함됨)에 기재된 바와 같은 상응하는 아넬로바이러스 ORF 서열에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드를 포함한다. 실시형태들에서, 유전 요소는 아넬로바이러스 ORF1, 또는 그의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본원에 기재된 바와 같은 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 ORF1 핵산(예를 들어, 아넬로바이러스 ORF1 분자 또는 그의 스플라이스 변이체 또는 기능적 단편)에 의해 인코딩된 폴리펩티드를 포함한다.

[0250] 일부 실시형태에서, 아넬로백터는 단백질성 외부(예를 들어, 본원에 기재된 바와 같은) 내에 유전 요소(예를 들어, 본원에 기재된 바와 같은)를 봉입함으로써 어셈블된다. 일부 실시형태에서, 유전 요소는 숙주 세포(예를 들어, 곤충 세포, 예를 들어, Sf9 세포)에서 단백질성 외부 내에 봉입되어 있다. 일부 실시형태에서, 숙주 세포는 단백질성 외부에 포함된 하나 이상의 폴리펩티드(예를 들어, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드, 예를 들어, ORF1 분자)를 발현한다. 예를 들어, 일부 실시형태에서, 숙주 세포는 아넬로바이러스 ORF1 분자, 예를 들어, 아넬로바이러스 ORF1 폴리펩티드의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 ORF1 단백질 또는 야생형 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드)을 인코딩하는 핵산 서열을 포함한다. 실시형태들에서, 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열은 숙주 세포에 포함된 핵산 작제물(예를 들어, 플라스미드, 바이러스 벡터, 바이러스, 미니서클, 박미드 또는 인공 염색체)에 포함된다. 실시형태들에서, 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열은 숙주 세포의 게놈 내로 통합된다.

[0251] 일부 실시형태에서, 숙주 세포는 유전 요소 및/또는 유전 요소의 서열을 포함하는 핵산 작제물을 포함한다. 일부 실시형태에서, 핵산 작제물은 플라스미드, 바이러스 핵산, 미니서클, 박미드 또는 인공 염색체로부터 선택된다. 일부 실시형태에서, 유전 요소는 핵산 작제물(예를 들어, 박미드)로부터 절제되고, 선택적으로, (예를 들어, 변성에 의해) 이중-가닥 형태로부터 단일-가닥 형태로 전환된다. 일부 실시형태에서, 유전 요소는 핵산 작제물(예를 들어, 박미드)의 주형 서열에 기반한 중합효소에 의해 생성된다. 일부 실시형태에서, 중합효소는 유전 요소 서열의 단일-가닥 카피를 생성하며, 이는 본원에 기재된 바와 같은 유전 요소를 형성하기 위해 선택적으로 환화될 수 있다.

[0252] 일부 실시형태에서, 숙주 세포는 유전 요소 작제물(예를 들어, 박미드, 플라스미드 또는 미니서클) 및 아넬로바이러스 ORF 분자(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및/또는 ORF1/2 ORF 분자) 또는 그의 기능적 단편을 인코딩하는 하나 이상의 서열을 포함하는 박미드를 포함한다. 일부 실시형태에서, 단백질성 외부 단백질은 박미드로부터 발현된다. 실시형태들에서, 박미드로부터 발현된 단백질성 외부 단백질은 유전 요소를 봉입하며, 이에 따라 아넬로백터를 형성한다. 일부 실시형태에서, 박미드는 곤충 세포(예를 들어, Sf9 세포)에서의 핵산 작제물의 복제에 적합한 백본, 예를 들어, 배칼로바이러스 백본 영역을 포함한다. 일부 실시형태에서, 박미드는 박테리아 세포(예를 들어, 이. 콜라이 세포, 예를 들어, DH 10Bac 세포)에서의 유전 요소 작제물의 복제에 적합한 백본 영역을 포함한다. 일부 실시형태에서, 유전 요소 작제물은 곤충 세포(예를 들어, Sf9 세포)에서의 핵산 작제물의 복제에 적합한 백본, 예를 들어, 배칼로바이러스 백본 영역을 포함한다. 일부 실시형태에서, 유전 요소 작제물은 박테리아 세포(예를 들어, 이. 콜라이 세포, 예를 들어, DH 10Bac 세포)에서의 유전 요소 작제물의 복제에 적합한 백본 영역을 포함한다. 일부 실시형태에서, 박미드는 배칼로바이러스 입자를 통해 숙주 세포 내로 도입된다. 실시형태들에서, 박미드는 생성자 세포, 예를 들어, 곤충 세포(예를 들어, Sf9 세포) 또는 박테리아 세포(예를 들어, 이. 콜라이 세포, 예를 들어, DH 10Bac 세포)에 의해 생성된다. 실시형태들에서, 생성자 세포는 예를 들어, 본원에 기재된 바와 같은 박미드 및/또는 공여자 벡터를 포함한다. 실시형태들에서, 생성자 세포는 박미드 및/또는 공여자 벡터의 복제를 위한 충분한 세포 기구를 추가로 포함한다.

[0253] **예를 들어, 아넬로백터의 어셈블리를 위한 ORF1 분자**

[0254] 아넬로백터는 예를 들어, 단백질성 외부 내에 유전 요소를 봉입함으로써 제조될 수 있다. 일부 실시형태에서, 봉입은 무세포 시스템 또는 세포에서 발생한다. 아넬로백터의 단백질성 외부는 일반적으로 아넬로바이러스 ORF1

핵산(예를 들어, 본원에 기재된 바와 같은 예를 들어, 아넬로바이러스 ORF1 분자 또는 그의 스플라이스 변이체 또는 기능적 단편)에 의해 인코딩된 폴리펩티드를 포함한다. ORF1 분자는, 일부 실시형태에서, 다음 중 하나 이상을 포함할 수 있다: 아르기닌 풍부 영역, 예를 들어, 적어도 60% 염기성 잔기(예를 들어, 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 또는 100% 염기성 잔기; 예를 들어, 60% 내지 90%, 60% 내지 80%, 70% 내지 90% 또는 70 내지 80% 염기성 잔기)를 갖는 영역을 포함하는 제1 영역, 및 젤리-롤 도메인, 예를 들어, 적어도 6개의 베타 가닥(예를 들어, 4, 5, 6, 7, 8, 9, 10, 11 또는 12개의 베타 가닥)을 포함하는 제2 영역. 실시형태들에서, 단백질성 외부는 하나 이상(예를 들어, 1, 2, 3, 4 또는 5개 모두)의 아넬로바이러스 ORF1 아르기닌-풍부 영역, 젤리-롤 영역, N22 도메인, 추가변 영역 및/또는 C-말단 도메인을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 ORF1 젤리-롤 영역(예를 들어, 본원에 기재된 바와 같음)을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 ORF1 아르기닌-풍부 영역(예를 들어, 본원에 기재된 바와 같음)을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 ORF1 N22 도메인(예를 들어, 본원에 기재된 바와 같음)을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 추가변 영역(예를 들어, 본원에 기재된 바와 같음)을 포함한다. 일부 실시형태에서, 단백질성 외부는 아넬로바이러스 ORF1 C-말단 도메인(예를 들어, 본원에 기재된 바와 같음)을 포함한다.

[0255] 일부 실시형태에서, 아넬로바이러스는 ORF1 분자 및/또는 ORF1 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질) 또는 그의 기능적 단편의 구조적 특징 및/또는 활성을 갖는 폴리펩티드를 포함한다. 일부 실시형태에서, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)에 비하여 절단(truncation)을 포함한다. 일부 실시형태에서, ORF1 분자는 아넬로바이러스 ORF1 단백질의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700개의 아미노산에 의해 절단된다. 일부 실시형태에서, ORF1 분자는 예를 들어, 본원에 기재된 바와 같은, 알파토르크바이러스, 베타토르크바이러스 또는 감마토르크바이러스 ORF1 단백질에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다. ORF1 분자는 일반적으로, 핵산 분자, 예컨대, DNA(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 유전 요소)에 결합할 수 있다. 일부 실시형태에서, ORF1 분자는 세포의 핵에 국소화된다. 특정 실시형태에서, ORF1 분자는 세포의 핵소체에 국소화된다.

[0256] 이론에 결부시키지 않고, ORF1 분자는 다른 ORF1 분자에 결합하여, 예를 들어, (예를 들어, 본원에 기재된 바와 같은) 단백질성 외부를 형성하는 능력이 있을 수 있다. 이러한 ORF1 분자는 캡시드를 형성하는 능력을 갖는 것으로서 설명될 수 있다. 일부 실시형태에서, 단백질성 외부는 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 봉입할 수 있다. 일부 실시형태에서, 복수의 ORF1 분자는 다합체를 형성하여, 예를 들어, 단백질성 외부를 생성할 수 있다. 일부 실시형태에서, 다합체는 동종다합체일 수 있다. 다른 실시형태에서, 다합체는 이종다합체일 수 있다.

[0257] **예를 들어, 아넬로바이러스의 어셈블리를 위한 ORF2 분자**

[0258] 본원에 기재된 조성물 또는 방법을 사용하여 아넬로바이러스를 생성하는 것은 아넬로바이러스 ORF2 분자(예를 들어, 본원에 기재된 바와 같음), 또는 그의 스플라이스 변이체 또는 기능적 단편의 발현을 수반할 수 있다. 일부 실시형태에서, 아넬로바이러스는 ORF2 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편, 및/또는 ORF2 분자를 인코딩하는 핵산, 또는 그의 스플라이스 변이체 또는 기능적 단편을 포함한다. 일부 실시형태에서, 아넬로바이러스는 ORF2 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편, 및/또는 ORF2 분자를 인코딩하는 핵산, 또는 그의 스플라이스 변이체 또는 기능적 단편을 포함하지 않는다. 일부 실시형태에서, 아넬로바이러스를 생성하는 것은 ORF2 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편의 발현을 포함하지만, ORF2 분자는 아넬로바이러스 내로 통합되지 않는다.

[0259] **유전 요소**

[0260] **RNA를 포함하는 유전 요소**

[0261] 일부 실시형태에서, 유전 요소는 핵산이거나 이를 포함한다. 일부 실시형태에서, 유전 요소는 단일-가닥 폴리뉴클레오티드이다. 일부 실시형태에서, 유전 요소는 하나 이상의 이중 가닥 영역을 포함한다. 일부 실시형태에서, 유전 요소는 RNA를 포함한다. 일부 실시형태에서, 유전 요소는 RNA 헤어핀 구조를 포함한다. 일부 실시형태에서, 유전 요소는 mRNA, 예를 들어, 화학적으로 변형된 mRNA이다. 일부 실시형태에서, 유전 요소는 적어도 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% RNA로 이루어진다. 일부 실시형태에서, 유전 요소는 DNA 가닥 및 RNA 가닥을 포함하며, 예를 들어, DNA 가닥의 적

어도 일부는 RNA 가닥의 적어도 일부에 혼성화된다.

[0262] 일부 실시형태에서, 유전 요소는 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 또는 ORF2t/3 중 임의의 것을 인코딩하지 않는다.

[0263] 일부 실시형태에서, RNA 유전 요소는 이펙터, 예를 들어, 이펙터 단백질을 인코딩한다.

[0264] 일부 실시형태에서, RNA 유전 요소는 이펙터, 예를 들어, 기능적 RNA이거나 이를 포함한다. 일부 실시형태에서, RNA는 mRNA, rRNA, tRNA(예를 들어, TREM), 조절 RNA, 비-코딩 RNA, 긴 비-코딩 RNA(lncRNA), 환형 RNA(circRNA), 이중 가닥 RNA(dsRNA), 가이드 RNA(gRNA), 작은 간섭 RNA(siRNA), 짧은 헤어핀 RNA(shRNA), 피위-상호작용 RNA(piRNA), 작은 핵소체 RNA(snoRNA), 작은 핵 RNA(snRNA), 세포외 RNA(exRNA), 작은 카잘체-특이적 RNA(scaRNA), 마이크로RNA(miRNA) 및 기타 RNAi 분자로 이루어진 군으로부터 선택된다.

[0265] 일부 실시형태에서, 유전 요소는 RNA, 예를 들어, 화학적으로 변형된 RNA를 포함한다. 일부 실시형태에서, 유전 요소의 RNA의 하나 이상의 뉴클레오티드는 화학적으로 변형된다. 일부 실시형태에서, RNA는 하나 이상의 염기에 대하여 하나 이상의 화학적 변형을 포함한다. 일부 실시형태에서, RNA는 하나 이상의 당에 대하여 하나 이상의 화학적 변형을 포함한다. 일부 실시형태에서, 유전 요소의 RNA의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오티드는 화학적으로 변형된다. 일부 실시형태에서, RNA는 하나 이상의 백본 변형을 포함한다. 일부 실시형태에서, 변형은 비-자연 발생 변형, 예를 들어, 표 5 내지 9 중 어느 하나에 기재된 변형을 포함한다. 비-자연 발생 변형은 당염기에 알려진 방법에 따라 이루어질 수 있다.

[0266] 일부 실시형태에서, 본원에 기재된 유전 요소는 표 5에 제공된 비-자연 발생 변형 또는 이들의 조합을 포함한다.

[0267] 표 5: 예시적인 비-자연 발생 변형

명칭	기호	염기	자연 발생
7-데아자-아데노신	--	A	아니오
N1-메틸-아데노신	--	A	아니오
N6, N6(디메틸)아데닌	--	A	아니오
N6-시스-하이드록시-이소펜테닐-아데노신	--	A	아니오
a-티오-아데노신	--	A	아니오
2(아미노)아데닌	--	A	아니오
2(아미노프로필)아데닌	--	A	아니오
2(메틸티오)N6(이소펜테닐)아데닌	--	A	아니오
2-(알킬)아데닌	--	A	아니오
2-(아미노알킬)아데닌	--	A	아니오
2-(아미노프로필)아데닌	--	A	아니오
2-(할로)아데닌	--	A	아니오
2-(할로)아데닌	--	A	아니오
2-(프로필)아데닌	--	A	아니오
2'-아미노-2'-데옥시-ATP	--	A	아니오
2'-아지도-2'-데옥시-ATP	--	A	아니오
2'-데옥시-2'-a-아미노아데노신 TP	--	A	아니오
2'-데옥시-2'-a-아지도아데노신 TP	--	A	아니오
6(알킬)아데닌	--	A	아니오
6(메틸)아데닌	--	A	아니오

[0268]

6-(알킬)아데닌	--	A	아니오
6-(메틸)아데닌	--	A	아니오
7(데아자)아데닌	--	A	아니오
8(알케닐)아데닌	--	A	아니오
8(알킬닐)아데닌	--	A	아니오
8(아미노)아데닌	--	A	아니오
8(티오알킬)아데닌	--	A	아니오
8-(알케닐)아데닌	--	A	아니오
8-(알킬)아데닌	--	A	아니오
8-(알킬닐)아데닌	--	A	아니오
8-(아미노)아데닌	--	A	아니오
8-(할로)아데닌	--	A	아니오
8-(하이드록실)아데닌	--	A	아니오
8-(티오알킬)아데닌	--	A	아니오
8-(티올)아데닌	--	A	아니오
8-아지도-아데노신	--	A	아니오
아자 아데닌	--	A	아니오
데아자 아데닌	--	A	아니오
N6 (메틸)아데닌	--	A	아니오
N6-(이소펜틸)아데닌	--	A	아니오
7-데아자-8-아자-아데노신	--	A	아니오
7-메틸아데닌	--	A	아니오
1-데아자아데노신 TP	--	A	아니오
2'플루오로-N6-Bz-데옥시아데노신 TP	--	A	아니오
2'-OMe-2-아미노-ATP	--	A	아니오
2'O-메틸-N6-Bz-데옥시아데노신 TP	--	A	아니오
2'-a-에틸닐아데노신 TP	--	A	아니오
2-아미노아데닌	--	A	아니오
2-아미노아데노신 TP	--	A	아니오
2-아미노-ATP	--	A	아니오
2'-a-트리플루오로메틸아데노신 TP	--	A	아니오
2-아지도아데노신 TP	--	A	아니오

[0269]

2' -b-에티닐아데노신 TP	--	A	아니오
2-브로모아데노신 TP	--	A	아니오
2'-b-트리플루오로메틸아데노신 TP	--	A	아니오
2-클로로아데노신 TP	--	A	아니오
2' -데옥시-2',2'-디플루오로아데노신 TP	--	A	아니오
2' -데옥시-2'-a-머캅토아데노신 TP	--	A	아니오
2' -데옥시-2'-a-티오메톡시아데노신 TP	--	A	아니오
2' -데옥시-2'-b-아미노아데노신 TP	--	A	아니오
2' -데옥시-2'-b-아지도아데노신 TP	--	A	아니오
2' -데옥시-2'-b-브로모아데노신 TP	--	A	아니오
2' -데옥시-2'-b-클로로아데노신 TP	--	A	아니오
2' -데옥시-2'-b-플루오로아데노신 TP	--	A	아니오
2' -데옥시-2'-b-아이오도아데노신 TP	--	A	아니오
2' -데옥시-2'-b-머캅토아데노신 TP	--	A	아니오
2' -데옥시-2'-b-티오메톡시아데노신 TP	--	A	아니오
2-플루오로아데노신 TP	--	A	아니오
2-아이오도아데노신 TP	--	A	아니오
2-머캅토아데노신 TP	--	A	아니오
2-메톡시-아데닌	--	A	아니오
2-메틸티오-아데닌	--	A	아니오
2-트리플루오로메틸아데노신 TP	--	A	아니오
3-데아자 -3-브로모아데노신 TP	--	A	아니오
3-데아자 -3-클로로아데노신 TP	--	A	아니오
3-데아자 -3-플루오로아데노신 TP	--	A	아니오
3-데아자 -3-아이오도아데노신 TP	--	A	아니오
3-데아자아데노신 TP	--	A	아니오
4' -아지도아데노신 TP	--	A	아니오
4'-카보사이클릭 아데노신 TP	--	A	아니오
4' -에티닐아데노신 TP	--	A	아니오
5'-호모-아데노신 TP	--	A	아니오
8- 아자-ATP	--	A	아니오
8-브로모-아데노신 TP	--	A	아니오

[0270]

8-트리플루오로메틸아데노신 TP	--	A	아니오
9-데아자아데노신 TP	--	A	아니오
2-아미노퓨린	--	A/G	아니오
7-데아자-2,6-디아미노퓨린	--	A/G	아니오
7-데아자-8-아자-2,6-디아미노퓨린	--	A/G	아니오
7-데아자-8-아자-2-아미노퓨린	--	A/G	아니오
2,6-디아미노퓨린	--	A/G	아니오
7-데아자-8-아자-아데닌, 7-데아자-2-아미노퓨린	--	A/G	아니오
4-메틸시티딘	--	C	아니오
5-아자-시티딘	--	C	아니오
슈도-이소-시티딘	--	C	아니오
피플로-시티딘	--	C	아니오
a-티오-시티딘	--	C	아니오
2-(티오)시토신	--	C	아니오
2'-아미노-2'-데옥시-CTP	--	C	아니오
2'-아지도-2'-데옥시-CTP	--	C	아니오
2'-데옥시-2'-a-아미노시티딘 TP	--	C	아니오
2'-데옥시-2'-a-아지도시티딘 TP	--	C	아니오
3 (데아자) 5 (아자)시토신	--	C	아니오
3 (메틸)시토신	--	C	아니오
3-(알킬)시토신	--	C	아니오
3-(데아자) 5 (아자)시토신	--	C	아니오
3-(메틸)시티딘	--	C	아니오
4,2'-O-디메틸시티딘	--	C	아니오
5(할로)시토신	--	C	아니오
5(메틸)시토신	--	C	아니오
5(프로피닐)시토신	--	C	아니오
5(트리플루오로메틸)시토신	--	C	아니오
5-(알킬)시토신	--	C	아니오
5-(알킬니!)시토신	--	C	아니오
5-(할로)시토신	--	C	아니오
5-(프로피닐)시토신	--	C	아니오

[0271]

5-(트리플루오로메틸)시토신	--	C	아니오
5-브로모-시티딘	--	C	아니오
5-아이오도-시티딘	--	C	아니오
5-프로피닐 시토신	--	C	아니오
6-(아조)시토신	--	C	아니오
6-아자-시티딘	--	C	아니오
아자 시토신	--	C	아니오
데아자 시토신	--	C	아니오
N4(아세틸)시토신	--	C	아니오
1-메틸-1-데아자-슈도이소시티딘	--	C	아니오
1-메틸-슈도이소시티딘	--	C	아니오
2-메톡시-5-메틸-시티딘	--	C	아니오
2-메톡시-시티딘	--	C	아니오
2-티오-5-메틸-시티딘	--	C	아니오
4-메톡시-1-메틸-슈도이소시티딘	--	C	아니오
4-메톡시-슈도이소시티딘	--	C	아니오
4-티오-1-메틸-1-데아자-슈도이소시티딘	--	C	아니오
4-티오-1-메틸-슈도이소시티딘	--	C	아니오
4-티오-슈도이소시티딘	--	C	아니오
5-아자-제볼라린	--	C	아니오
5-메틸-제볼라린	--	C	아니오
피롤로-슈도이소시티딘	--	C	아니오
제볼라린	--	C	아니오
(E)-5-(2-브로모-비닐)시티딘 TP	--	C	아니오
2,2'-안하이드로-시티딘 TP 하이드로클로라이드	--	C	아니오
2'플루오르-N4-Bz-시티딘 TP	--	C	아니오
2'플루오르-N4-아세틸-시티딘 TP	--	C	아니오
2'-O-메틸-N4-아세틸-시티딘 TP	--	C	아니오
2'O-메틸-N4-Bz-시티딘 TP	--	C	아니오
2'-a-에틸닐시티딘 TP	--	C	아니오
2'-a-트리플루오로메틸시티딘 TP	--	C	아니오
2'-b-에틸닐시티딘 TP	--	C	아니오

[0272]

2'-b-트리플루오로메틸시티딘 TP	--	C	아니오
2'-데옥시-2',2'-디플루오로시티딘 TP	--	C	아니오
2'-데옥시-2'-a-머캅토시티딘 TP	--	C	아니오
2'-데옥시-2'-a-티오메톡시티딘 TP	--	C	아니오
2'-데옥시-2'-b-아미노시티딘 TP	--	C	아니오
2'-데옥시-2'-b-아지도시티딘 TP	--	C	아니오
2'-데옥시-2'-b-브로모시티딘 TP	--	C	아니오
2'-데옥시-2'-b-클로로시티딘 TP	--	C	아니오
2'-데옥시-2'-b-플루오로시티딘 TP	--	C	아니오
2'-데옥시-2'-b-아이오도시티딘 TP	--	C	아니오
2'-데옥시-2'-b-머캅토시티딘 TP	--	C	아니오
2'-데옥시-2'-b-티오메톡시티딘 TP	--	C	아니오
2'-O-메틸-5-(1-프로피닐)시티딘 TP	--	C	아니오
3'-에티닐시티딘 TP	--	C	아니오
4'-아지도시티딘 TP	--	C	아니오
4'-카보사이클릭 시티딘 TP	--	C	아니오
4'-에티닐시티딘 TP	--	C	아니오
5-(1-프로피닐)아라-시티딘 TP	--	C	아니오
5-(2-클로로-페닐)-2-티오시티딘 TP	--	C	아니오
5-(4-아미노-페닐)-2-티오시티딘 TP	--	C	아니오
5-아미노알릴-CTP	--	C	아니오
5-시아노시티딘 TP	--	C	아니오
5-에티닐아라-시티딘 TP	--	C	아니오
5-에티닐시티딘 TP	--	C	아니오
5'-호모-시티딘 TP	--	C	아니오
5-메톡시티딘 TP	--	C	아니오
5-트리플루오로메틸-시티딘 TP	--	C	아니오
N4-아미노-시티딘 TP	--	C	아니오
N4-벤조일-시티딘 TP	--	C	아니오
슈도이소시티딘	--	C	아니오
6-티오-구아노신	--	G	아니오
7-테아자-구아노신	--	G	아니오

[0273]

8-옥소-구아노신	--	G	아니오
N1-메틸-구아노신	--	G	아니오
a-티오-구아노신	--	G	아니오
2 (프로필)구아닌	--	G	아니오
2-(알킬)구아닌	--	G	아니오
2'-아미노-2'-데옥시-GTP	--	G	아니오
2'-아지도-2'-데옥시-GTP	--	G	아니오
2'-데옥시-2'-a-아미노구아노신 TP	--	G	아니오
2'-데옥시-2'-a-아지도구아노신 TP	--	G	아니오
6(메틸)구아닌	--	G	아니오
6-(알킬)구아닌	--	G	아니오
6-(메틸)구아닌	--	G	아니오
6-메틸-구아노신	--	G	아니오
7(알킬)구아닌	--	G	아니오
7(테아자)구아닌	--	G	아니오
7(메틸)구아닌	--	G	아니오
7-(알킬)구아닌	--	G	아니오
7-(테아자)구아닌	--	G	아니오
7-(메틸)구아닌	--	G	아니오
8(알킬)구아닌	--	G	아니오
8(알킬닐)구아닌	--	G	아니오
8(할로)구아닌	--	G	아니오
8(티오알킬)구아닌	--	G	아니오
8-(알케닐)구아닌	--	G	아니오
8-(알킬)구아닌	--	G	아니오
8-(알킬닐)구아닌	--	G	아니오
8-(아미노)구아닌	--	G	아니오
8-(할로)구아닌	--	G	아니오
8-(하이드록실)구아닌	--	G	아니오
8-(티오알킬)구아닌	--	G	아니오
8-(티올)구아닌	--	G	아니오
야자구아닌	--	G	아니오

[0274]

데아자 구아닌	--	G	아니오
N(메틸)구아닌	--	G	아니오
N-(메틸)구아닌	--	G	아니오
1-메틸-6-티오-구아노신	--	G	아니오
6-메톡시-구아노신	--	G	아니오
6-티오 -7-데아자-8-아자-구아노신	--	G	아니오
6-티오 -7-데아자-구아노신	--	G	아니오
6-티오 -7-메틸-구아노신	--	G	아니오
7-데아자-8-아자-구아노신	--	G	아니오
7-메틸-8-옥소-구아노신	--	G	아니오
N2,N2-디메틸-6-티오-구아노신	--	G	아니오
N2-메틸-6-티오-구아노신	--	G	아니오
1-Me-GTP	--	G	아니오
2'플루오로-N2-이소부틸-구아노신 TP	--	G	아니오
2'O-메틸-N2-이소부틸-구아노신 TP	--	G	아니오
2'-a-에티닐구아노신 TP	--	G	아니오
2'-a-트리플루오로메틸구아노신 TP	--	G	아니오
2'-b-에티닐구아노신 TP	--	G	아니오
2'-b-트리플루오로메틸구아노신 TP	--	G	아니오
2'-데옥시-2',2'-디플루오로구아노신 TP	--	G	아니오
2'-데옥시-2'-a-머캅토구아노신 TP	--	G	아니오
2'-데옥시-2'-a-티오메톡시구아노신 TP	--	G	아니오
2'-데옥시-2'-b-아미노구아노신 TP	--	G	아니오
2'-데옥시-2'-b-아지도구아노신 TP	--	G	아니오
2'-데옥시-2'-b-브로모구아노신 TP	--	G	아니오
2'-데옥시-2'-b-클로로구아노신 TP	--	G	아니오
2'-데옥시-2'-b-플루오로구아노신 TP	--	G	아니오
2'-데옥시-2'-b-아이오도구아노신 TP	--	G	아니오
2'-데옥시-2'-b-머캅토구아노신 TP	--	G	아니오
2'-데옥시-2'-b-티오메톡시구아노신 TP	--	G	아니오
4'-아지도구아노신 TP	--	G	아니오
4'-카보사이클릭 구아노신 TP	--	G	아니오

[0275]

4' - 에틸닐구아노신 TP	--	G	아니오
5'-호모-구아노신 TP	--	G	아니오
8-브로모-구아노신 TP	--	G	아니오
9-데아자구아노신 TP	--	G	아니오
N2-이소부틸-구아노신 TP	--	G	아니오
7-메틸이노신		A	아니오
알리아미노- 티미딘	--	T	아니오
아자 티미딘	--	T	아니오
데아자 티미딘	--	T	아니오
데옥시-티미딘	--	T	아니오
5-프로피닐 우라실	--	U	아니오
a-티오-우리딘	--	U	아니오
1 (아미노알킬아미노-카보닐에틸레닐)- 2(티오)-슈도우라실	--	U	아니오
1 (아미노알킬아미노카보닐에틸레닐)- 2,4-(디티오)슈도우라실	--	U	아니오
1 (아미노알킬아미노카보닐에틸레닐)-4 (티오)슈도우라실	--	U	아니오
1 (아미노알킬아미노카보닐에틸레닐)-슈도우라실	--	U	아니오
1 (아미노카보닐에틸레닐)-2(티오)슈도우라실	--	U	아니오
1 (아미노카보닐에틸레닐)-2,4- (디티오)슈도우라실	--	U	아니오
1 (아미노카보닐에틸레닐)-4(티오)슈도우라실	--	U	아니오
1 (아미노카보닐에틸레닐)-슈도우라실	--	U	아니오
1 치환된 2(티오)-슈도우라실	--	U	아니오
1 치환된 2,4-(디티오)슈도우라실	--	U	아니오
1 치환된 4 (티오)슈도우라실	--	U	아니오
1 치환된 슈도우라실	--	U	아니오
1-(아미노알킬아미노-카보닐에틸레닐)- 2- (티오)슈도우라실	--	U	아니오
1-메틸-3-(3-아미노-3-카복시프로필)슈도우리딘 TP	--	U	아니오
1-메틸-3-(3-아미노-3-카복시프루롤)슈도 -UTP	--	U	아니오
1-메틸-슈도 -UTP	--	U	아니오

[0276]

2 (티오)슈도우라실	--	U	아니오
2' -데옥시 우리딘	--	U	아니오
2' -플루오로우리딘	--	U	아니오
2-(티오)우라실	--	U	아니오
2,4-(디티오)슈도우라실	--	U	아니오
2' -메틸, 2' -아미노, 2' -아지도, 2' -플루로-구아노신	--	U	아니오
2' -아미노-2' -데옥시-UTP	--	U	아니오
2' -아지도-2' -데옥시-UTP	--	U	아니오
2' -아지도-데옥시우리딘 TP	--	U	아니오
2'-O-메틸슈도우리딘	--	U	아니오
2' -데옥시 우리딘	2'dU	U	아니오
2' -플루오로우리딘	--	U	아니오
2' -데옥시-2'-a-아미노우리딘 TP	--	U	아니오
2' -데옥시-2'-a-아지도우리딘 TP	--	U	아니오
2-메틸슈도우리딘	m3'P	U	아니오
3 (3 아미노-3 카복시프로필)우라실	--	U	아니오
4 (티오)슈도우라실	--	U	아니오
4-(티오)슈도우라실	--	U	아니오
4-(티오)우라실	--	U	아니오
4-티오우라실	--	U	아니오
5 (1,3-디아졸-1-알킬)우라실	--	U	아니오
5 (2-아미노프로필)우라실	--	U	아니오
5 (아미노알킬)우라실	--	U	아니오
5 (디메틸아미노알킬)우라실	--	U	아니오
5 (구아니디늄알킬)우라실	--	U	아니오
5 (메톡시카보닐메틸)-2-(티오)우라실	--	U	아니오
5 (메톡시카보닐-메틸)우라실	--	U	아니오
5 (메틸) 2 (티오)우라실	--	U	아니오
5 (메틸) 2,4 (디티오)우라실	--	U	아니오
5 (메틸) 4 (티오)우라실	--	U	아니오
5 (메틸아미노메틸)-2 (티오)우라실	--	U	아니오
5 (메틸아미노메틸)-2,4 (디티오)우라실	--	U	아니오

[0277]

5 (메틸아미노메틸)-4 (티오)우라실	--	U	아니오
5 (프로피닐)우라실	--	U	아니오
5 (트리플루오로메틸)우라실	--	U	아니오
5-(2-아미노프로필)우라실	--	U	아니오
5-(알킬)-2-(티오)슈도우라실	--	U	아니오
5-(알킬)-2,4 (디티오)슈도우라실	--	U	아니오
5-(알킬)-4 (티오)슈도우라실	--	U	아니오
5-(알킬)슈도우라실	--	U	아니오
5-(알킬)우라실	--	U	아니오
5-(알키닐)우라실	--	U	아니오
5-(알릴아미노)우라실	--	U	아니오
5-(시아노알킬)우라실	--	U	아니오
5-(디아킬아미노알킬)	--	U	아니오
5-(디메틸아미노알킬)우라실	--	U	아니오
5-(구아니디늄알킬)우라실	--	U	아니오
5-(할로)우라실	--	U	아니오
5-(1,3-디아졸-1-알킬)우라실	--	U	아니오
5-(메톡시)우라실	--	U	아니오
5-(메톡시카보닐메틸)-2-(티오)우라실	--	U	아니오
5-(메톡시카보닐-메틸)우라실	--	U	아니오
5-(메틸) 2(티오)우라실	--	U	아니오
5-(메틸) 2,4 (디티오)우라실	--	U	아니오
5-(메틸) 4 (티오)우라실	--	U	아니오
5-(메틸)-2-(티오)슈도우라실	--	U	아니오
5-(메틸)-2,4 (디티오)슈도우라실	--	U	아니오
5-(메틸)-4 (티오)슈도우라실	--	U	아니오
5-(메틸)슈도우라실	--	U	아니오
5-(메틸아미노메틸)-2 (티오)슈도우라실	--	U	아니오
5-(메틸아미노메틸)-2,4(디티오)슈도우라실	--	U	아니오
5-(메틸아미노메틸)-4-(티오)슈도우라실	--	U	아니오
5-(프로피닐)우라실	--	U	아니오
5-(트리플루오로메틸)우라실	--	U	아니오

[0278]

5-아미노알릴-우리딘	--	U	아니오
5-브로모-우리딘	--	U	아니오
5-아이오도-우리딘	--	U	아니오
5-우라실	--	U	아니오
6 (아조)우라실	--	U	아니오
6-(아조)우라실	--	U	아니오
6-아자-우리딘	--	U	아니오
알리아미노-우라실	--	U	아니오
아자 우라실	--	U	아니오
데아자 우라실	--	U	아니오
N3 (메틸)우라실	--	U	아니오
슈도-UTP-1-2-에탄산	--	U	아니오
슈도우라실	--	U	아니오
4-티오-슈도-UTP	--	U	아니오
1-카복시메틸-슈도우리딘	--	U	아니오
1-메틸-1-데아자-슈도우리딘	--	U	아니오
1-프로피닐-우리딘	--	U	아니오
1-타우리노메틸-1-메틸-우리딘	--	U	아니오
1-타우리노메틸-4-티오-우리딘	--	U	아니오
1-타우리노메틸-슈도우리딘	--	U	아니오
2-메톡시-4-티오-슈도우리딘	--	U	아니오
2-티오-1-메틸-1-데아자-슈도우리딘	--	U	아니오
2-티오-1-메틸-슈도우리딘	--	U	아니오
2-티오-5-아자-우리딘	--	U	아니오
2-티오-디하이드로슈도우리딘	--	U	아니오
2-티오-디하이드로우리딘	--	U	아니오
2-티오-슈도우리딘	--	U	아니오
4-메톡시-2-티오-슈도우리딘	--	U	아니오
4-메톡시 슈도우리딘	--	U	아니오
4-티오-1-메틸-슈도우리딘	--	U	아니오
4-티오-슈도우리딘	--	U	아니오
5-아자-우리딘	--	U	아니오

[0279]

디하이드로슈도우리딘	--	U	아니오
(±)1-(2-하이드록시프로필)슈도우리딘 TP	--	U	아니오
(2R)-1-(2-하이드록시프로필)슈도우리딘 TP	--	U	아니오
(2S)-1-(2-하이드록시프로필)슈도우리딘 TP	--	U	아니오
(E)-5-(2-브로모-비닐)아라-우리딘 TP	--	U	아니오
(E)-5-(2-브로모-비닐)우리딘 TP	--	U	아니오
(Z)-5-(2-브로모-비닐)아라-우리딘 TP	--	U	아니오
(Z)-5-(2-브로모-비닐)우리딘 TP	--	U	아니오
1-(2,2,2-트리플루오로에틸)-슈도-UTP	--	U	아니오
1-(2,2,3,3,3-펜타플루오로프로필)슈도우리딘 TP	--	U	아니오
1-(2,2-디에톡시메틸)슈도우리딘 TP	--	U	아니오
1-(2,4,6-트리메틸벤질)슈도우리딘 TP	--	U	아니오
1-(2,4,6-트리메틸-벤질)슈도-UTP	--	U	아니오
1-(2,4,6-트리메틸-페닐)슈도-UTP	--	U	아니오
1-(2-아미노-2-카복시에틸)슈도-UTP	--	U	아니오
1-(2-아미노-에틸)슈도-UTP	--	U	아니오
1-(2-하이드록시에틸)슈도우리딘 TP	--	U	아니오
1-(2-메톡시에틸)슈도우리딘 TP	--	U	아니오
1-(3,4-비스-트리플루오로메톡시벤질)슈도우리딘 TP	--	U	아니오
1-(3,4-디메톡시벤질)슈도우리딘 TP	--	U	아니오
1-(3-아미노-3-카복시프로필)슈도-UTP	--	U	아니오
1-(3-아미노-프로필)슈도-UTP	--	U	아니오
1-(3-사이클로프로필-프로프-2-이닐)슈도우리딘 TP	--	U	아니오
1-(4-아미노-4-카복시부틸)슈도-UTP	--	U	아니오
1-(4-아미노-벤질)슈도-UTP	--	U	아니오
1-(4-아미노-부틸)슈도-UTP	--	U	아니오
1-(4-아미노-페닐)슈도-UTP	--	U	아니오
1-(4-아지도벤질)슈도우리딘 TP	--	U	아니오
1-(4-브로모벤질)슈도우리딘 TP	--	U	아니오

[0280]

1-(4-클로로벤질)슈도우리딘 TP	--	U	아니오
1-(4-플루오로벤질)슈도우리딘 TP	--	U	아니오
1-(4-아이오도벤질)슈도우리딘 TP	--	U	아니오
1-(4-메탄설폰닐벤질)슈도우리딘 TP	--	U	아니오
1-(4-메톡시벤질)슈도우리딘 TP	--	U	아니오
1-(4-메톡시-벤질)슈도-UTP	--	U	아니오
1-(4-메톡시-페닐)슈도-UTP	--	U	아니오
1-(4-메틸벤질)슈도우리딘 TP	--	U	아니오
1-(4-메틸-벤질)슈도-UTP	--	U	아니오
1-(4-니트로벤질)슈도우리딘 TP	--	U	아니오
1-(4-니트로-벤질)슈도-UTP	--	U	아니오
1-(4-니트로-페닐)슈도-UTP	--	U	아니오
1-(4-티오메톡시벤질)슈도우리딘 TP	--	U	아니오
1-(4-트리플루오로메톡시벤질)슈도우리딘 TP	--	U	아니오
1-(4-트리플루오로메틸벤질)슈도우리딘 TP	--	U	아니오
1-(5-아미노-펜틸)슈도-UTP	--	U	아니오
1-(6-아미노-헥실)슈도-UTP	--	U	아니오
1,6-디메틸-슈도-UTP	--	U	아니오
1-[3-(2-{2-[2-(2-아미노에톡시)-에톡시]-에톡시}-프로피오닐)슈도우리딘 TP	--	U	아니오
1-{3-[2-(2-아미노에톡시)-에톡시]-프로피오닐}슈도우리딘 TP	--	U	아니오
1-아세틸슈도우리딘 TP	--	U	아니오
1-알킬-6-(1-프로피닐)-슈도-UTP	--	U	아니오
1-알킬-6-(2-프로피닐)-슈도-UTP	--	U	아니오
1-알킬-6-알릴-슈도-UTP	--	U	아니오
1-알킬-6-에틸닐-슈도-UTP	--	U	아니오
1-알킬-6-호모알릴-슈도-UTP	--	U	아니오
1-알킬-6-비닐-슈도-UTP	--	U	아니오
1-알릴슈도우리딘 TP	--	U	아니오
1-아미노메틸-슈도-UTP	--	U	아니오

[0281]

1-벤조일슈도우리딘 TP	--	U	아니오
1-벤질옥시메틸슈도우리딘 TP	--	U	아니오
1-벤질-슈도-UTP	--	U	아니오
1-비오틴-PEG2-슈도우리딘 TP	--	U	아니오
1-비오틴슈도우리딘 TP	--	U	아니오
1-부틸-슈도-UTP	--	U	아니오
1-시아노메틸슈도우리딘 TP	--	U	아니오
1-사이클로부틸메틸-슈도-UTP	--	U	아니오
1-사이클로부틸-슈도-UTP	--	U	아니오
1-사이클로헥실메틸-슈도-UTP	--	U	아니오
1-사이클로헥실-슈도-UTP	--	U	아니오
1-사이클로헥실메틸-슈도-UTP	--	U	아니오
1-사이클로헥실-슈도-UTP	--	U	아니오
1-사이클로옥틸메틸-슈도-UTP	--	U	아니오
1-사이클로옥틸-슈도-UTP	--	U	아니오
1-사이클로펜틸메틸-슈도-UTP	--	U	아니오
1-사이클로펜틸-슈도-UTP	--	U	아니오
1-사이클로프로필메틸-슈도-UTP	--	U	아니오
1-사이클로프로필-슈도-UTP	--	U	아니오
1-에틸-슈도-UTP	--	U	아니오
1-헥실-슈도-UTP	--	U	아니오
1-호모알릴슈도우리딘 TP	--	U	아니오
1-하이드록시메틸슈도우리딘 TP	--	U	아니오
1-이소-프로필-슈도-UTP	--	U	아니오
1-Me-2-티오-슈도-UTP	--	U	아니오
1-Me-4-티오-슈도-UTP	--	U	아니오
1-Me-알파-티오-슈도-UTP	--	U	아니오
1-메탄설폰닐메틸슈도우리딘 TP	--	U	아니오
1-메톡시메틸슈도우리딘 TP	--	U	아니오
1-메틸-6-(2,2,2-트리플루오로에틸)슈도-UTP	--	U	아니오
1-메틸-6-(4-모르폴리노)-슈도-DTP	--	U	아니오

[0282]

1-메틸-6-(4-티오모르폴리노)-슈도-UTP	--	U	아니오
1-메틸-6-(치환된 페닐)슈도-UTP	--	U	아니오
1-메틸-6-아미노-슈도-UTP	--	U	아니오
1-메틸-6-아지도-슈도-UTP	--	U	아니오
1-메틸-6-브로모-슈도-UTP	--	U	아니오
1-메틸-6-부틸-슈도-UTP	--	U	아니오
1-메틸-6-클로로-슈도-UTP	--	U	아니오
1-메틸-6-시아노-슈도-UTP	--	U	아니오
1-메틸-6-디메틸아미노-슈도-UTP	--	U	아니오
1-메틸-6-에톡시-슈도-UTP	--	U	아니오
1-메틸-6-에틸카복실레이트-슈도-UTP	--	U	아니오
1-메틸-6-에틸-슈도-UTP	--	U	아니오
1-메틸-6-플루오로-슈도-UTP	--	U	아니오
1-메틸-6-포르밀-슈도-UTP	--	U	아니오
1-메틸-6-하이드록시아미노-슈도-UTP	--	U	아니오
1-메틸-6-하이드록시-슈도-UTP	--	U	아니오
1-메틸-6-아이오도-슈도-UTP	--	U	아니오
1-메틸-6-이소-프로필-슈도-UTP	--	U	아니오
1-메틸-6-메톡시-슈도-UTP	--	U	아니오
1-메틸-6-메틸아미노-슈도-UTP	--	U	아니오
1-메틸-6-페닐-슈도-UTP	--	U	아니오
1-메틸-6-프로필-슈도-UTP	--	U	아니오
1-메틸-6-tert-부틸-슈도-UTP	--	U	아니오
1-메틸-6-트리플루오로메톡시-슈도-UTP	--	U	아니오
1-메틸-6-트리플루오로메틸-슈도-UTP	--	U	아니오
1-모르폴리노메틸슈도우리딘 TP	--	U	아니오
1-펜틸-슈도-UTP	--	U	아니오
1-페닐-슈도-UTP	--	U	아니오
1-피알로일슈도우리딘 TP	--	U	아니오
1-프로파르길슈도우리딘 TP	--	U	아니오
1-프로필-슈도-UTP	--	U	아니오

[0283]

1-프로피닐-슈도우리딘	--	U	아니오
1-p-톨릴-슈도-UTP	--	U	아니오
1-tert-부틸-슈도-UTP	--	U	아니오
1-티오메톡시메틸슈도우리딘 TP	--	U	아니오
1-티오모르폴리노메틸슈도우리딘 TP	--	U	아니오
1-트리플루오로아세틸슈도우리딘 TP	--	U	아니오
1-트리플루오로메틸-슈도-UTP	--	U	아니오
1-비닐슈도우리딘 TP	--	U	아니오
2,2'-안하이드로-우리딘 TP	--	U	아니오
2' -브로모-데옥시우리딘 TP	--	U	아니오
2' -F-5-메틸-2'-데옥시-UTP	--	U	아니오
2' -OMe-5-Me-UTP	--	U	아니오
2' -OMe-슈도-UTP	--	U	아니오
2' -a-에틸우리딘 TP	--	U	아니오
2' -a-트리플루오로메틸우리딘 TP	--	U	아니오
2' -b-에틸우리딘 TP	--	U	아니오
2' -b-트리플루오로메틸우리딘 TP	--	U	아니오
2' -데옥시-2',2'-디플루오로우리딘 TP	--	U	아니오
2' -데옥시-2'-a-머캅토우리딘 TP	--	U	아니오
2' -데옥시-2'-a-티오메톡시우리딘 TP	--	U	아니오
2' -데옥시-2'-b-아미노우리딘 TP	--	U	아니오
2' -데옥시-2'-b-아지도우리딘 TP	--	U	아니오
2' -데옥시-2'-b-브로모우리딘 TP	--	U	아니오
2' -데옥시-2'-b-클로로우리딘 TP	--	U	아니오
2' -데옥시-2'-b-플루오로우리딘 TP	--	U	아니오
2' -데옥시-2'-b-아이오도우리딘 TP	--	U	아니오
2' -데옥시-2'-b-머캅토우리딘 TP	--	U	아니오
2' -데옥시-2'-b-티오메톡시우리딘 TP	--	U	아니오
2-메톡시-4-티오-우리딘	--	U	아니오
2-메톡시우리딘	--	U	아니오
2' -O-메틸-5-(1-프로피닐)우리딘 TP	--	U	아니오
3-알킬-슈도-UTP	--	U	아니오

[0284]

4'-아지도우리딘 TP	--	U	아니오
4'-카보사이클릭 우리딘 TP	--	U	아니오
4'-에티닐우리딘 TP	--	U	아니오
5-(1-프로피닐)아라-우리딘 TP	--	U	아니오
5-(2-푸라닐)우리딘 TP	--	U	아니오
5-시아노우리딘 TP	--	U	아니오
5-디메틸아미노우리딘 TP	--	U	아니오
5'-호모-우리딘 TP	--	U	아니오
5-아이오도-2'-플루오로-데옥시우리딘 TP	--	U	아니오
5-페닐에티닐우리딘 TP	--	U	아니오
5-트리뷰테로메틸-6-뷰테로우리딘 TP	--	U	아니오
5-트리플루오로메틸-우리딘 TP	--	U	아니오
5-비닐아라우리딘 TP	--	U	아니오
6-(2,2,2-트리플루오로에틸)-슈도-UTP	--	U	아니오
6-(4-모르폴리노)-슈도-DTP	--	U	아니오
6-(4-티오모르폴리노)-슈도-UTP	--	U	아니오
6-(치환된-페닐)-슈도-UTP	--	U	아니오
6-아미노-슈도-UTP	--	U	아니오
6-아지도-슈도-UTP	--	U	아니오
6-브로모-슈도-UTP	--	U	아니오
6-부틸-슈도-UTP	--	U	아니오
6-클로로-슈도-UTP	--	U	아니오
6-시아노-슈도-UTP	--	U	아니오
6-디메틸아미노-슈도-UTP	--	U	아니오
6-에톡시-슈도-UTP	--	U	아니오
6-에틸카복실레이트-슈도-UTP	--	U	아니오
6-에틸-슈도-UTP	--	U	아니오
6-플루오로-슈도-UTP	--	U	아니오
6-포르밀-슈도-UTP	--	U	아니오
6-하이드록시아미노-슈도-UTP	--	U	아니오
6-하이드록시-슈도-UTP	--	U	아니오
6-아이오도-슈도-UTP	--	U	아니오

[0285]

6-이소-프로필-슈도-UTP	--	U	아니오
6-메톡시-슈도-UTP	--	U	아니오
6-메틸아미노-슈도-UTP	--	U	아니오
6-메틸-슈도-UTP	--	U	아니오
6-페닐-슈도-UTP	--	U	아니오
6-페닐-슈도-UTP	--	U	아니오
6-프로필-슈도-UTP	--	U	아니오
6-tert-부틸-슈도-UTP	--	U	아니오
6-트리플루오로메톡시-슈도-UTP	--	U	아니오
6-트리플루오로메틸-슈도-UTP	--	U	아니오
알파-티오-슈도-UTP	--	U	아니오
슈도우리딘 1-(4-메틸벤젠설포산) TP	--	U	아니오
슈도우리딘 1-(4-메틸벤조산)TP	--	U	아니오
슈도우리딘 TP 1-[3-(2-에톡시)]프로피온산	--	U	아니오
슈도우리딘 TP 1-[3-{2-(2-[2-(2-에톡시)-에톡시]-에톡시)-에톡시}]프로피온산	--	U	아니오
슈도우리딘 TP 1-[3-{2-(2-[2-{2(2-에톡시)-에톡시]-에톡시)-에톡시}-에톡시}]프로피온산	--	U	아니오
슈도우리딘 TP 1-[3-{2-(2-[2-에톡시]-에톡시)-에톡스브}]	--	U	아니오
슈도우리딘 TP 1-[3-{2-(2-에톡시)-에톡스브}] 프로피온산	--	U	아니오
슈도우리딘 TP 1-메틸포스포산	--	U	아니오
슈도우리딘 TP 1-메틸포스포산 디에틸 에스테르	--	U	아니오
슈도-UTP-N1-3-프로피온산	--	U	아니오
슈도-UTP-N1-4-부탄산	--	U	아니오
슈도-UTP-N 1-5-펜탄산	--	U	아니오
슈도-UTP-N1-6-헥산산	--	U	아니오
슈도-UTP-N1-7-헵탄산	--	U	아니오
슈도-UTP-N1-메틸-p-벤조산	--	U	아니오
슈도-UTP-N1-p-벤조산	--	U	아니오

[0286]

[0287]

일부 실시형태에서, 본원에 기재된 유전 요소는 표 6에 제공된 변형 또는 이들의 조합을 포함한다. 표 6에 제공된 변형은 RNA에서 자연적으로 발생하며, 자연에서 발생하지 않는 위치에서의 유전 요소의 변형이 본원에서 사용될 수 있다.

[0288] 표 6: 추가의 예시적인 변형

명칭	기호	염기	자연 발생
2-메틸티오-N6-(시스-호브드록스비소펜테닐)아데노신	ms2i6A	A	예
2-메틸티오-N6-메틸아데노신	ms2m6A	A	예
2-메틸티오-N6-트레오닐카바모일아데노신	ms2t6A	A	예
N6-글리시닐카바모일아데노신	g6A	A	예
N6-이소펜테닐아데노신	i6A	A	예
N6-메틸아데노신	m6A	A	예
N6-트레오닐카바모일아데노신	t6A	A	예
1,2'-O-디메틸아데노신	m1Am	A	예
1-메틸아데노신	m1A	A	예
2'-O-메틸아데노신	Am	A	예
2'-O-리보실아데노신 (인산염)	Ar(p)	A	예
2-메틸아데노신	m2A	A	예
2-메틸티오-N6 이소펜테닐아데노신	ms2i6A	A	예
2-메틸티오-N6-하이드록시노르발릴카바모일아데노신	ms2hn6A	A	예
2'-O-메틸아데노신	m6A	A	예
2'-O-리보실아데노신 (인산염)	Ar(p)	A	예
이소펜테닐아데노신	Iga	A	예
N6-(시스-하이드록시이소펜테닐)아데노신	io6A	A	예
N6,2'-O-디메틸아데노신	m6Am	A	예
N6,2'-O-디메틸아데노신	m <sup>6</sup> Am	A	예
N6,N6,2'-O-트리메틸아데노신	m62Am	A	예
N6,N6-디메틸아데노신	m62A	A	예
N6-아세틸아데노신	ac6A	A	예
N6-하이드록시노르발릴카바모일아데노신	hn6A	A	예
N6-메틸-N6-트레오닐카바모일아데노신	m6t6A	A	예
2-메틸아데노신	m <sup>2</sup> A	A	예
2-메틸티오-N <sup>6</sup> -이소펜테닐아데노신	ms <sup>2</sup> i <sup>6</sup> A	A	예
2-티오시티딘	s2C	C	예

[0289]

3-메틸시티딘	m3C	C	예
5-포르밀시티딘	f5C	C	예
5-하이드록시메틸시티딘	hm5C	C	예
5-메틸시티딘	m5C	C	예
N4-아세틸시티딘	ac4C	C	예
2'-O-메틸시티딘	Cm	C	예
2'-O-메틸시티딘	Cm	C	예
5,2'-O-디메틸시티딘	m5Cm	C	예
5-포르밀-2'-O-메틸시티딘	f5Cm	C	예
리시티딘	k2C	C	예
N4,2'-O-디메틸시티딘	m4Cm	C	예
N4-아세틸-2'-O-메틸시티딘	ac4Cm	C	예
N4-메틸시티딘	m4C	C	예
N4,N4-디메틸-2'-OMe-시티딘 TP	--	C	예
7-메틸구아노신	m7G	G	예
N2,2'-O-디메틸구아노신	m2Gm	G	예
N2-메틸구아노신	m2G	G	예
와이오신	imG	G	예
1,2'-O-디메틸구아노신	m1Gm	G	예
1-메틸구아노신	m1G	G	예
2'-O-메틸구아노신	Gm	G	예
2'-O-리보실구아노신 (인산염)	Gr(p)	G	예
2'-O-메틸구아노신	Gm	G	예
2'-O-리보실구아노신 (인산염)	Gr(p)	G	예
7-아미노메틸-7-데아자구아노신	preQ1	G	예
7-시아노-7-데아자구아노신	preQ0	G	예
아캐오신	G+	G	예
메틸와이오신	mimG	G	예
N2,7-디메틸구아노신	m2,7G	G	예
N2,N2,2'-O-트리메틸구아노신	m22Gm	G	예
N2,N2,7-트리메틸구아노신	m2,2,7G	G	예
N2,N2-디메틸구아노신	m22G	G	예

[0290]

N2, 7,2 '-O-트리메틸쿠아노신	m2 7 ' Gm	G	예
1-메틸이노신	mI	A	예
이노신	I	A	예
1,2'-O-디메틸이노신	mIim	A	예
2'-O-메틸이노신	Im	A	예
2' -O-메틸이노신	Im	A	예
에폭시 케오신	oQ	G	예
갈락토실- 케오신	galQ	G	예
만노실- 케오신	manQ	G	예
2' -O-메틸우리딘	--	U	예
2-티오우리딘	s2U	U	예
3-메틸우리딘	m3U	U	예
5-카복시메틸우리딘	cm5U	U	예
5-하이드록시우리딘	ho5U	U	예
5-메틸우리딘	m5U	U	예
5-타우리노메틸-2-티오우리딘	rm5s2U	U	예
5-타우리노메틸우리딘	rm5U	U	예
디하이드로우리딘	D	U	예
슈도우리딘	Q	U	예
(3-(3-아미노-3-카복시프로필)우리딘	acp3U	U	예
1-메틸-3-(3-아미노-5- 카복시 프로필)슈도우리딘	mIacp3'P	U	예
1-메틸슈도우리딘	mI'P	U	예
1-메틸-슈도우리딘	--	U	예
2'-0-메틸우리딘	Um	U	예
2'-0-메틸슈도우리딘	'Pm	U	예
2' -0-메틸우리딘	Um	U	예
2-티오-2'-0-메틸우리딘	s2Um	U	예
3-(3-아미노-3-카복시 프로필)우리딘	acp3U	U	예
3,2'-0-디메틸우리딘	m3Um	U	예
3-메틸-슈도-우리딘 TP	--	U	예
4-티오우리딘	s4U	U	예

[0291]

5-(카복시하이드록시메틸)우리딘	chm5U	U	예
5-(카복시하이드록시메틸)우리딘 메틸 에스테르	mchm5U	U	예
5,2'-0-디메틸우리딘	m5Um	U	예
5,6-디하이드로-우리딘	--	U	예
5-아미노메티 1-2-티오우리딘	nm5s2U	U	예
5-카바모일메틸-2'-0-메틸우리딘	ncm5Um	U	예
5-카바모일메틸우리딘	ncm5U	U	예
5-카복시하이드록시메틸우리딘	--	U	예
5-카복시하이드록시메틸우리딘 메틸 에스테르	--	U	예
5-카복시메틸아미노메틸-2'-0-메틸우리딘	cmnm5Um	U	예
5-카복시메틸아미노메틸-2- 티오우리딘	cmnm5s2U	U	예
5-카복시메틸아미노메틸-2- 티오우리딘	--	U	예
5-카복시메틸아미노메틸우리딘	cmnm5U	U	예
5-카복시메틸아미노메틸우리딘	--	U	예
5-카바모일메틸우리딘 TP	--	U	예
5-메톡시카보닐메틸-2'-0-메틸우리딘	mcm5Um	U	예
5-메톡시카보닐메틸-2-티오우리딘	mcm5s2U	U	예
5-메톡시카보닐메틸우리딘	mcm5U	U	예
5-메톡시우리딘	mo5U	U	예
5-메틸-2-티오우리딘	m5s2U	U	예
5-메틸아미노메틸-2-셀레노우리딘	mm5se2U	U	예
5-메틸아미노메틸-2-티오우리딘	mm5s2U	U	예
5-메틸아미노메틸우리딘	mm5U	U	예
5-메틸디하이드로우리딘	--	U	예
5-0 시아세트산- 우리딘 TP	--	U	예
5-0 시아세트산-메틸 에스테르-우리딘 TP	--	U	예
N1-메틸-슈도-우리딘	--	U	예
우리딘 5-옥시아세트산	cmo5U	U	예
우리딘 5-옥시아세트산 메틸 에스테르	mcmo5U	U	예
3-(3-아미노-3-카복시프로필)-우리딘	--	U	예

[0292]

TP			
5-(이소-펜테닐아미노메틸)- 2- 티오우리딘 TP	--	U	예
5-(이소-펜테닐아미노메틸)-2'-0- 메틸우리딘 TP	--	U	예
5-(이소-펜테닐아미노메틸)우리딘 TP	--	U	예
와이부토신	yW	A/T	예
하이드록시와이부토신	OHyW	A/T	예
이소와이오신	imG2	A/T	예
퍼옥시와이부토신	o2yW	A/T	예
미변형된 하이드록시와이부토신	OHyW*	A/T	예
4-데메틸와이오신	imG-14	A/T	예

[0293]

[0294]

일 실시형태에서, 본원에 기재된 유전 요소는 표 7에 제공된 비-자연 발생 변형 또는 이들의 조합을 포함한다.

[0295] 표 7: 추가의 예시적인 비-자연 발생 변형

명칭
2,6-(디아미노)퓨린
1-(아자)-2-(티오)-3-(아자)-페녹사진-1-일
1,3-(디아자)-2-(옥소)-펜티아진-1-일
1,3-(디아자)-2-(옥소)-페녹사진-1-일
1,3,5-(트리아자)-2,6-(디옥사)-나프탈렌
2 (아미노)퓨린
2,4,5-(트리메틸)페닐
2' 메틸, 2'아미노, 2'아지도, 2'플루로-시티딘
2' 메틸, 2'아미노, 2'아지도, 2'플루로-아데닌
2'메틸, 2'아미노, 2'아지도, 2'플루로-우리딘
2'-아미노-2'-데옥시리보스
2-아미노-6-클로로-퓨린
2-아자-이노시닐
2'-아지도-2'-데옥시리보스
2'플루오로-2'-데옥시리보스
2'-플루오로-변형된 염기
2'-O-메틸-리보스
2-옥소-7-아미노피리도피리미딘-3-일
2-옥소-피리도피리미딘-3-일
2- 피리디논

[0296]

3 니트로피롤
3-(메틸)-7-(프로피닐)이소카보스티릴릴
3-(메틸)이소카보스티릴릴
4-(플루오로)-6-(메틸)벤즈이미다졸
4-(메틸)벤즈이미다졸
4-(메틸)인돌릴
4,6-(디메틸)인돌릴
5 니트로인돌
5 치환된 피리미딘
5-(메틸)이소카보스티릴릴
5-니트로인돌
6-(아자)피리미딘
6-(아조)티민
6-(메틸)-7-(아자)인돌릴
6-클로로-퓨린
6-페닐-피롤로-피리미딘-2-온-3-일
7-(아미노알킬 하이드록시)-1-(아자)-2-(티오)-3-(아자)- 펜티아진-1-일
7-(아미노알킬 하이드록시)-1-(아자)-2-(티오)-3-(아자)- 페녹사진-1-일
7-(아미노알킬 하이드록시)-1,3-(디아자)-2-(옥소)- 페녹사진-1-일
7-(아미노알킬 하이드록시)-1,3-(디아자)-2-(옥소)- 펜티아진-1-일
7-(아미노알킬 하이드록시)-1,3-(디아자)-2-(옥소)- 페녹사진-1-일
7-(아자)인돌릴
7-(구아니디늄알킬 하이드록시)-1-(아자)-2-(티오)-3- (아자)- 페녹사진 1-일
7-(구아니디늄알킬 하이드록시)-1-(아자)-2-(티오)-3- (아자)-펜티아진-1-일
7-(구아니디늄알킬 하이드록시)-1-(아자)-2-(티오)-3- (아자)- 페녹사진-1-일
7-(구아니디늄알킬 하이드록시)-1,3-(디아자)-2-(옥소)- 페녹사진-1-일
7-(구아니디늄알킬-하이드록시)-1,3-(디아자)-2-(옥소)- 펜티아진-1-일
7-(구아니디늄알킬 하이드록시)-1,3-(디아자)-2-(옥소)- 페녹사진-1-일
7-(프로피닐)이소카보스티릴릴
7-(프로피닐)이소카보스티릴릴, 프로피닐-7- (아자)인돌릴
7-데아자-이노시닐
7-치환된 1-(아자)-2-(티오)-3-(아자)-페녹사진-1-일

[0297]

7-치환된 1,3-(디아자)-2-(옥소)-페녹사진-1-일
9-(메틸)-이미디조피리디닐
아미노인돌릴
안트라세닐
비스-오르토-(아미노알킬하이드록시)-6-페닐-피롤로- 느브리미딘(nvrimidin)-2-온-3-일
비스-오르토-치환된-6-페닐-피롤로-피리미딘-2-온-3-일
디플루오로톨릴
하이포크산틴
이미디조피리디닐
이노시닐
이소카보스티릴릴
이소구아닌
N2-치환된 퓨린
N6-메틸-2-아미노-퓨린
N6-치환된 퓨린
N-알킬화된 유도체
나프탈레닐
니트로벤즈이미다졸릴
니트로이미다졸릴
니트로인다졸릴
니트로피라졸릴
누볼라린
O6-치환된 퓨린
O-알킬화된 유도체
오르토-(아미노알킬하이드록시)-6-페닐-피롤로- 피리미딘-2-온-3-일
오르토-치환된-6-페닐-피롤로-피리미딘-2-온-3-일
옥소 포르마이신 TP
파라-(아미노알킬하이드록시)-6-페닐-피롤로-피리미딘- 2-온-3-일
파라-치환된-6-페닐-피롤로-피리미딘-2-온-3-일
펜타세닐
페난트라세닐
페닐
프로피닐-7-(아자)인돌릴
피렌일
피리도피리미딘-3-일
피리도피리미딘-3-일, 2-옥소-7-아미노-피리도피리미딘- 3-일
피롤로-피리미딘-2-온-3-일

[0298]

피롤로피리미디닐
피롤로피리지닐
스틸벤질
치환된 1,2,4-트리아졸
테트라세닐
투베르시딘
크산틴
크산토신-5'-TP
2-티오-제볼라린
5-아자-2-티오-제볼라린
7-데아자-2-아미노-퓨린
피리딘-4-온 리보뉴클레오시드
2-아미노-리보시드-TP
포르마이신 A TP
포르마이신 B TP
피롤로신 TP
2'-OH-아라-아데노신 TP
2'-OH-아라-시티딘 TP
2'-OH-아라-우리딘 TP
2'-OH-아라-구아노신 TP
5-(2-카보메톡시비닐)우리딘 TP
N6-(19-아미노-펜타옥사노나데실)아데노신 TP

[0299]

[0300]

일 실시형태에서, 본원에 기재된 유전 요소는 표 8에 제공된 비-자연 발생 변형 또는 이들의 조합을 포함한다.

[0301] 표 8: 예시적인 백본 변형

명칭
3'-알킬렌 포스포네이트
3'-아미노 포스포라미데이트
알켄 함유 백본
아미노알킬 포스포라미데이트
아미노알킬 포스포트리에스테르
보라노포스페이트
-CH <sub>2</sub> -O-N(CH <sub>3</sub> )-CH <sub>2</sub> -
-CH <sub>2</sub> -N(CH <sub>3</sub> )-N(CH <sub>3</sub> )-CH <sub>2</sub> -
-CH <sub>2</sub> -NH-CH <sub>2</sub> -
키랄 포스포네이트
키랄 포스포로티오에이트
포름아세틸 및 티오포름아세틸 백본
메틸렌 (메틸이미노)
메틸렌 포름아세틸 및 티오포름아세틸 백본
메틸렌이미노 및 메틸렌하이드라지노 백본
모르폴리노 연결
-N(CH <sub>3</sub> )-CH <sub>2</sub> -CH <sub>2</sub> -
헤테로원자 뉴클레오시드간 연결을 갖는 올리고뉴클레오시드
포스피네이트
포스포라미데이트
포스포로티오에이트
포스포로티오에이트 뉴클레오시드간 연결
포스포로티오에이트
포스포트리에스테르
PNA
실록산 백본
실파메이트 백본
실파이드 실폭사이드 및 실폰 백본
실포네이트 및 실폰아미드 백본
티오노알킬 포스포네이트
티오노알킬 포스포트리에스테르
티오노포스포라미데이트

[0302]

[0303] 일 실시형태에서, 본원에 기재된 유전 요소는 표 9에 제공된 비-자연 발생 변형 또는 이들의 조합을 포함한다.

[0304] 표 9: 예시적인 비-자연 발생 백본 변형

<b>합성 백본 변형의 명칭</b>
포스포로티오에이트
제약된 핵산(CNA)
2' O-메틸화
2'-O-메톡시에틸리보스(MOE)
2' 플루오로
잠금 핵산(LNA)
(S)-제약된 에틸( <i>cEt</i> )
플루오로 핵시톨 핵산(FHNA)
5'포스포로티오에이트
포스포로디아미데이트 모르폴리노 올리고머(PMO)
트리사이클로-DNA(tcDNA)
(S) 5'-C-메틸
(E)-비닐포스포네이트
메틸 포스포네이트
인산염을 갖는 (S) 5'-C-메틸

[0305]

[0306]

일부 실시형태에서, 유전 요소는 캡을 포함한다. 캡은 전형적으로 mRNA의 5' 말단에 위치하지만, 캡은 RNA의 3' 말단에도 위치할 수 있다. 일부 실시형태에서, 캡은 역소뉴클레아제 분해로부터 유전 요소를 보호하고, 세포 내 운반 및/또는 국소화를 도울 수 있다. 캡은 5'-말단(5'-캡) 또는 3'-말단(3'-캡)에 존재할 수 있거나, 두 말단 모두에 존재할 수 있다. 5'-캡의 비-제한적인 예는 글리세릴, 역전 데옥시 탈염기성 잔기(모이어티); 4',5'-메틸렌 뉴클레오티드; 1-(베타-D-에리트로푸라노실) 뉴클레오티드, 4'-티오 뉴클레오티드; 카보사이클릭 뉴클레오티드; 1,5-안하이드로핵시톨(anhydrohexitol) 뉴클레오티드; L-뉴클레오티드; 알파-뉴클레오티드; 변형된 염기 뉴클레오티드; 포스포로디티오에이트 연결; 트레오-펜토푸라노실 뉴클레오티드; 비환식 3',4'-세코 뉴클레오티드; 비환식 3,4-디하이드록시부틸 뉴클레오티드; 비환식 3,5-디하이드록시펜틸 뉴클레오티드, 3'-3'-역전 뉴클레오티드 모이어티; 3'-3'-역전 탈염기성 모이어티; 3'-2'-역전 뉴클레오티드 모이어티; 3'-2'-역전 탈염기성 모이어티; 1,4-부탄디올 포스페이트; 3'-포스포르아미데이트; 핵실포스페이트; 아미노핵실포스페이트; 3'-포스페이트; 3'-포스포로티오에이트; 포스포로디티오에이트; 또는 가교 또는 비-가교 메틸포스포네이트 모이어티를 포함하나, 이에 제한되지 않는다.

[0307]

3'-캡의 비-제한적인 예는 글리세릴, 역전 데옥시 탈염기성 잔기(모이어티), 4',5'-메틸렌 뉴클레오티드; 1-(베타-D-에리트로푸라노실) 뉴클레오티드; 4'-티오 뉴클레오티드, 카보사이클릭 뉴클레오티드; 5'-아미노-알킬 포스페이트; 1,3-디아미노-2-프로필 포스페이트; 3-아미노프로필 포스페이트; 6-아미노핵실포스페이트; 1,2-아미노도데실 포스페이트; 하이드록시프로필 포스페이트; 1,5-안하이드로핵시톨 뉴클레오티드; L-뉴클레오티드; 알파-뉴클레오티드; 변형된 염기 뉴클레오티드; 포스포로디티오에이트; 트레오-펜토푸라노실 뉴클레오티드; 비환식 3',4'-세코 뉴클레오티드; 3,4-디하이드록시부틸 뉴클레오티드; 3,5-디하이드록시펜틸 뉴클레오티드, 5'-5'-역전 뉴클레오티드 모이어티; 5'-5'-역전 탈염기성 모이어티; 5'-포스포르아미데이트; 5'-포스포로티오에이트; 1,4-부탄디올 포스페이트; 5'-아미노; 가교 및/또는 비-가교 5'-포스포르아미데이트, 포스포로티오에이트 및/또는 포스포로디티오에이트, 가교 또는 비-가교 메틸포스포네이트 및 5'-머캡토 모이어티(상세한 내용은 본원에 참조로 포함된 문헌[Beaucage and Iyer, 1993, Tetrahedron 49, 1925] 참조)를 포함하나, 이에 제한되지 않는다.

[0308]

일부 실시형태에서, 유전 요소는 폴리-A 꼬리를 포함한다. 일부 실시형태에서, 폴리-A 꼬리는 적어도 약 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 또는 100개의 아데노신 길이를 포함한다. 일부 실시형태에서, RNA에는 폴리-A 꼬리가 결여되어 있다. RNA에 폴리-A 꼬리가 결여된 일부 실시형태에서, RNA는 약 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 또는 100개 이하의 순차적인 아데노신을 포함한다.

[0309]

일부 실시형태에서, 유전 요소는 선형이다. 일부 실시형태에서, 유전 요소는 환형이다. 일부 실시형태에서, 유전 요소는 제1 영역, 및 제1 영역과 혼성화될 수 있는 제2 영역을 포함한다. 일부 실시형태에서, 유전 요소는

제1 영역, 및 원을 형성하기 위해 제1 영역과 혼성화될 수 있는 제2 영역을 포함한다. 일부 실시형태에서, 유전 요소는 5' 또는 3' 말단을 포함하지 않는다. 일부 실시형태에서, 유전 요소는 유리 인산염 및 유리 당 중 하나 또는 둘 모두를 포함하지 않는다. 일부 실시형태에서, 유전 요소의 모든 인산염은 인산염에 포함된 제1 산소 원자에 의해 제1 당에 공유적으로 연결되고, 인산염에 포함된 제2 산소 원자에 의해 제2 당에 공유적으로 연결된다. 일부 실시형태에서, 유전 요소 내의 모든 당은 당에 포함된 제1 탄소 원자에 의해 제1 인산염에 공유적으로 연결되고, 당에 포함된 제2 탄소 원자에 의해 제2 인산염에 공유적으로 연결된다. 일부 실시형태에서, 유전 요소는 선형 RNA를 환화함으로써 생성된다. 환형 RNA는 예를 들어, 그의 전문이 본원에 참조로 포함되는 미국 특허 공개공보 20200306286호에 기재되어 있다.

[0310] 일부 실시형태에서, 유전 요소는 약 10 내지 20, 20 내지 30, 30 내지 40, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 125, 125 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 900 내지 1000, 1000 내지 1500, 1500 내지 2000, 2000 내지 2500, 2500 내지 3000, 3000 내지 3500, 3500 내지 4000 또는 4000 내지 4500개의 뉴클레오티드 길이이다. 일부 실시형태에서, 유전 요소는 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000 또는 4500개의 뉴클레오티드 길이이다.

[0311] **RNA-전용 유전 요소**

[0312] 일부 실시형태에서, 유전 요소는 RNA로 이루어지거나 본질적으로 이로 이루어진다. 예를 들어, 일부 실시형태에서, 유전 요소는 DNA가 실질적으로 없다. 일부 실시형태에서, 유전 요소는 단일 가닥 RNA를 포함한다. 일부 실시형태에서, 유전 요소는 적어도 하나의 이중 가닥 영역을 포함한다. 일부 실시형태에서, 유전 요소의 이중 가닥 영역은 RNA와 쌍을 형성하는 RNA의 영역을 포함한다.

[0313] **혼성화된 RNA-ssDNA 유전 요소**

[0314] 일부 실시형태에서, 유전 요소는 DNA 영역을 포함한다. 일부 실시형태에서, RNA를 포함하는 유전 요소는 DNA 영역을 추가로 포함한다. 예를 들어, 유전 요소는 단일 가닥일 수 있으며, 여기서 단일 가닥의 제1 부분은 리보뉴클레오티드를 포함하고, 단일 가닥의 제2 부분은 데옥시리보뉴클레오티드를 포함한다. 일부 실시형태에서, DNA 영역을 포함하는 유전 요소는 화학적 변형을 갖는 하나 이상의 DNA 뉴클레오티드를 포함한다. 일부 실시형태에서, 유전 요소는 DNA 영역을 포함하며, 여기서 DNA 영역의 모든 뉴클레오티드는 화학적으로 변형된다.

[0315] 일부 실시형태에서, 유전 요소의 적어도 일부는 단일 가닥이다. 일부 실시형태에서, 유전 요소는 단일 가닥이다. 일부 실시형태에서, 유전 요소는 ssDNA를 포함한다. 일부 실시형태에서, 유전 요소는 이중 가닥 영역을 포함한다. 일부 실시형태에서, 유전 요소의 이중 가닥 영역은 RNA와 쌍을 형성하는 RNA의 영역을 포함한다. 이러한 일부 실시형태에서, 유전 요소의 이중 가닥 영역은 RNA와 쌍을 형성하는 DNA의 영역을 포함한다. 일부 실시형태에서, DNA 영역의 적어도 일부는 유전 요소의 RNA의 적어도 일부에 혼성화된다.

[0316] 일부 실시형태에서, DNA 영역은 약 10 내지 20, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 또는 900 내지 1000개의 뉴클레오티드 길이이다.

[0317] **RNA/DNA 컨주게이트**

[0318] 일부 실시형태에서, 유전 요소는 DNA 영역을 포함한다. 일부 실시형태에서, RNA를 포함하는 유전 요소는 DNA 영역을 추가로 포함한다. 일부 실시형태에서, DNA 영역을 포함하는 유전 요소는 화학적 변형을 갖는 하나 이상의 DNA 뉴클레오티드를 포함한다. 일부 실시형태에서, 유전 요소는 DNA 영역을 포함하며, 여기서 DNA 영역의 모든 뉴클레오티드는 화학적으로 변형된다.

[0319] 일부 실시형태에서, 유전 요소의 적어도 일부는 단일 가닥이다. 일부 실시형태에서, 유전 요소는 단일 가닥이다. 일부 실시형태에서, 유전 요소는 ssDNA를 포함한다. 일부 실시형태에서, 유전 요소는 이중 가닥 영역을 포함한다. 일부 실시형태에서, 유전 요소의 이중 가닥 영역은 RNA와 쌍을 형성하는 RNA의 영역을 포함한다. 이러한 일부 실시형태에서, 유전 요소의 이중 가닥 영역은 RNA와 쌍을 형성하는 DNA의 영역을 포함한다. 유전 요소가 RNA를 포함하는 일부 실시형태에서, DNA 영역은 유전 요소의 RNA에 공유적으로 연결된다.

[0320] 일부 실시형태에서, DNA 영역은 약 10 내지 20, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 60, 60 내지

70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 150, 150 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900 또는 900 내지 1000개의 뉴클레오티드 길이이다.

[0321] **유전 요소 작제물**

[0322] 일부 실시형태에서, 유전 요소는 유전 요소 작제물로부터 생성된다. 예를 들어, 일부 실시형태에서, 유전 요소 작제물은 DNA, 예를 들어, 이중 가닥 DNA이고, 유전 요소는 RNA 유전 요소를 생성하는 전사에 의해 생성될 수 있다.

[0323] 본원에 기재된 바와 같은 아벨로벡터의 유전 요소는 유전 요소 영역 및 선택적으로 (예를 들어, 배콜로바이러스 계놈 또는 그의 단편, 예를 들어, 하나 이상의 배콜로바이러스 요소를 포함하는) 박미드 또는 공여자 벡터 백본과 같은 다른 서열을 포함하는 유전 요소 작제물로부터 생성될 수 있다. 일부 실시형태에서, 유전 요소 작제물은 아벨로바이러스 5' UTR(예를 들어, 본원에 기재된 바와 같음)을 포함한다. 유전 요소 작제물은 유전 요소가 단백질성 외부 내에 봉입될 수 있는 숙주 세포 또는 무세포 시스템 내로의 유전 요소의 서열의 운반에 적합한 임의의 핵산 작제물일 수 있다. 일부 실시형태에서, 유전 요소 작제물은 프로모터를 포함한다. 일부 실시형태에서, 유전 요소 작제물로부터의 전사는 RNA 유전 요소를 생성한다.

[0324] 일부 실시형태에서, 유전 요소 작제물은 선형 핵산 분자이다. 일부 실시형태에서, 유전 요소 작제물은 환형 핵산 분자(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 플라스미드, 박미드, 공여자 벡터 또는 미니서클)이다. 유전 요소 작제물은, 일부 실시형태에서, 이중-가닥일 수 있다. 다른 실시형태에서, 유전 요소는 단일-가닥이다. 일부 실시형태에서, 유전 요소 작제물은 DNA를 포함한다. 일부 실시형태에서, 유전 요소 작제물은 RNA를 포함한다. 일부 실시형태에서, 유전 요소 작제물은 하나 이상의 변형된 뉴클레오티드를 포함한다.

[0325] **플라스미드**

[0326] 일부 실시형태에서, 유전 요소 작제물은 플라스미드이다. 플라스미드는 일반적으로 본원에 기재된 바와 같은 유전 요소의 서열뿐만 아니라 숙주 세포에서의 복제에 적합한 복제 원점(예를 들어, 박테리아 세포에서의 복제를 위한 박테리아 복제 원점) 및 선택 가능한 마커(예를 들어, 항생제 저항성 유전자)를 포함할 것이다. 일부 실시형태에서, 유전 요소의 서열은 플라스미드로부터 절제될 수 있다. 일부 실시형태에서, 플라스미드는 박테리아 세포에서 복제될 수 있다. 일부 실시형태에서, 플라스미드는 포유동물 세포(예를 들어, 인간 세포)에서 복제될 수 있다. 일부 실시형태에서, 플라스미드는 적어도 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000 또는 5000 bp의 길이이다. 일부 실시형태에서, 플라스미드는 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 또는 10,000 bp 미만의 길이이다. 일부 실시형태에서, 플라스미드는 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 900 내지 1000, 1000 내지 1500, 1500 내지 2000, 2000 내지 2500, 2500 내지 3000, 3000 내지 4000 또는 4000 내지 5000 bp의 길이를 갖는다.

[0327] **작은 환형 핵산 작제물**

[0328] 일부 실시형태에서, 유전 요소 작제물은 예를 들어, 벡터 백본이 결여된(예를 들어, 박테리아 복제 원점 및/또는 선택가능한 마커가 결여된) 환형 핵산 작제물이다. 실시형태들에서, 유전 요소는 단일-가닥 또는 이중-가닥 환형 핵산 작제물이다. 실시형태들에서, 환형 핵산 작제물은 예를 들어, 본원에 기재된 바와 같이 시험관 내 환화(IVC)에 의해 생성된다. 실시형태들에서, 이중-가닥 환형 핵산 작제물은 숙주 세포 내로 도입될 수 있으며, 여기에서 이는 예를 들어, 본원에 기재된 바와 같이 단일-가닥 환형 유전 요소를 생성하기 위한 주형으로 전환되거나 주형으로서 사용될 수 있다. 일부 실시형태에서, 환형 핵산 작제물은 플라스미드 백본 또는 그의 기능적 단편을 포함하지 않는다. 일부 실시형태에서, 환형 핵산 작제물은 적어도 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400 또는 4500 bp의 길이이다. 일부 실시형태에서, 환형 핵산 작제물은 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5500 또는 6000 bp 미만의 길이이다. 일부 실시형태에서, 환형 핵산 작제물은 2000 내지 2100, 2100 내지 2200, 2200 내지 2300, 2300 내지 2400, 2400 내지 2500, 2500 내지 2600, 2600 내지 2700, 2700 내지 2800, 2800 내지 2900, 2900 내지 3000, 3000 내지 3100, 3100 내지 3200, 3200 내지 3300, 3300 내지 3400, 3400 내지 3500, 3500 내지 3600, 3600 내지 3700, 3700 내지 3800, 3800 내지 3900, 3900 내지 4000, 4000 내지 4100, 4100 내지 4200, 4200 내지 4300, 4300 내지 4400 또는 4400 내지 4500 bp 길이이다. 일부 실시형태

에서, 환형 핵산 작제물은 미니서클이다.

[0329] *시스/트랜스 작제물*

[0330] 일부 실시형태에서, 본원에 기재된 바와 같은 유전 요소 작제물(예를 들어, 박미드 또는 공여자 벡터)은 하나 이상의 아넬로바이러스 ORF, 예를 들어, 단백질성 외부 구성성분(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드)을 인코딩하는 하나 이상의 서열을 포함한다. 예를 들어, 유전 요소 작제물은 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열을 포함할 수 있다. 이러한 유전 요소 작제물은 유전 요소 및 아넬로바이러스 ORF(들)를 시스로 숙주 세포 내로 도입하는데 적합할 수 있다. 다른 실시형태에서, 본원에 기재된 바와 같은 유전 요소 작제물은 하나 이상의 아넬로바이러스 ORF, 예를 들어, 단백질성 외부 구성성분(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드)을 인코딩하는 서열을 포함하지 않는다. 예를 들어, 유전 요소 작제물은 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열을 포함하지 않을 수 있다. 이러한 유전 요소 작제물은 (예를 들어, 아넬로바이러스 ORF 중 하나 이상을 인코딩하는 제2 핵산 작제물의 도입을 통해, 또는 숙주 세포의 게놈 내로 통합된 아넬로바이러스 ORF 카세트를 통해) 트랜스로 제공될 하나 이상의 아넬로바이러스 ORF와 함께 유전 요소를 숙주 세포 내로 도입하기에 적합할 수 있다. 일부 실시형태에서, 유전 요소 작제물은 곤충 세포(예를 들어, Sf9 세포)에서의 핵산 작제물의 복제에 적합한 백본, 예를 들어, 배칼로바이러스 백본 영역을 포함한다. 일부 실시형태에서, 유전 요소 작제물은 박테리아 세포(예를 들어, 이. 콜라이 세포, 예를 들어, DH 10Bac 세포)에서의 유전 요소 작제물의 복제에 적합한 백본 영역을 포함한다.

[0331] 일부 실시형태에서, 유전 요소 작제물(예를 들어, 박미드 또는 공여자 벡터)은 아넬로바이러스 ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 실시형태들에서, 유전 요소의 서열을 포함하지 않는 유전 요소의 일부는 (예를 들어, 프로모터, 및 아넬로바이러스 ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편을 인코딩하는 서열을 포함하는 카세트에서) 아넬로바이러스 ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편을 인코딩하는 서열을 포함한다. 추가 실시형태에서, 유전 요소의 서열을 포함하는 작제물의 일부는 아넬로바이러스 ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 실시형태들에서, (예를 들어, 본원에 기재된 바와 같은) 단백질성 외부에서의 이러한 유전 요소의 봉입은 복제-구성성분인 아넬로벡터(예를 들어, 세포를 감염시킬 때, 세포가 예를 들어, 본원에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF를 인코딩하는 추가의 핵산 작제물을 세포 내로 도입하지 않고 아넬로벡터의 추가 카피를 생성할 수 있게 하는 아넬로벡터)를 생성한다.

[0332] 다른 실시형태에서, 유전 요소는 아넬로바이러스 ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함하지 않는다. 실시형태들에서, 단백질성 외부(예를 들어, 본원에 기재된 바와 같은)에서 이러한 유전 요소의 봉입은 복제-불능 아넬로벡터(예를 들어, 세포를 감염시킬 때, 감염된 세포가 예를 들어, 본원에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF를 인코딩하는 예를 들어, 하나 이상의 추가 작제물의 부재 하에 추가 아넬로벡터를 생성할 수 없게 하는 아넬로벡터)를 생성한다.

[0333] *발현 카세트*

[0334] 일부 실시형태에서, 유전 요소 작제물(예를 들어, 박미드 또는 공여자 벡터)은 폴리펩티드 또는 비코딩 RNA(예를 들어, miRNA 또는 siRNA)의 발현을 위한 하나 이상의 카세트를 포함한다. 일부 실시형태에서, 유전 요소 작제물은 이펙터(예를 들어, 외인성 또는 내인성 이펙터), 예를 들어, 본원에 기재된 바와 같은 폴리펩티드 또는 비코딩 RNA의 발현을 위한 카세트를 포함한다. 일부 실시형태에서, 유전 요소 작제물은 아넬로바이러스 단백질(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 또는 ORF1/2 또는 그의 기능적 단편)의 발현을 위한 카세트를 포함한다. 발현 카세트는, 일부 실시형태에서, 유전 요소 서열 내에 위치할 수 있다. 실시형태들에서, 이펙터용 발현 카세트는 유전 요소 서열 내에 위치한다. 실시형태들에서, 아넬로바이러스 단백질용 발현 카세트는 유전 요소 서열 내에 위치한다. 다른 실시형태에서, 발현 카세트는 유전 요소의 서열 외부에 있는 유전 요소 작제물 내의 위치(예를 들어, 백본)에 위치한다. 실시형태들에서, 아넬로바이러스 단백질용 발현 카세트는 유전 요소의 서열 외부에 있는 유전 요소 작제물 내의 위치(예를 들어, 백본)에 위치한다.

[0335] 폴리펩티드 발현 카세트는 일반적으로 프로모터, 및 폴리펩티드, 예를 들어, 이펙터(예를 들어, 본원에 기재된 바와 같은 외인성 또는 내인성 이펙터) 또는 아넬로바이러스 단백질을 인코딩하는 코딩 서열(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, ORF1/2 또는 그의 기능적 단편을 인코딩하는 서열)을

포함한다. (예를 들어, 폴리펩티드의 발현을 구동하기 위해) 폴리펩티드 발현 카세트에 포함될 수 있는 예시적인 프로모터는 제한 없이, 예를 들어, 본원에 기재된 바와 같은, 구성적 프로모터(예를 들어, CMV, RSV, PGK, EF1a 또는 SV40), 세포 또는 조직-특이적 프로모터(예를 들어, 골격 α-액틴 프로모터, 미오신 경쇄 2A 프로모터, 디스트로핀 프로모터, 근육 크레아틴 키나제 프로모터, 간 알부민 프로모터, B형 간염 바이러스 코어 프로모터, 오스테오칼신 프로모터, 골 시알로단백질 프로모터, CD2 프로모터, 면역글로불린 중쇄 프로모터, T 세포 수용체 α쇄 프로모터, 뉴런-특이적 에놀라제(NSE) 프로모터 또는 신경필라멘트 경쇄 프로모터) 및 유도성 프로모터(예를 들어, 아연-유도성 양 메탈로티오닌(MT) 프로모터; 텍사메타손(Dex)-유도성 마우스 유방 종양 바이러스(MMTV) 프로모터; T7 중합효소 프로모터 시스템, 테트라사이클린-억제성 시스템, 테트라사이클린-유도성 시스템, RU486-유도성 시스템, 라파마이신-유도성 시스템)을 포함한다. 일부 실시형태에서, 발현 카세트는 예를 들어, 본원에 기재된 바와 같은 인헌서를 추가로 포함한다.

[0336] **RNA-기반 유전 요소의 생성**

[0337] RNA-기반 유전 요소는 다양한 방법에 의해 생성될 수 있다. 예를 들어, DNA를 포함하는 유전 요소 작제물은 예를 들어, 위에 기재된 바와 같이 RNA를 포함하는 유전 요소를 생성하기 위해 전사될 수 있다. 전사는 예를 들어, 세포 또는 무세포 시스템에서 일어날 수 있다. RNA는 예를 들어, 고체 상 합성에 의해 시험관 내에서 합성될 수 있다.

[0338] **단백질 구성성분의 생성**

[0339] 아넬로백터의 단백질 구성성분, 예를 들어, ORF1은 본원에 기재된 다양한 방식으로 생성될 수 있다.

[0340] **배칼로바이러스 발현 시스템**

[0341] 바이러스 발현 시스템, 예를 들어, 배칼로바이러스 발현 시스템은 예를 들어, 본원에 기재된 바와 같이 (예를 들어, 아넬로백터의 생성을 위한) 단백질을 발현하기 위해 사용될 수 있다. 배칼로바이러스는 환형의 슈퍼코일링된 이중-가닥 DNA 게놈을 갖는 막대 모양의 바이러스이다. 배칼로바이러스의 속은 다음을 포함한다: 알파배칼로바이러스(레피도프테라(*Lepidoptera*)로부터 단리된 뉴클레오폴리헤드로바이러스(NPV)), 베타배칼로바이러스(레피도프테라로부터 단리된 과립바이러스(GV)), 감마배칼로바이러스(하이메노프테라(*Hymenoptera*)로부터 단리된 NPV) 및 델타배칼로바이러스(디프테라(*Diptera*)로부터 단리된 NPV). GV는 전형적으로 외피당 하나의 뉴클레오택시드만을 함유하지만, NPV는 전형적으로 외피당 단일(SNPV) 또는 다중(MNPV) 뉴클레오택시드를 함유한다. 외피 비리온은 GV의 과립 매트릭스와 NPV의 폴리헤드린에서 더 폐색된다. 배칼로바이러스는 전형적으로 용해 및 폐색 수명 주기 둘 모두를 가지고 있다. 일부 실시형태에서, 용해 및 폐색 수명 주기는 바이러스 복제의 3개의 단계: 초기, 후기 및 극후기(very late) 단계 전반에 걸쳐 독립적으로 나타난다. 일부 실시형태에서, 초기 단계 동안에, 바이러스 DNA 복제는 숙주 세포 내로의 바이러스 진입, 초기 바이러스 유전자 발현 및 숙주 유전자 발현 기구의 차단 후에 발생한다. 일부 실시형태에서, 후기 단계에서 바이러스 DNA 복제를 인코딩하는 후기 유전자가 발현되고, 바이러스 입자가 어셈블되고, 세포외 바이러스(EV)가 숙주 세포에 의해 생성된다. 일부 실시형태에서, 극후기 단계에서 폴리헤드린 및 p10 유전자가 발현되고, 폐색된 바이러스(OV)가 숙주 세포에 의해 생성되고, 숙주 세포가 용해된다. 배칼로바이러스는 곤충 중을 감염시키기 때문에, 이들은 배칼로바이러스-허용 곤충 세포 또는 유충에서 외인성 단백질을 생성하는 생물학적 제제로서 사용될 수 있다. 아우토그라파 칼리포니카(*Autographa californica*) 다중 핵다각체병 바이러스(AcMNPV) 및 봄빅스 모리(*Bombyx mori*) 핵다각체병 바이러스(BmNPV)와 같은 배칼로바이러스의 상이한 단리물은 외인성 단백질 발현에 사용될 수 있다. 다양한 배칼로바이러스 발현 시스템은 예를 들어, ThermoFisher로부터 상업적으로 이용 가능하다.

[0342] 일부 실시형태에서, 본원에 기재된 단백질(예를 들어, 아넬로바이러스 ORF 분자, 예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 또는 ORF1/2, 또는 그의 기능적 단편 또는 스플라이스 변이체)은 본원에 기재된 하나 이상의 구성성분을 포함하는 배칼로바이러스 발현 벡터(예를 들어, 박미드)를 사용하여 발현될 수 있다. 예를 들어, 배칼로바이러스 발현 벡터는 선택 가능한 마커(예를 들어, kanR), 복제 원점(예를 들어, 박테리아 복제 원점 및 곤충 세포 복제 원점 중 하나 또는 둘 모두), 재조합효소 인식 부위(예를 들어, att 부위) 및 프로모터 중 하나 이상(예를 들어, 전부)을 포함할 수 있다. 일부 실시형태에서, 배칼로바이러스 발현 벡터(예를 들어, 본원에 기재된 바와 같은 박미드)는 배칼로바이러스 폐색체를 인코딩하는 자연 발생 야생형 폴리헤드린 유전자를 본원에 기재된 단백질을 인코딩하는 유전자로 교체함으로써 생성될 수 있다. 일부 실시형태에서, 본원에 기재된 단백질을 인코딩하는 유전자는 배칼로바이러스 프로모터를 함유하는 배칼로바이러스 발현 벡터(예를 들어, 본원에 기재된 바와 같은 박미드) 내로 클로닝된다. 일부 실시형태에서, 배칼로바이러스 벡터는 하나 이상의 비-배칼로바이러스 프로모터, 예를 들어, 포유동물 프로모터 또는 아넬로바이러스 프로모터를 포함한다. 일부 실

시형태에서, 본원에 기재된 단백질을 인코딩하는 유전자는 공여자 벡터(예를 들어, 본원에 기재된 바와 같은) 내로 클로닝된 다음, 빈 배칼로바이러스 발현 벡터(예를 들어, 빈 박미드)와 접촉되며, 그에 따라 본원에 기재된 단백질을 인코딩하는 유전자는 공여자 벡터로부터 배칼로바이러스 발현 벡터(예를 들어, 박미드) 내로(예를 들어, 상동 재조합 또는 트랜스포사제 활성화에 의해) 전달된다. 일부 실시형태에서, 배칼로바이러스 프로모터는 비필수 폴리헤드린 유전자의 유전자좌로부터의 배칼로바이러스 DNA에 축적된다. 일부 실시형태에서, 본원에 기재된 단백질은 바이러스 복제의 극후기 단계에서 AcNPV 폴리헤드린 프로모터의 전사 제어 하에 있다. 일부 실시형태에서, 곤충 세포에서의 배칼로바이러스 발현에 사용하기에 적합한 강한 프로모터는 배칼로바이러스 p10 프로모터, 폴리헤드린(polh) 프로모터, p6.9 프로모터 및 캡시드 단백질 프로모터를 포함하나, 이에 제한되지 않는다. 곤충 세포에서의 배칼로바이러스 발현에 사용하기에 적합한 약한 프로모터는 배칼로바이러스의 ie1, ie2, ie0, et1, 39K(일명 pp31) 및 gp64 프로모터를 포함한다.

[0343] 일부 실시형태에서, 재조합 배칼로바이러스는 배칼로바이러스 게놈(예를 들어, 야생형 또는 돌연변이체 배칼로바이러스 게놈)과 전달 벡터 간의 상동 재조합에 의해 생성된다. 일부 실시형태에서, 본원에 기재된 단백질을 인코딩하는 하나 이상의 유전자는 전달 벡터 내로 클로닝된다. 일부 실시형태에서, 전달 벡터는 비필수 유전자의 유전자좌, 예를 들어, 폴리헤드린 유전자로부터의 DNA에 축적되는 배칼로바이러스 프로모터를 추가로 함유한다. 일부 실시형태에서, 본원에 기재된 단백질을 인코딩하는 하나 이상의 유전자는 배칼로바이러스 게놈과 전달 벡터 간의 상동성 재조합에 의해 배칼로바이러스 게놈 내로 삽입된다. 일부 실시형태에서, 배칼로바이러스 게놈은 하나 이상의 독특한 부위에서 선형화된다. 일부 실시형태에서, 선형화된 부위는 본원에 기재된 단백질을 인코딩하는 유전자를 배칼로바이러스 게놈 내로 삽입하기 위한 표적 부위 근처에 위치한다. 일부 실시형태에서, 유전자, 예를 들어, 폴리헤드린 유전자로부터 하류의 배칼로바이러스 게놈의 단편이 누락된 선형화된 배칼로바이러스 게놈은 상동 재조합에 사용될 수 있다. 일부 실시형태에서, 배칼로바이러스 게놈 및 전달 벡터는 곤충 세포 내로 공동-형질주입된다. 일부 실시형태에서, 재조합 배칼로바이러스를 생성하는 방법은 본원에 기재된 하나 이상의 단백질을 인코딩하는 유전자를 함유하는 전달 벡터와 상동 재조합을 수행하기 위한 배칼로바이러스 게놈을 제조하는 단계, 및 전달 벡터 및 배칼로바이러스 게놈 DNA를 곤충 세포 내로 공동-형질주입하는 단계를 포함한다. 일부 실시형태에서, 배칼로바이러스 게놈은 전달 벡터의 영역과 상동인 영역을 포함한다. 이들 상동 영역은 배칼로바이러스 게놈과 전달 벡터 간의 재조합의 가능성을 향상시킬 수 있다. 일부 실시형태에서, 전달 벡터의 상동성 영역은 프로모터의 상류 또는 하류에 위치한다. 일부 실시형태에서, 상동 재조합을 유도하기 위해, 배칼로바이러스 게놈 및 전달 벡터는 약 1:1 내지 10:1의 중량비로 혼합된다.

[0344] 일부 실시형태에서, 재조합 배칼로바이러스는 Tn7을 이용한 부위-특이적 전위를 포함하는 방법에 의해 생성되며, 예를 들어, 이에 의해 본원에 기재된 단백질을 인코딩하는 유전자는 박미드 DNA에 삽입되고, 예를 들어, 박테리아, 예를 들어, 이. 콜라이(예를 들어, DH 10Bac 세포)에서 증식된다. 일부 실시형태에서, 본원에 기재된 단백질을 인코딩하는 유전자는 pFASTBAC® 벡터 내로 클로닝되고, 미니-attTn7 표적 부위를 갖는 박미드 DNA를 함유하는 적격 세포, 예를 들어, DH10BAC® 적격 세포 내로 형질전환된다. 일부 실시형태에서, 배칼로바이러스 발현 벡터, 예를 들어, pFASTBAC® 벡터는 프로모터, 예를 들어, 이중 프로모터(예를 들어, 폴리헤드린 프로모터, p10 프로모터)를 가질 수 있다. 상업적으로 이용 가능한 pFASTBAC® 공여자 플라스미드는 다음을 포함한다: pFASTBAC 1, pFASTBAC HT 및 pFASTBAC DUAL. 일부 실시형태에서, 재조합 박미드 DNA 함유-콜로니가 식별되고, 박미드 DNA는 곤충 세포를 형질주입시키기 위해 단리된다.

[0345] 일부 실시형태에서, 배칼로바이러스 벡터는 헬퍼 핵산과 함께 곤충 세포 내로 도입된다. 도입은 동시적이거나 순차적일 수 있다. 일부 실시형태에서, 헬퍼 핵산은 예를 들어, 배칼로바이러스 벡터의 패키징을 촉진하기 위해 하나 이상의 배칼로바이러스 단백질을 제공한다. 일부 실시형태에서, (예를 들어, 상동 재조합에 의해) 곤충 세포에서 생성된 재조합 배칼로바이러스는 확장되고, 재조합 단백질 발현을 위해 곤충 세포를 감염시키기 위해(예를 들어, 중간-대수 성장 단계에서) 사용된다. 일부 실시형태에서, 박테리아, 예를 들어, 이. 콜라이에서 부위-특이적 전위에 의해 생성된 재조합 박미드 DNA는 곤충 세포를 형질주입제, 예를 들어, Cellfectin® II로 형질주입시키는 데 사용된다. 배칼로바이러스 발현 시스템에 대한 추가 정보는 본원에 참조로 포함된, 미국 특허 출원 제14/447,341호, 제14/277,892호 및 제12/278,916호에서 논의된다.

[0346] **곤충 세포 시스템**

[0347] 본원에 기재된 단백질은 예를 들어, 위에 기재된 바와 같은 재조합 배칼로바이러스 또는 박미드 DNA로 감염되거나 형질주입된 곤충 세포에서 발현될 수 있다. 일부 실시형태에서, 곤충 세포는 다음을 포함한다: 스포도프테라 프루기페르다(*Spodoptera frugiperda*)로부터 유래된 Sf9 및 Sf21 세포 및 트리코플루시아 니(*Trichoplusia ni*)로부터 유래된 Tn-368 및 High Five™ BTI-TN-5B1-4 세포(Hi5 세포로도 지칭됨). 일부 실시형태에서, 번데기

가을 거염벌레(*army worm*)인 스포도프테라 프루기페르다의 난소로부터 유래된 곤충 세포주 Sf21 및 Sf9는 배칼로바이러스 발현 시스템을 사용하는 재조합 단백질의 발현에 사용될 수 있다. 일부 실시형태에서, Sf21 및 Sf9 곤충 세포는 상업적으로 이용 가능한 혈청-보충 배지 또는 무혈청 배지에서 배양될 수 있다. 곤충 세포 배양에 적합한 배지는 다음을 포함한다: 그레이스 보충물(Grace's Supplemented)(TNM-FH), IPL-41, TC-100, 슈나이더 드로스필라(Schneider's Drosophila), SF-900 II SFM 및 EXPRESS-FIVE™ SFM. 일부 실시형태에서, 일부 무혈청 배지 제형은 배양 및 재조합 단백질 생성 둘 모두를 위해 6.0 내지 6.4 범위의 배양 pH를 유지하기 위한 인산염 완충 시스템을 활용한다(문헌[Licari et al. Insect cell hosts for baculovirus expression vectors contain endogenous exoglycosidase activity. Biotechnology Progress 9: 146-152 (1993) 및 Drugmand et al. Insect cells as factories for biomanufacturing. Biotechnology Advances 30:1140-1157 (2012)]). 일부 실시형태에서, 다양한 곤충 세포주를 배양하기 위해 6.0 내지 6.8의 pH가 사용될 수 있다. 일부 실시형태에서, 곤충 세포는 통기를 사용하여 25°C 내지 30°C의 온도에서 현탁액 내에서 또는 단층으로서 배양된다. 곤충 세포에 대한 추가 정보는 예를 들어, 이들 각각이 본원에 참조로 포함된, 미국 특허 출원 제14/564,512호 및 제14/775,154호에 논의된다.

**[0348] 포유동물 세포 시스템**

**[0349]** 일부 실시형태에서, 본원에 기재된 단백질은 예를 들어, 본원에 기재된 바와 같은 단백질을 인코딩하는 벡터로 감염되거나 형질주입된 동물 세포주에서 시험관 내 발현될 수 있다. 본 개시내용의 맥락에서 예상되는 동물 세포주는 돼지 세포주, 예를 들어, 불멸화된 돼지 세포주, 예컨대, 비제한적으로 돼지 신장 상피 세포주 PK-15 및 SK, 단일골수성(monomyeloid) 세포주 3D4/31 및 정소 세포주 ST를 포함한다. 또한, 다른 포유동물 세포주, 예컨대, CHO 세포(중국 햄스터 난소), MARC-145, MDBK, RK-13, EEL이 포함된다. 추가로 또는 대안적으로, 본 발명의 방법의 특정 실시형태는 상피 세포주, 즉, 상피 계통의 세포의 세포주인 동물 세포주를 사용한다. 본원에 기재된 단백질을 발현하기에 적합한 세포주는 인간 또는 영장류 기원의 세포주, 예컨대, 인간 또는 영장류 신장 암종 세포주를 포함하나, 이에 제한되지 않는다.

**[0350] 이펙터**

**[0351]** 본원에 기재된 조성물 및 방법은 예를 들어, 본원에 기재된 바와 같은 이펙터(예를 들어, 외인성 이펙터 또는 내인성 이펙터)를 인코딩하는 서열을 포함하는 아넬로벡터의 유전 요소를 생성하기 위해 사용될 수 있다. 일부 실시형태에서, 유전 요소는 이펙터이며, 예를 들어, 유전 요소는 기능적 RNA이다. 이펙터는, 일부 예에서, 내인성 이펙터 또는 외인성 이펙터일 수 있다. 일부 실시형태에서, 이펙터는 치료적 이펙터이다. 일부 실시형태에서, 이펙터는 폴리펩티드(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 치료적 폴리펩티드 또는 펩티드)를 포함한다. 일부 실시형태에서, 이펙터는 비-코딩 RNA(예를 들어, miRNA, siRNA, shRNA, mRNA, lncRNA, RNA, DNA, 안티센스 RNA 또는 gRNA)를 포함한다. 일부 실시형태에서, 이펙터는 예를 들어, 본원에 기재된 바와 같은 조절 핵산을 포함한다.

**[0352] 시험관 내 어셈블리 방법**

**[0353]** 아넬로벡터는 예를 들어, 시험관 내 어셈블리에 의해 생성될 수 있다. 일부 실시형태에서, 유전 요소는 어셈블리를 허용하는 조건 하에서 시험관 내에서 ORF1과 접촉된다.

**[0354]** 일부 실시형태에서, 배칼로바이러스 작제물은 아넬로바이러스 단백질을 생성하기 위해 사용된다. 그런 다음, 이들 단백질은 예를 들어, 유전 요소, 예를 들어, RNA를 포함하는 유전 요소를 캡시드화하기 위한 시험관 내 어셈블리에 사용될 수 있다. 일부 실시형태에서, 하나 이상의 아넬로바이러스 단백질을 인코딩하는 폴리뉴클레오티드는 숙주 세포, 예를 들어, 곤충 또는 동물 세포에서의 발현을 위해 프로모터에 융합된다. 일부 실시형태에서, 폴리뉴클레오티드는 배칼로바이러스 발현 시스템 내로 클로닝된다. 일부 실시형태에서, 숙주 세포, 예를 들어, 곤충 세포는 배칼로바이러스 발현 시스템으로 감염되고 일정 기간 동안 항온처리된다. 일부 실시형태에서, 감염된 세포는 약 1, 2, 3, 4, 5, 10, 15 또는 20일 동안 항온처리된다. 일부 실시형태에서, 감염된 세포는 아넬로바이러스 단백질을 회수하기 위해 용해된다.

**[0355]** 일부 실시형태에서, 단리된 아넬로바이러스 단백질은 정제된다. 일부 실시형태에서, 아넬로바이러스 단백질은 비제한적으로 킬레이팅 정제, 헤파린 정제, 구배 침강 정제 및/또는 SEC 정제를 포함하는 정제 기술을 사용하여 정제된다. 일부 실시형태에서, 정제된 아넬로바이러스 단백질은 유전 요소, 예를 들어, RNA를 포함하는 유전 요소를 캡시드화하기 위해 유전 요소와 혼합된다. 일부 실시형태에서, 유전 요소는 ORF1 단백질, ORF2 단백질 또는 그의 변형된 버전을 사용하여 캡시드화된다. 일부 실시형태에서 2개의 핵산이 캡시드화된다. 예를 들어, 제1

핵산은 mRNA, 예를 들어, 화학적으로 변형된 mRNA일 수 있고, 제2 핵산은 DNA일 수 있다.

[0356] 일부 실시형태에서, 아넬로바이러스(AV) ORF1(예를 들어, 야생형 ORF1 단백질, 예를 들어, 어셈블리 효율, 수율 또는 안정성을 개선하기 위한 돌연변이를 보유하는 ORF1 단백질, 키메라 ORF1 단백질 또는 그의 단편)을 인코딩하는 DNA는 곤충 세포주(예를 들어, Sf9 및/또는 HighFive), 동물 세포주(예를 들어, 닭 세포주(MDCC)), 박테리아 세포(예를 들어, 이. 콜라이) 및/또는 포유동물 세포주(예를 들어, 293expi 및/또는 MOLT4)에서 발현된다. 일부 실시형태에서, AV ORF1을 인코딩하는 DNA는 태깅되지 않을 수 있다. 일부 실시형태에서, AV ORF1을 인코딩하는 DNA는 N-말단 및/또는 C-말단에 융합된 태그를 함유할 수 있다. 일부 실시형태에서, AV ORF1을 인코딩하는 DNA는, 예를 들어, 면역염색 검정(ELISA 또는 웨스턴 블롯을 포함하나 이에 제한되지 않음)을 통한 예를 들어, 정제 및/또는 동일성 결정을 돕기 위한 태그를 도입하기 위해 ORF1 단백질 내에 돌연변이, 삽입 또는 결실을 보유할 수 있다. 일부 실시형태에서, AV ORF1을 인코딩하는 DNA는 단독으로 또는 임의의 수의 헬퍼 단백질과 조합하여 발현될 수 있다. 일부 실시형태에서, AV ORF1을 인코딩하는 DNA는 AV ORF2 및/또는 ORF3 단백질과 조합하여 발현된다.

[0357] 일부 실시형태에서, 어셈블리 효율을 개선시키기 위한 돌연변이를 보유하는 ORF1 단백질은 입자 어셈블리를 촉발하도록 pH 민감성 핵산 결합을 허용하는 ARG 아암의 pI를 변경하기 위해 N-말단 아르기닌 아암(ARG 아암) 내로 도입된 돌연변이를 보유하는 ORF1 단백질을 포함할 수 있으나, 이에 제한되지 않는다(SEQ ID 3 내지 5). 일부 실시형태에서, 안정성을 개선시키는 돌연변이를 보유하는 ORF1 단백질은 프로토머 표면의 소수성 상태를 변경하고 캡시드 형성의 열역학적 우호성을 개선시키기 위해 표준 젤리를 베타-배럴의 베타 가닥 F 및 G와 접촉하는 인터프로토머에 대한 돌연변이를 포함할 수 있다.

[0358] 일부 실시형태에서, 키메라 ORF1 단백질은 다른 캡시드 단백질, 예를 들어, 부리 및 깃털 질병 바이러스(BFDV) 캡시드 단백질, 또는 E형 간염 캡시드 단백질로부터의 비슷한 부분으로 교체된 ORF1 단백질의 서열의 일부 또는 일부들을 갖는 ORF1 단백질, 예를 들어, BFDV 캡시드 단백질로부터의 비슷한 구성성분으로 교체된 고리 9 ORF1의 ARG 아암 또는 F 및 G 베타 가닥을 포함할 수 있으나, 이에 제한되지 않는다. 일부 실시형태에서, 키메라 ORF1 단백질은 다른 AV ORF1 단백질의 비슷한 부분으로 교체된 ORF1 단백질의 서열의 일부 또는 일부들(예를 들어, 고리 9 ORF1의 비슷한 부분으로 교체된 고리 2 ORF1의 젤리를 단편 또는 C-말단 부분)을 갖는 ORF1 단백질을 포함할 수도 있다.

[0359] 일부 실시형태에서, 본 개시내용은 아넬로백터를 제조하는 방법을 설명하며, 방법은 (a) (i) RNA를 포함하는 유전 요소, 및 (ii) ORF1 분자를 포함하는 혼합물을 제공하는 단계; 및 (b) ORF1 분자를 포함하는 단백질성 외부 내에 유전 요소를 봉입하기에 적합한 조건 하에서 혼합물을 향온처리하며, 이에 따라 아넬로백터를 제조하는 단계를 포함하며; 선택적으로, 혼합물은 세포에 포함되지 않는다. 일부 실시형태에서, 방법은 (a)를 제공하기 전에, 예를 들어, 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)에서 ORF1 분자를 발현시키는 단계를 추가로 포함한다. 일부 실시형태에서, 발현은 ORF1 분자를 생성하기에 적합한 조건 하에서 ORF1 분자를 인코딩하는 핵산 분자(예를 들어, 배컬로바이러스 발현 벡터)를 포함하는 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)를 향온처리하는 것을 포함한다. 일부 실시형태에서, 방법은 (a)를 제공하기 전에, 숙주 세포에 의해 발현된 ORF1 분자를 정제하는 단계를 추가로 포함한다. 일부 실시형태에서, 방법은 무세포 시스템에서 수행된다. 일부 실시형태에서, 본 개시내용은 다음을 포함하는, 아넬로백터 조성물의 제조 방법을 설명한다: (a) 선행하는 실시형태 중 어느 하나에 따른 복수의 아넬로백터 또는 조성물을 제공하는 단계; (b) 선택적으로, 다음 중 하나 이상에 대하여 복수를 평가하는 단계: 본원에 기재된 오염물질, 광학 밀도 측정치(예를 들어, OD 260), (예를 들어, HPLC에 의한) 입자 개수, 감염성(예를 들어, 형광 및/또는 ELISA에 의해 결정되는 바와 같은, 예를 들어, 입자:감염성 유닛 비); 및 (c) 예를 들어, (b)의 파라미터 중 하나 이상이 명시된 임계값을 만족한다면, 복수의 아넬로백터를 예를 들어, 대상체로의 투여에 적합한 약제학적 조성물로서 제형화하는 단계.

[0360] **농축 및 정제**

[0361] 수확된 아넬로백터는 예를 들어, 아넬로백터 제제를 생성하기 위해 정제 및/또는 농축될 수 있다. 일부 실시형태에서, 수확된 아넬로백터는 예를 들어, 바이러스 입자를 정제하기 위해 당업계에 알려진 방법(예를 들어, 침강, 크로마토그래피 및/또는 한외여과에 의한 정제)을 사용하여 수확 용액에 존재하는 다른 구성요소(constituent) 또는 오염물질로부터 분리된다. 일부 실시형태에서, 수확된 아넬로백터는 친화도 정제(예를 들어, 헤파린 친화도 정제)에 의해 정제된다. 일부 실시형태에서, 수확된 아넬로백터는 크기 배제 크로마토그래피(예를 들어, Tris 완충액 이동상을 사용함)에 의해 정제된다. 일부 실시형태에서, 수확된 아넬로백터는 음이온 교환 크로마토그래피(예를 들어, Mustang Q 막 크로마토그래피)에 의해 정제된다. 일부 실시형태에서, 수확

된 아넬로바이러스는 혼합 모드 크로마토그래피(예를 들어, 혼합 모드 수지, 예를 들어, Cato700 수지를 사용함)에 의해 정제된다. 일부 실시형태에서, 정제 단계는 제제로부터 혈청, 숙주 세포 DNA, 숙주 세포 단백질, 유전 요소가 결합된 입자 및/또는 페놀 레드 중 하나 이상을 제거하는 단계를 포함한다. 일부 실시형태에서, 수확된 아넬로바이러스는, 예를 들어, 바이러스 입자를 농축시키기 위한 당업계에 알려진 방법을 사용하여, 수확 용액에 존재하는 다른 구성요소 또는 오염물질에 비하여 농축된다.

[0362] 일부 실시형태에서, 생성된 제제 또는 제제를 포함하는 약제학적 조성물은 허용 가능한 기간 및 온도에 걸쳐 안정할 것이고/것이거나, 요망되는 투여 경로 및/또는 이러한 투여 경로가 필요로 할 임의의 디바이스, 예를 들어, 니들 또는 주사기와 양립 가능할 것이다.

[0363] **II. 아넬로바이러스**

[0364] 일부 양태에서, 본 개시내용은 아넬로바이러스를 사용 및 제조하는 조성물 및 방법, 아넬로바이러스 제제, 및 치료적 조성물을 제공한다. 일부 실시형태에서, 아넬로바이러스는 아넬로바이러스(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스), 또는 그의 단편 또는 부분, 또는 다른 실질적으로 비-병원성인 바이러스, 예를 들어, 상리공생 바이러스(symbiotic virus), 편리공생 바이러스(commensal virus), 고유 바이러스에 기반한 서열, 구조 및/또는 기능을 포함하는 하나 이상의 핵산 또는 폴리펩티드를 포함한다. 일부 실시형태에서, 아넬로바이러스-기반의 아넬로바이러스는 아넬로바이러스에 대하여 외인성인 적어도 하나의 요소, 예를 들어, 외인성 이펙터 또는 아넬로바이러스의 유전 요소 내에 배치된 외인성 이펙터를 인코딩하는 핵산 서열을 포함한다. 일부 실시형태에서, 아넬로바이러스-기반의 아넬로바이러스는 아넬로바이러스 유래의 또 다른 요소에 대하여 이종인 적어도 하나의 요소, 예를 들어, 또 다른 연결된 핵산 서열, 예컨대 프로모터 요소에 대하여 이종인 이펙터-인코딩 핵산 서열을 포함한다. 일부 실시형태에서, 아넬로바이러스는 유전 요소의 나머지 및/또는 단백질성 외부에 비하여 이종인 적어도 하나의 요소(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 이펙터를 인코딩하는 외인성 요소)를 포함하는 유전 요소(예를 들어, 환형 DNA, 예를 들어, 단일 가닥 DNA)를 포함한다. 아넬로바이러스는 페이로드를 위한 숙주, 예를 들어, 인간 내로의 운반 비히클(예를 들어, 실질적으로 비병원성 운반 비히클)일 수 있다. 일부 실시형태에서, 아넬로바이러스는 진핵 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포에서 복제 가능하다. 일부 실시형태에서, 아넬로바이러스는 포유동물(예를 들어, 인간) 세포에서 실질적으로 비병원성 및/또는 실질적으로 비통합성이다. 일부 실시형태에서, 아넬로바이러스는 포유동물, 예를 들어, 인간에서 실질적으로 비-면역원성이다. 일부 실시형태에서, 아넬로바이러스는 복제-결핍이다. 일부 실시형태에서, 아넬로바이러스는 복제-적격이다.

[0365] 일부 실시형태에서, 아넬로바이러스는 예를 들어, 그의 전문이 본원에 참조로 포함되는 PCT 출원 제 PCT/US2018/037379호에 기재된 바와 같은, 큐론 또는 그의 구성성분(예를 들어, 이펙터를 인코딩하는 서열, 및/또는 단백질성 외부)를 포함하는, 예를 들어, 유전 요소)을 포함한다. 일부 실시형태에서, 아넬로바이러스는 예를 들어, 그의 전문이 본원에 참조로 포함되는 PCT 출원 제 PCT/US19/65995호에 기재된 바와 같은 아넬로바이러스 또는 그의 구성성분(예를 들어, 이펙터를 인코딩하는 서열, 및/또는 단백질성 외부)를 포함하는, 예를 들어, 유전 요소)을 포함한다.

[0366] 일 양태에서, 본 발명은 (i) 프로모터 요소, 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터, 예를 들어, 페이로드)를 인코딩하는 서열 및 단백질 결합 서열(예를 들어, 외부 단백질 결합 서열, 예를 들어, 패키징 신호)을 포함하는 유전 요소로서, 유전 요소가 단일-가닥 DNA이며, 다음의 특성 중 하나 또는 둘 모두를 갖는 유전 요소: 환형이고/환형이거나 세포에 유입되는 유전 요소의 약 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5% 또는 2% 미만의 빈도로 진핵 세포의 게놈 내로 통합됨; 및 (ii) 단백질성 외부를 포함하는 아넬로바이러스를 포함하며; 유전 요소는 단백질성 외부 내에 봉입되며; 아넬로바이러스는 유전 요소를 진핵 세포 내로 운반할 수 있다.

[0367] 본원에 기재된 아넬로바이러스의 일부 실시형태에서, 유전 요소는 세포에 유입되는 유전 요소의 약 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5% 또는 2% 미만의 빈도로 통합된다. 일부 실시형태에서, 대상체에 투여되는 복수의 아넬로바이러스로부터 약 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4% 또는 5% 미만의 유전 요소는 대상체 내의 하나 이상의 숙주 세포의 게놈 내로 통합될 것이다. 일부 실시형태에서, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스의 집단의 유전 요소는 유사한 AAV 바이러스의 집단의 것보다 더 낮은 빈도로, 예를 들어, 유사한 AAV 바이러스의 집단보다 약 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 100% 이상 더 낮은 빈도로 숙주 세포의 게놈 내로 통합된다.

[0368] 일 양태에서, 본 발명은 (i) 프로모터 요소 및 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터, 예를 들어, 페이로드)를 인코딩하는 서열, 및 단백질 결합 서열(예를 들어, 외부 단백질 결합 서열)을 포함하는 유전

요소로서, 유전 요소가 야생형 아넬로바이러스 서열(예를 들어, 야생형 토르크 테노 바이러스(TTV), 토르크 테노 미니 바이러스(TTMV) 또는 TTMDV 서열, 예를 들어, 본원에 기재된 바와 같은 야생형 아넬로바이러스 서열)에 대하여 적어도 75%(예를 들어, 적어도 75, 76, 77, 78, 79, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100%) 서열 동일성을 갖는 유전 요소; 및 (ii) 단백질성 외부를 포함하는 아넬로백터를 포함하며; 유전 요소는 단백질성 외부 내에 봉입되며; 아넬로백터는 유전 요소를 진핵 세포 내로 운반할 수 있다.

- [0369] 일 양태에서, 본 발명은
- [0370] a) (i) 외부 단백질(예를 들어, 비-병원성 외부 단백질)을 인코딩하는 서열, (ii) 유전 요소를 비-병원성 외부 단백질에 결합시키는 외부 단백질 결합 서열, 및 (iii) 이펙터(예를 들어, 내인성 또는 외인성 이펙터)를 인코딩하는 서열을 포함하는 유전 요소; 및
- [0371] b) 유전 요소와 회합된, 예를 들어, 이를 둘러싸거나 봉입하는 단백질성 외부를 포함하는 아넬로백터를 포함한다.
- [0372] 일부 실시형태에서, 아넬로백터는 비-외피, 환형, 단일-가닥 DNA 바이러스로부터의(또는 이와 70% 초과, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100%의 상동성을 갖는) 서열 또는 발현 생성물을 포함한다. 동물 환형 단일-가닥 DNA 바이러스는 일반적으로 진핵 비-식물 숙주를 감염시키고, 환형 계놈을 갖는 단일 가닥 DNA(ssDNA) 바이러스의 하위군을 지칭한다. 따라서, 동물 환형 ssDNA 바이러스는 원핵생물을 감염시키는 ssDNA 바이러스(즉, 마이크로바이러스과 및 이노바이러스과) 및 식물을 감염시키는 ssDNA 바이러스(즉, 체미니바이러스과 및 나노바이러스과)와 구별 가능하다. 그들은 또한 비-식물 진핵생물을 감염시키는 선형 ssDNA 바이러스(즉, 파보바이러스과)와 구별 가능하다.
- [0373] 일부 실시형태에서, 아넬로백터는 숙주 세포 기능을 예를 들어, 일시적으로 또는 장기간 조절한다. 특정 실시형태에서, 세포 기능은 안정하게 변경되며, 예컨대 조절은 적어도 약 1시간 내지 약 30일, 또는 적어도 약 2시간, 6시간, 12시간, 18시간, 24시간, 2일, 3일, 4일, 5일, 6일, 7일, 8일, 9일, 10일, 11일, 12일, 13일, 14일, 15일, 16일, 17일, 18일, 19일, 20일, 21일, 22일, 23일, 24일, 25일, 26일, 27일, 28일, 29일, 30일, 60일 이상 또는 그 사이의 임의의 시간 동안 지속된다. 특정 실시형태에서, 세포 기능은 일시적으로 변경되며, 예를 들어, 예컨대 조절은 약 30분 이하 내지 약 7일 또는 약 1시간 이하, 2시간, 3시간, 4시간, 5시간, 6시간, 7시간, 8시간, 9시간, 10시간, 11시간, 12시간, 13시간, 14시간, 15시간, 16시간, 17시간, 18시간, 19시간, 20시간, 21시간, 22시간, 24시간, 36시간, 48시간, 60시간, 72시간, 4일, 5일, 6일, 7일 또는 그 사이의 임의의 시간 동안 지속된다.
- [0374] 일부 실시형태에서, 유전 요소는 프로모터 요소를 포함한다. 실시형태들에서, 프로모터 요소는 RNA 중합효소 II-의존성 프로모터, RNA 중합효소 III-의존성 프로모터, PGK 프로모터, CMV 프로모터, EF-1 $\alpha$  프로모터, SV40 프로모터, CAGG 프로모터 또는 UBC 프로모터, TTV 바이러스 프로모터, 조직 특이적, U6(po11III), 활성화 단백질(TetR-VP16, Gal4-VP16, dCas9-VP16 등)에 대한 상류 DNA 결합 부위가 있는 최소 CMV 프로모터로부터 선택된다. 실시형태들에서, 프로모터 요소는 TATA 박스를 포함한다. 실시형태들에서, 프로모터 요소는 예를 들어, 본원에 기재된 바와 같은 야생형 아넬로바이러스에 대하여 내인성이다.
- [0375] 일부 실시형태에서, 유전 요소는 다음의 특징 중 하나 이상을 포함한다: 단일-가닥, 환형, 음성 가닥 및/또는 DNA. 실시형태들에서, 유전 요소는 에피솜을 포함한다. 일부 실시형태에서, 이펙터를 배제한 유전 요소의 부분은 약 2.5 내지 5 kb(예를 들어, 약 2.8 내지 4 kb, 약 2.8 내지 3.2 kb, 약 3.6 내지 3.9 kb 또는 약 2.8 내지 2.9 kb), 약 5 kb 미만(예를 들어, 약 2.9 kb, 3.2 kb, 3.6 kb, 3.9 kb 또는 4 kb 미만) 또는 적어도 100개의 뉴클레오티드(예를 들어, 적어도 1 kb)의 조합된 크기를 갖는다.
- [0376] 일부 실시형태에서, 아넬로백터 또는 아넬로백터에 포함되는 유전 요소는 세포(예를 들어, 인간 세포) 내로 도입된다. 일부 실시형태에서, 예를 들어, 일단 아넬로백터 또는 유전 요소가 세포 내로 도입되면, 예를 들어, 아넬로백터의 유전 요소에 의해 인코딩되는 이펙터(예를 들어, RNA, 예를 들어, miRNA)는 세포(예를 들어, 인간 세포)에서 발현된다. 실시형태들에서, 세포 내로의 아넬로백터 또는 거기에 포함된 유전 요소의 도입은 예를 들어, 세포에 의한 표적 분자의 발현 수준을 변경시킴으로써 세포에서 표적 분자(예를 들어, 표적 핵산, 예를 들어, RNA 또는 표적 폴리펩티드)의 수준을 조절한다(예를 들어, 증가시키거나 감소시킨다). 실시형태들에서, 아넬로백터 또는 거기에 포함된 유전 요소의 도입은 세포에 의해 생성되는 인터페론의 수준을 감소시킨다. 실시형태들에서, 세포 내로의 아넬로백터 또는 거기에 포함된 유전 요소의 도입은 세포의 기능을 조절한다(예를 들어, 증가시키거나 감소시킨다). 실시형태들에서, 세포 내로의 아넬로백터 또는 거기에 포함된 유전 요소의 도입은

세포의 생존력을 조절한다(예를 들어, 증가시키거나 감소시킨다). 실시형태들에서, 세포 내로의 아넬로바이러스 또는 거기에 포함된 유전 요소의 도입은 세포(예를 들어, 암 세포)의 생존력을 감소시킨다.

[0377] 일부 실시형태에서, 본원에 기재된 아넬로바이러스(예를 들어, 합성 아넬로바이러스)는 70% 미만의 항체 출현율(예를 들어, 약 60%, 50%, 40%, 30%, 20% 또는 10% 미만의 항체 출현율)을 유도한다. 실시형태들에서, 항체 출현율은 해당 분야에 알려진 방법에 따라 결정된다. 실시형태들에서, 항체 출현율은 예를 들어, 문헌[Tsuda et al., 1999; *J. Virol. Methods* 77: 199-206](본원에 참조로 포함됨)에 기재된 항-TTV 항체 검출 방법 및/또는 문헌[Kakkola et al., 2008; *Virology* 382: 182-189](본원에 참조로 포함됨)에 기재된 항-TTV IgG 혈청학적 출현율의 결정 방법에 따라 생물학적 시료에서 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음) 또는 그에 기초한 아넬로바이러스에 대한 항체를 검출함으로써 결정된다. 아넬로바이러스 또는 그에 기초한 아넬로바이러스에 대한 항체는 또한 항-바이러스 항체를 검출하기 위한 해당 분야의 방법, 예를 들어, 문헌[Calcedo et al., 2013; *Front. Immunol.* 4(341): 1-7](본원에 참조로 포함됨)에 기재된 바와 같은 예를 들어, 항-AAV 항체의 검출 방법에 의해 검출될 수 있다.

[0378] 일부 실시형태에서, 복제 결핍, 복제 결함 또는 복제 비적격 유전 요소는 유전 요소의 복제에 필요한 필수 기구 또는 구성성분의 전부를 인코딩하지 않는다. 일부 실시형태에서, 복제 결함 유전 요소는 복제 인자를 인코딩하지 않는다. 일부 실시형태에서, 복제 결함 유전 요소는 하나 이상의 ORF(예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 및/또는 ORF2t/3)를 인코딩하지 않는다. 일부 실시형태에서, 유전 요소에 의해 인코딩되지 않는 기구 또는 구성성분은 (예를 들어, 숙주 세포에 의해 포함되는, 예를 들어, 숙주 세포의 게놈 내로 통합되는 핵산에 인코딩되어) 트랜스로 제공되어, 예를 들어, 유전 요소가 트랜스로 제공되는 기구 또는 구성성분의 존재 하에 복제를 겪을 수 있게 할 수 있다.

[0379] 일부 실시형태에서, 패키징 결핍, 패키징 결함 또는 패키징 비적격 유전 요소는 단백질성 외부(예를 들어, 단백질성 외부는 예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함하는, 캡시드 또는 그의 부분을 포함함) 내에 패키징될 수 없다. 일부 실시형태에서, 패키징 결핍 유전 요소는 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스에 비하여 10% 미만(예를 들어, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01% 또는 0.001% 미만)의 효율로 단백질성 외부 내로 패키징된다. 일부 실시형태에서, 패키징 결함 유전 요소는 심지어 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스의 유전 요소의 패키징을 허용할 인자(예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 또는 ORF2t/3)의 존재 하에서도 단백질성 외부 내에 패키징될 수 없다. 일부 실시형태에서, 패키징 결핍 유전 요소는 심지어 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스의 유전 요소의 패키징을 허용할 인자(예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 또는 ORF2t/3)의 존재 하에서도, (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스에 비하여 10% 미만(예를 들어, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01% 또는 0.001% 미만)의 효율로 단백질성 외부 내로 패키징된다.

[0380] 일부 실시형태에서, 패키징 적격 유전 요소는 단백질성 외부(예를 들어, 단백질성 외부는 예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함하는, 캡시드 또는 그의 부분을 포함함) 내로 패키징될 수 있다. 일부 실시형태에서, 패키징 적격 유전 요소는 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스에 비하여 적어도 20%(예를 들어, 적어도 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100% 이상)의 효율로 단백질성 외부 내에 패키징된다. 일부 실시형태에서, 패키징 적격 유전 요소는 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스의 유전 요소의 패키징을 허용할 인자(예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 또는 ORF2t/3)의 존재 하에서, 단백질성 외부 내로 패키징될 수 있다. 일부 실시형태에서, 패키징 적격 유전 요소는 (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스의 유전 요소의 패키징을 허용할 인자(예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 또는 ORF2t/3)의 존재 하에서, (예를 들어, 본원에 기재된 바와 같은) 야생형 아넬로바이러스에 비하여 적어도 20%(예를 들어, 적어도 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100% 이상)의 효율로 단백질성 외부 내로 패키징된다.

[0381] **아넬로바이러스**

[0382] 일부 실시형태에서, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스는 아넬로바이러스로부터 유래된 서열 또는 발현 생성물을 포함한다. 일부 실시형태에서, 아넬로바이러스는 아넬로바이러스에 비하여 외인성인 하나 이상의 서열 또는 발현 생성물을 포함한다. 일부 실시형태에서, 아넬로바이러스는 아넬로바이러스에 비하여 내인성인 하나 이상의 서열 또는 발현 생성물을 포함한다. 일부 실시형태에서, 아넬로바이러스는 아넬로바이러스 내의 하나 이상의 다른

서열 또는 발현 생성물에 비하여 이중인 하나 이상의 서열 또는 발현 생성물을 포함한다. 아넬로바이러스는 일반적으로 음의 극성을 갖는 단일-가닥 환형 DNA 계통을 갖는다.

- [0383] 일부 실시형태에서, 유전 요소는 아미노산 서열 또는 그의 기능적 단편, 또는 본원에 기재된 아미노산 서열, 예를 들어, 아넬로바이러스 아미노산 서열 중 어느 하나에 대하여 적어도 약 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 인코딩하는 뉴클레오티드 서열을 포함한다.
- [0384] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로바이러스 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 서열 또는 그의 단편에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 하나 이상의 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 포함한다.
- [0385] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로바이러스 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스의 TATA 박스, 캡 부위, 개시자 요소, 전사 시작 부위, 5' UTR 보존된 도메인, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, 3개의 오픈-리딩 프레임 영역, 폴리(A) 신호, GC-풍부 영역, 또는 이들의 임의의 조합 중 하나 이상에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 하나 이상의 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 포함한다. 일부 실시형태에서, 핵산 분자는 캡시드 단백질을 인코딩하는 서열, 예를 들어, 본원에 기재된 아넬로바이러스 중 임의의 것의 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 서열을 포함한다. 실시형태들에서, 핵산 분자는 아넬로바이러스 ORF1 단백질(또는 그의 스플라이스 변이체 또는 기능적 단편) 또는 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 캡시드 단백질을 인코딩하는 서열을 포함한다.
- [0386] 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF1 핵산 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF1/1 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF1/2 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF2 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF2/2 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF2/3 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 ORF2t/3 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 TATA 박스 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 개시자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 5' UTR 보존된 도메인 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스의 3개의 오픈-리딩 프레임 영역의 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 폴리(A) 신호 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 A1의 아넬로바이러스 GC-풍부 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다.
- [0387] 실시형태들에서, 핵산 분자는 표 B1의 아넬로바이러스 ORF1 핵산 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분





어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 E1의 아넬로바이러스 폴리(A) 신호 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 핵산 분자는 표 E1의 아넬로바이러스 GC-풍부 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다.

[0391] 일부 실시형태에서, 유전 요소는 아미노산 서열 또는 그의 기능적 단편, 또는 본원에 기재된 아미노산 서열 중 어느 하나, 예를 들어, 아넬로바이러스 아미노산 서열에 대하여 적어도 약 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 인코딩하는 뉴클레오티드 서열을 포함한다.

[0392] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로바이러스, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 서열 또는 그의 단편에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 하나 이상의 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 포함한다. 실시형태들에서, 아넬로바이러스는 표 A1 내지 M2 중 임의의 것에 나타낸 바와 같은 서열 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열로부터 선택되는 핵산 서열을 포함한다. 실시형태들에서, 아넬로바이러스는 표 A2 내지 M2 중 임의의 것에 나타낸 바와 같은 서열 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드를 포함한다.

[0393] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로바이러스는 본원에 기재된 아넬로바이러스(예를 들어, 표 A 내지 M 중 임의의 것에 주석이 달리거나, 그에 열거된 서열에 의해 인코딩되는 바와 같은 아넬로바이러스 서열) 중 임의의 것의 TATA 박스, 캡 부위, 개시인자 요소, 전사 시작 부위, 5' UTR 보존된 도메인, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, 3개의 오픈-리딩 프레임 영역, 폴리(A) 신호, GC-풍부 영역 중 하나 이상 또는 그의 임의의 조합에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 하나 이상의 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 포함한다. 일부 실시형태에서, 핵산 분자는 캡시드 단백질을 인코딩하는 서열, 예를 들어, 본원에 기재된 아넬로바이러스(예를 들어, 표 A 내지 M 중 임의의 것에 주석이 달리거나, 그에 열거된 서열에 의해 인코딩되는 바와 같은 아넬로바이러스 서열) 중 임의의 것의 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 서열을 포함한다. 실시형태들에서, 핵산 분자는 아넬로바이러스 ORF1 또는 ORF2 단백질(예를 들어, 표 A2 내지 M2 중 임의의 것에 나타낸 바와 같은 ORF1 또는 ORF2 아미노산 서열, 또는 표 A1 내지 M1 중 임의의 것에 나타낸 바와 같은 핵산 서열에 의해 인코딩되는 ORF1 또는 ORF2 아미노산 서열)에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 캡시드 단백질을 인코딩하는 서열을 포함한다. 실시형태들에서, 핵산 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 표 A2 내지 M2 중 임의의 것에 나타낸 바와 같은 ORF1 아미노산 서열 또는 표 A1 내지 M1 중 임의의 것에 나타낸 바와 같은 핵산 서열에 의해 인코딩되는 ORF1 아미노산 서열)에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 캡시드 단백질을 인코딩하는 서열을 포함한다.

[0394] 일부 실시형태에서, 본원에 기재된 바와 같은 아넬로바이러스는 키메라 아넬로바이러스이다. 일부 실시형태에서, 키메라 아넬로바이러스는 아넬로바이러스 이외의 바이러스로부터의 하나 이상의 요소, 폴리펩티드 또는 핵산을 추가로 포함한다.

[0395] 실시형태들에서, 키메라 아넬로바이러스는 (예를 들어, 본원에 기재된 바와 같은) 복수의 상이한 아넬로바이러스로부터의 서열을 포함하는 복수의 폴리펩티드(예를 들어, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 및/또는 ORF2t/3)를 포함한다. 예를 들어, 키메라 아넬로바이러스는 하나의 아넬로바이러스로부터의 ORF1 분자(예를 들어, 고리1 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자) 및 상이한 아넬로바이러스로부터의 ORF2 분자(예를 들어, 고리2 ORF2 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2 분자)를 포함할 수 있다. 또 다른 예에서, 키메라 아넬로바이러스는 하나의 아넬로바이러스로부터의 제1 ORF1 분자(예를 들어, 고리1 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자) 및 상이한 아넬로바이러스로부터의 제2 ORF1 분자(예를 들어, 고리2 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자)를 포함할 수 있다.

[0396] 일부 실시형태에서, 아넬로백터는 예를 들어, 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 적어도 하나의 부분 및 상이한 바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 적어도 하나의 부분을 포함하는 키메라 폴리펩티드(예를 들어, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 및/또는 ORF2t/3)를 포함한다.

[0397] 일부 실시형태에서, 아넬로백터는 예를 들어, 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 적어도 하나의 부분 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 적어도 하나의 부분을 포함하는 키메라 폴리펩티드(예를 들어, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 및/또는 ORF2t/3)를 포함한다. 실시형태들에서, 아넬로백터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자의 적어도 하나의 부분을 포함하는 키메라 ORF1 분자를 포함한다. 실시형태들에서, 키메라 ORF1 분자는 하나의 아넬로바이러스로부터의 ORF1 젤리-롤 도메인 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스로부터의 ORF1 아미노산 하위서열(예를 들어, 본원에 기재된 바와 같음) 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 키메라 ORF1 분자는 하나의 아넬로바이러스로부터의 ORF1 아르기닌-풍부 영역 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스로부터의 ORF1 아미노산 하위서열(예를 들어, 본원에 기재된 바와 같음) 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 키메라 ORF1 분자는 하나의 아넬로바이러스로부터의 ORF1 N22 도메인 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스로부터의 ORF1 아미노산 하위서열(예를 들어, 본원에 기재된 바와 같음) 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 키메라 ORF1 분자는 하나의 아넬로바이러스로부터의 ORF1 C-말단 도메인 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스로부터의 ORF1 아미노산 하위서열(예를 들어, 본원에 기재된 바와 같음) 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 아넬로백터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1/1 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1/1 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1/1 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1/1 분자의 적어도 하나의 부분을 포함하는 키메라 ORF1/1 분자를 포함한다. 실시형태들에서, 아넬로백터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1/2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1/2 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF1/2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF1/2 분자의 적어도 하나의 부분을 포함하는 키메라 ORF1/2 분자를 포함한다. 실시형태들에서, 아넬로백터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2 분자의 적어도 하나의 부분을 포함하는 키메라 ORF2 분자를 포함한다. 실시형태들에서, 아넬로백터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2/2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2/2 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2/2 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2/2 분자의 적어도 하나의 부분을 포함하는 키메라 ORF2/2 분자를 포함한다. 실시형

태들에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2/3 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2/3 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2/3 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2/3 분자의 적어도 하나의 부분을 포함하는 키메라 ORF2/3 분자를 포함한다. 실시형태들에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2T/3 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2T/3 분자의 적어도 하나의 부분, 및 상이한 아넬로바이러스(예를 들어, 본원에 기재된 바와 같음)로부터의 ORF2T/3 분자 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 아미노산 서열 동일성을 갖는 ORF2T/3 분자의 적어도 하나의 부분을 포함하는 키메라 ORF2T/3 분자를 포함한다.

[0398] 본원에 포함된 서열 또는 하위서열이 본원에 기재된 조성물 및 방법에서 활용될 수 있는 추가의 예시적인 아넬로바이러스 게놈은 예를 들어, PCT 출원 PCT/US2018/037379호 및 PCT/US19/65995호(그의 전문이 본원에 참조로 포함됨)에 기재되어 있다. 일부 실시형태에서, 예시적인 아넬로바이러스 서열은 본원에 참조로 포함된 PCT/US19/65995호의 표 A1, A3, A5, A7, A9, A11, B1 내지 B5, 1, 3, 5, 7, 9, 11, 13, 15 또는 17 중 임의의 것에 열거된 바와 같은 핵산 서열을 포함한다. 일부 실시형태에서, 예시적인 아넬로바이러스 서열은 본원에 참조로 포함된 PCT/US19/65995호의 표 A2, A4, A6, A8, A10, A12, C1 내지 C5, 2, 4, 6, 8, 10, 12, 14, 16 또는 18 중 임의의 것에 열거된 바와 같은 아미노산 서열을 포함한다. 일부 실시형태에서, 예시적인 아넬로바이러스 서열은 예를 들어, 본원에 참조로 포함된 PCT/US19/65995호의 표 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10 또는 37A 내지 37C 중 임의의 것에 열거된 바와 같은, ORF1 분자 서열, 또는 ORF1 분자 서열을 인코딩하는 핵산 서열을 포함한다.

[0399] 표 A1. 예시적인 아넬로바이러스 핵산 서열(알파토크크바이러스, 클레이드 3)

명칭	고리 1
속/클레이드	알파토크크바이러스, 클레이드 3
수탁 번호	AJ620231.1
전체 서열: 3753 bp	
	1            10            20            30            40            50
	TGCTACGTCACCTAACCCACGTGCTCCTACAGGCCAATCGCAGTCTATGT
	CGTGCACCTTCCTGGGCATGGTCTACATAATTATATAAATGCTTGCACTTC
	CGAATGGCTGAGTTTTTGCTGCCCGTCCGCGGAGAGAGCCACGGCAGGG
	GATCCGAACGTCCTGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGAAG
	TCAAGGGGCAATTCGGGCTCAGGACTGGCCGGGCTTTGGGCAAGGCTCTT
	AAAAATGCACTTTTCTCGAATAAGCAGAAAGAAAAGGAAAGTGTACTGCT
	TTTGCGTGCCAGCAGCTAAGAAAAAACAACCTGCTATGAGCTTCTGGAAA
	CCTCCGGTACACAATGTACCGGGATCCAACGCATGTGGTATGAGTCCTT
	TCACCGTGGCCACGCTTCTTTTGTGGTTGTGGGAATCTATACTTCACA
	TTACTGCACCTTGCTGAAACATATGGCCATCCAACAGGCCCGAGACCTTCT
	GGGCCACCGGGAGTAGACCCCAACCCACATCCGTAGAGCCAGGCCTGC
	CCCGGCCGCTCCGGAGCCCTCACAGGTTGATTGAGACCAGCCCTGACAT
	GGCATGGGGATGGTGAAGCGACGGAGGCGCTGGTGGTTCCGGAAGCGGT
	GGACCCGTGGCAGACTTCGCAGACGATGGCCTCGATCAGCTCGTCCCGC
	CCTAGACGACGAAGAGTAAGGAGGCGCAGACGGTGGAGGAGGGGAGACG
	AAAAACAAGGACTTACAGACGCAGGAGACGCTTTAGACGCAGGGGACGAA

[0400]

AAGCAAACTTATAATAAACTGTGGCAACCTGCAGTAATTTAAAAGATGC  
AGAATAAAGGGATACATACCCTGATTATAAGTGGGAACGGTACCTTTGC  
CACAACTTTACCAGTCACATAAATGACAGAATAATGAAAGGCCCTTCG  
GGGAGGACACAGCCTATGAGGTTGAGCCTCTACATTTTGTGAGGAG  
CACCTCAGACACATGAACTTCTGGACCAGAAGCAACGATAACCTAGAGCT  
AACCAGATACTTGGGGCTTCAGTAAAAATATACAGGCACCCAGACCAAG  
ACTTTATAGTAATATACAACAGAAGAACCCTCTAGGAGGCAACATCTAC  
ACAGCACCTCTCTACACCCAGGCAATGCCATTTTAGCAAAACACAAAAT  
ATTAGTACCAAGTTTACAGACAAGACCAAGGGTAGAAAAGCAATTAGAC  
TAAGAATAGCACCCCCACACTCTTTACAGACAAGTGGTACTTTCAAAG  
GACATAGCCGACCTCACCTTTTCAACATCATGGCAGTTGAGGCTGACTT  
GCGGTTTCCGTTCTGCTCACCAAACTGACAACACTGCATCAGCTTCC  
AGGTCCTTAGTTCGTTTACAACACTACCTCAGTATTAATACCTTTAAT  
AATGACAACCTCAGACTCAAAGTTAAAAGAATTTTAAATAAAGCATTTC  
AACCAAGGCACAAAAGGAACAAGTTTAAATGCACTAAATACATTTAGAA  
CAGAAGGATGCATAAGTCACCCACAACATAAAAAAACCAACCACAAATA  
AACAAACCATTAGAGTCACAATACTTTGCACCTTAGATGCCCTCTGGGG  
AGACCCCATATACTATAATGATCTAAATGAAAACAAAAGTTTGAACGATA  
TCATTGAGAAAATACTAATAAAAAACATGATTACATACCATGCAAAACTA  
AGAGAATTTCCAAATTCATACCAAGGAAACAAGGCCTTTTGCACCTAAC  
AGGCATATACAGCCACCATACTAAACCAAGGCAGAAATATCTCCAGAAA  
TATTTGGACTGTACACAGAAATAATTTACAACCCTTACACAGACAAAGGA  
ACTGGAACAAAGTATGGATGGACCCACTAACTAAAGAGAACAACATATA  
TAAAGAAGGACAGAGCAAATGCCTACTGACTGACATGCCCTATGGACTT  
TACTTTTGGATATACAGACTGGTGTAAAAGGACACTAATAACTGGGAC  
TTACCACTAACTACAGACTAGTACTAATATGCCCTTATACCTTTCCAAA  
ATTGTACAATGAAAAAGTAAAAGACTATGGGTACATCCCGTACTCTACA  
AATTCGGAGCGGGTCAGATGCCAGACGGCAGCAACTACATACCCTTTCAG  
TTTAGAGCAAAGTGGTACCCACAGTACTACACCAGCAACAGGTAATGGA  
GGACATAAGCAGGAGCGGGCCCTTTGCACCTAAGGTAGAAAACCAAGCA  
CTCAGCTGGTAATGAAGTACTGTTTTAACTTTAACTGGGGCGGTAACCCT  
ATCATTGAACAGATTGTTAAAGACCCAGCTTCCAGCCACCTATGAAAT  
ACCCGGTACCGGTAACATCCCTAGAAGAATACAAGTCATCGACCCGCGGG  
TCCTGGGACCGCACTACTCGTTCGGTTCATGGGACATGCGCAGACACACA  
TTTAGCAGAGCAAGTATTAAGAGAGTGTGAGAACAACAAGAACTTCTGA  
CCTTGTATTCTCAGGCCAAAAAGCCTCGGGTCGACATCCCAAAACAAG  
AAACCAAGAAGAAAGCTCACATTCACTCAAAGAGAATCGAGACCGTGG  
GAGACCGAGGAAGAAAGCGAGACAGAAGCCCTCTCGCAAGAGAGCCAAGA

[0401]

GGTCCCCTTCCAACAGCAGTTGCAGCAGCAGTACCAAGAGCAGCTCAAGC  
 TCAGACAGGGAATCAAAGTCTCTTCGAGCAGCTCATAAGGACCCAACAA  
 GGGGTCCATGTAAACCCATGCCTACGGTAGGTCCCAGGCAGTGGCTGTTT  
 CCAGAGAGAAAGCCAGCCCCAGCTCCTAGCAGTGGAGACTGGCCATGGA  
 GTTCTCGCAGCAAAAATATTTGATAGGCCAGTTAGAAGCAACCTTAAAG  
 ATACCCCTTACTACCCATATGTTAAAAACCAATACAATGTCTACTTTGAC  
 CTTAAATTTGAATAAACAGCAGCTTCAAACCTTGCAAGGCCGTGGGAGTTT  
 CACTGGTCCGGTGTCTACCTCTAAAGGTCACTAAGCACTCCGAGCGTAAGC  
 GAGGAGTGCAGCCCTCCCCCTGGAACAACCTTCTTCGGAGTCCGGCGCTA  
 CGCCTTCGGCTGCGCCGACACCTCAGACCCCCCTCCACCCGAAACGCT  
 TCGCGTTTCGGACCTTCGGCGTCGGGGGGTTCGGGAGCTTTATTAACG  
 GACTCCGAAGTGCTCTTGACACTGAGGGGTGAACAGCAACGAAAGTGA  
 GTGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTCCATTTGTCAGT  
 GTCCGGGTCGCCATAGGCTTCGGGCTCGTTTTTAGGCCTCCGGACTAC  
 AAAAAATCGCCATTTTGGTGACGTCACGGCCGCCATCTTAAGTAGTTGAGG  
 CGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAA  
 TGGTGGACAATTTCTTCGGGTCAAAGGTTACAGCCGCCATGTTAAACA  
 CGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACT  
 TCCTTCTCTTTTTCAAAAAAAGCGGAAGTGCCGCCGCGGGCGGGGG  
 GCGGCGGCTGCGCGCCGCCAGTAGGGGAGCCATGCGCCCCCCCC  
 GCGCATGCGGGGGCCCCCCCCCGGGGGGCTCCGCCCCCCGGCCCCC  
 CCG (SEQ ID NO: 16)

**주석:**

<i>추정되는 도메인</i>	<i>염기 범위</i>
TATA 박스	83 – 88
캡 부위	104 – 111
전사 시작 부위	111
5' UTR 보존된 도메인	170 – 240
ORF2	336 – 719
ORF2/2	336 – 715 ; 2363 – 2789
ORF2/3	336 – 715 ; 2565 – 3015
ORF2i/3	336 – 388 ; 2565 – 3015
ORF1	599 – 2830
ORF1/1	599 – 715 ; 2363 – 2830
ORF1/2	599 – 715 ; 2565 – 2789
3 개의 오픈-리딩 프레임 영역	2551 – 2786
폴리(A) 신호	3011 – 3016
GC-풍부 영역	3632 – 3753

[0402]

[0403]

[0404] 표 A2. 예시적인 아델로바이러스 아미노산 서열(알파토르크바이러스, 클레이드 3)

고리 1 (알파토르크바이러스 클레이드 3)	
ORF2	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAAPAEPSQVDSRPALTWHDGGSDDGGAGGSGSGGPVADFA DDGLDQLVAALDDEE (SEQ ID NO: 17)
ORF2/2	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAAPAEPSQVDSRPALTWHDGGSDDGGAGGSGSGGPVADFA DDGLDQLVAALDDEELLKTPASSPPMKYPVPTSLEEYKSSSTRGSWDRTRSGHGT CADTHLAEQVLRQCQNNKLLTLYSQAQKSLGSTSQNKKPKKKAHIHNSKENRDRG RPRKKARQKPSRKRAKRSNSSSCSSTKSSSSSDRESKSSSSS (SEQ ID NO: 18)
ORF2/3	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAAPAEPSQVDSRPALTWHDGGSDDGGAGGSGSGGPVADFA DDGLDQLVAALDDEEPKASGRHPKTRNPRRKLTFPKRIETVGDGRKRDRSPLA REPRGLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQWLF ERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPIYPYVKNQYNVYFDLKFE (SEQ ID NO: 19)
ORF2t/3	MSFWKPPVHNVTGIQRMWPKASGRHPKTRNPRRKLTFPKRIETVGDGRKRDR SPLAREPRGLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQ WLFPERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPIYPYVKNQYNVYFDLK FE (SEQ ID NO: 20)
ORF1	MAWGWWKRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRWRGRRK TRTYRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISNGTFATNFTSHINDR IMKGPFGGGHSTMRFSLYLFEHLRHMNFWTRSDNLELTRYLGASVKIYRHPDQ DFIVYNNRRTPLGGNIYTAAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFT DKWYFQKDIADLTLFNIMAVEADLRFPCSPQTDNTCISFQVLSVYNNYLSINTFN NDNSDKLKEFLNKAFTTGTGKTSLNALNTRTEGCISHPQLKKNPQINKPLESQ YFAPLDALWGDPYYNDLNNKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAF HLTGIYSPYLNQGRISPEIFGLYTEIYNPYTDKGTGNKVVMDPLTKENNIYKEGQS KCLLDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDY

[0405]

	GYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSRSGPFAPKVEKP STQLVMKYCFNFWGNGNPIEQVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSF RSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESHSLQRESR PWETEESETEALSQESQEVFPQQQLQQYQEQLKLRQGKIKVLFEQLIRTQQGVHV NPCLR (SEQ ID NO: 21)
ORF1/1	MAWGWWKRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRIVKDPSPQTYEIPG TGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPR VDIPKQETQEESHSLQRESRPWETEESETEALSQESQEVFPQQQLQQYQEQLKL RQGKIKVLFEQLIRTQQGVHVNPCLR (SEQ ID NO: 22)
ORF1/2	MAWGWWKRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRAQKSLGSTSQNKK PKKKAHIHNSKENRDRGRPRKKARQKPSRKRAKRSNSSSCSSTKSSSSSDRESKSS SSS (SEQ ID NO: 23)

[0406]

[0407] 표 B1. 예시적인 아넬로바이러스 핵산 서열(베타토르크바이러스)

명칭	고리 2
속/클레이드	<i>베타토르크바이러스</i>
수탁 번호	JX134045.1

전체 서열: 2797 bp

1	10	20	30	40	50
TAATAAATATTC AACAGGAAAACCACCTAATTAAATTGCCGACCACAAA					
CCGTCACTTAGTTCCCCTTTTGGCAACAACCTCTGCTTTTTCCAACTGC					
CGGAAAACCACATAATTTGCATGGCTAACCCAAAACCTGATATGCTAATTA					
ACTTCCACAAAACAACCTCCCCTTTAAAACCCACACCTACAAATTAATTA					
TTAAACACAGTCACATCCTGGGAGGTA CTACCACACTATAATACCAAGTG					
CACTTCCGAATGGCTGAGTTTATGCCGCTAGACGGAGAACGCATCAGTTA					
CTGACTGCGGACTGAACTGGGCGGGTGCCGAAGGTGAGTGAAACCACCG					
AAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAGAAA					
CTTAAATTTATTTTATTTTTCAGATGAGCGACTGCTTTAAACCAACATGC					
TACAACAACAAAACAAGCAAACCTCACTGGATTAATAACCTGCATTTAAC					
CCACGACCTGATCTGCTTCTGCCCAACACCAACTAGACACTTATTACTAG					
CTTTAGCAGAACAACAAGAAACAATTGAAGTGTCTAAACAAGAAAAAGAA					
AAAATAACAAGATGCCTTATTACTACAGAAGAAGACGGTACAACACTACAGA					

[0408]

CGTCCTAGATGGTATGGACGAGGTTGGATTAGACGCCCTTTTCGCAGAAG  
 ATTTCGAAGAAAAAGAAGGGTAAGACCTACTTATACTACTATTTCCTCTAA  
 AGCAATGGCAACCGCCATATAAAAGAACATGCTATATAAAAGGACAAGAC  
 TGTTTAATACTATAGCAACTTAAGACTGGGAATGAATAGTACAATGTA  
 TGAAAAAGTATTGTACCTGTACATTGGCCGGGAGGGGTTCTTTTCTG  
 TAAGCATGTTAACTTTAGATGCCTTGATGATATACATAAACTTTGTAGA  
 AACTGGTGGACATCCACAAACCAAGACTTACCCTAGTAAGATATAAAGG  
 ATGCAAAATAACATTTTATCAAAGCACATTTACAGACTACATAGTAAGAA  
 TACATACAGAACTACCAGCTAACAGTAACAACTAACATACCCAAACACA  
 CATCCACTAATGATGATGATGTCTAAGTACAAACACATTATACCTAGTAG  
 ACAAACAAGAAGAAAAAGAAACCATAACAAAAATATTTGTAACCAC  
 CTCCGCAATTTGAAAACAAATGGTACTTTGCTACAGACCTCTACAAAATT  
 CCATTACTACAAATACACTGCACAGCATGCAACTTACAAAACCCATTTGT  
 AAAACCAGACAAATTATCAAACAATGTTACATTATGGTCACTAAACACCA  
 TAAGCATACAAATAGAAACATGTCAGTGGATCAAGGACAATCATGGCCA  
 TTTAAATACTAGGAACACAAAGCTTTTATTTTACTTTTACACCGGAGC  
 AAACCTACCAGGTGACACAACACAAATACCAGTAGCAGACCTATTACCAC  
 TAACAAACCCAAGAATAAACAGACCAGGACAATCACTAAATGAGGCAAAA  
 ATTACAGACCATATTACTTTACAGAATACAAAAACAAATTTACAAATTA  
 TTGGGGTAACCCATTTAATAAACACATTCAAGAACACCTAGATATGATAC  
 TATACTCACTAAAAAGTCCAGAAGCAATAAAAAACGAATGGACAACAGAA  
 AACATGAAATGGAACCAATTAACAATGCAGGAACAATGGCATTAACACC  
 ATTTAACGAGCCAATATTCACACAAATACAATATAACCCAGATAGAGACA  
 CAGGAGAAGACACTCAATTATACCTACTCTCTAACGCTACAGGAACAGGA  
 TGGGACCCACCAGGAATCCAGAATTAATACTAGAAGGATTTCCACTATG  
 GTTAATATATTGGGGATTTGCAGACTTTCAAAAAACCTAAAAAAGTAA  
 CAAACATAGACACAAATTACATGTTAGTAGCAAAAACAAAATTTACACAA  
 AAACCTGGCACATTCTACTTAGTAATACTAAATGACACCTTTGTAGAAGG  
 CAATAGCCCATATGAAAAACAACCTTTACCTGAAGACAACATTAATGGT  
 ACCCAAGTACAATACCAATTAGAAGCACAAAACAACTACTACAAACT  
 GGGCCATTTACACCAACATACAAGGACAACCTATCAGACAATATATCAAT  
 GTTTTATAAATTTTACTTTAAATGGGGAGGAAGCCACCAAAGCAATTA  
 ATGTTGAAAATCCTGCCACCAGATTCAATATCCCATACCCCGTAACGAG  
 CATGAAAACAACTTCGTTACAGAGTCCAGGGGAAGCCCCAGAATCCATCTT

[0409]

ATACTCCTTCGACTATAGACACGGGAACTACACAACAACAGCTTTGTCCAC  
 GAATTAGCCAAGACTGGGCACTTAAAGACTGTCTTCTAAAATTACAGAG  
 CCAGATCGACAGCAACTGCTCAAACAAGCCCTCGAATGCCTGCAAATCTC  
 GGAAGAAACGCGAGGAGAAAAAGAAAAGAAGTACAGCAGCTCATCAGCA  
 ACCTCAGACAGCAGCAGCAGCTGTACAGAGAGCGAATAATATCATTATTA  
 AAGGACCAATAAAGTTTAACTGTGTAAAAAGGTGAAATGTTTTGATGAT  
 AAACCAAAAAACCGTAGATTTACACCTGAGGAATTTGAAACTGAGTTACA  
 AATAGCAAAATGGTTAAAGAGACCCCCAAGATCCTTTGTAAATGATCCTC  
 CCTTTTACCCATGGTTACCACCTGAACCTGTGTAAACTTTAAGCTTAAT  
 TTTACTGAATAAAGGCCAGCATTAAATTCACTTAAGGAGTCTGTTTATTTA  
 AGTTAAACCTTAATAAACGGTCACCGCCTCCCTAATACGCAGGCGCAGAA  
 AGGGGGCTCCGCCCCCTTAAACCCCAAGGGGCTCCGCCCCCTGAAACCC  
 CCAAGGGGGCTACGCCCCCTTACACCCCC (SEQ ID NO: 54)

**주석:**

<b>추정되는 도메인</b>	<b>염기 범위</b>
TATA 박스	237- 243
캡 부위	260 - 267
전사 시작 부위	267
5' UTR 보존된 도메인	323 - 393
ORF2	424 - 723
ORF2/2	424 - 719 ; 2274 - 2589
ORF2/3	424 - 719 ; 2449 - 2812
ORF1	612 - 2612
ORF1/1	612 - 719 ; 2274 - 2612
ORF1/2	612 - 719 ; 2449 - 2589
3 개의 오픈-리딩 프레임 영역	2441 - 2586
폴리(A) 신호	2808 - 2813
GC-풍부 영역	2868 - 2929

[0410]

[0411] 표 B2. 예시적인 아넬로바이러스 아미노산 서열(베타토르크바이러스)

고리 2 (베타토르크바이러스)	
ORF2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTVDVLDGMDEVGLDALFAEDFEEKEG (SEQ ID NO: 55)
ORF2/2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTVDVLDGMDEVGLDALFAEDFEEKEGFNIPYVTSMKQLRY RVQGGKPNPSYTPSTIDTGTQQQLCHELAKTGHLKTLFLKLQSQIDSNCNSNKPSNA CKSRKKRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 56)
ORF2/3	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTVDVLDGMDEVGLDALFAEDFEEKEGARSTATAQTSRMP ANLGRNAGEKRKRSTAAHQQPQTAAAQVQRANNIIKGPITFNCVKKVCLFDDKPK NRRFTPEEFETELQIAKWLKRPPRSFVNDPPFYPWLPPEPVVNFKLNfte (SEQ ID NO: 57)
ORF1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRFRFRKRRVRPTYTTIPLKQWQPPYKR TCYIKGQDCLIIYSNLRGMLNSTMYEKSIVPVHWPGGGSFSVSMMLTDALYDIHKL CRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHELPAANSNKLTYPNTHPLM MMMSKYKHIIIPSRQTRRKKKPYTKIFVKPPPQFENKWFATDLYKIPLLQIHCTACN LQNPVFKPKLSNNVTLWSLNTISIQNRNMSVDQGSWPFKILGTQSFYFYFYTG NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNK HIQEHLDMILYSLSPEAIKNEWTTENMKWNQLNAGTMALTPFNEPIFTQIQYNP DRDTGEDTQLYLLSNA GTGWDPPIELILEGFPLWLIYWGFAFQKLNKKVTNID TNYMLVAKTKFTQKPGTFYLVLNDTFVEGNSPYEKQPLPEDNIKWYPVQVYQLEA QNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPKAINVENPAHQIQYPIRNE HETTSLSQSPGEAPESILYSFDYRHGNYTTTALSRSQDWALKDTSKITEPDRQQLLK QALECLQISEETQEKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 58)
ORF1/1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRFRFRKRRRIQYPIRNEHETTSLSQSPGE APESILYSFDYRHGNYTTTALSRSQDWALKDTSKITEPDRQQLLKQALECLQISEE TQEKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 59)
ORF1/2	MPYYYRRRRYNYRRPRWYGRGWIRRPFRFRFRKRRSQIDSNCNSNKPSNACKSRK KRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 60)

[0412]

[0413] 표 C1. 예시적인 아넬로바이러스 핵산 서열(감마토르크바이러스)

명칭	고리 4
속/클레이드	감마토르크바이러스
수탁 번호	

[0414]

전체 서열: 3176 bp

```

1           10           20           30           40           50
|           |           |           |           |
TAAAAATGGCGGGAGCCAATCATTTTATACTTTCACTTTCCAATTAAAAAT
GGCCACGTCACAAAACAAGGGGTGGAGCCATTTAAACTATATAACTAAGTG
GGGTGGCGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCATGGGCGGGAGCCGAGGTGAG
TGAAACCACCGAGGTCTAGGGGCAATTCTGGGCTAGGGCAGTCTAGCGGAA
CGGGCAAGAAACTTAAAACAATATTTGTTTTACAGATGGTTAGTATATCC
TCAAGTGATTTTTTTAAGAAAACGAAATTTAATGAGGAGACGCAGAACCA
AGTATGGATGTCTCAAATGCTGACTCTCATGATAATATCTGCAGTTGCT
GGCATCCATTTGCTCACCTTCTTGCTTCCATATTTCTCTCTGGCCACAAA
GATCGTGATCTTACTATTAACCAAATTTCTTCTAAGAGATTATAAAGAAAA
ATGCCATTCTGGTGGAGAAGAAGGAGAAAAATCTGGACCAACAACAGGTT
TAATTACACAAAAGAAGAAGATATAGAAAAGATGGCCAGAAGGCGCC
GCAGAAGAAGACCATACAGACGCCCTGTTTCGCCGCCGCCGTAGAAAACTT
CGAAAGGTAAGAGAAAAAATACTTTAATTGTTAGACAATGGCAACC
AGACAGTATAAGAACTTGTAATAATATAGGACAGTCAGCTATAGTTGTTG
GGGCTGAGGAAAGCAAAATGTAATGTTACTGTTCAATAAGTTAATTAAT
GTGCCCCCAAAAACACCATATGGGGGAGGCTTTGGAGTAGACCAATACAC
ACTGAAATACTTATATGAAGAATACAGATTTGCACAAAACATTTGGACAC
AATCTAATGTACTGAAAGACTTATGCAGATACATAAATGTTAAGCTAATA
TTCTACAGAGACAACAAAACAGACTTTGTCTTTTCTATGACAGAAACCC
ACCTTTTCAACTAACAAAATTTACATACCCAGGAGCACCCACAACAAA
TCATGCTTCAAAAACACCACAAATTCATACATCACAAATGACAAAGCCT
AATGGAAGACTAACAAAAAACTCAAATTAACCTCCTAAACAAATGCT
TTCTAAATGGTTCTTTTCAAACAATCTGTAAATACCCTTTACTATCTC
TTAAAGCTTCTGCACCTAGACCTTAGGCACTCTTACCTAGGCTGCTGTAAT
GAAAATCCACAGGTATTTTTTTTATTATTTAAACCATGGATACTACACAAT
AACAACTGGGGAGCACAATCCTCAACAGCATAACAGACCTAACTCCAAGG
TGACAGACACAACATACTACAGATACAAAAATGACAGAAAAAATATTAAC
ATTAAGCCATGAATACGAAAAAGTATATCATATGAAAACGGTTATTT
TCAATCTAGTTTCTTACAAAACACAGTGCATATATACCAGTGAGCGTGGTG
AAGCCTGTATAGCAGAAAAACCACTAGGAATAGCTATTTACAATCCAGTA
AAAGACAATGGAGATGGTAATATGATATACCTGTAAGCACTCTAGCAAA
CACTTGGGACCAGCCTCCAAAAGACAGTGCATTTTTAATACAAGGAGTAC
CCATATGGCTAGGCTTATTTGGATATTTAGACTACTGTAGACAAATTTAA
GCTGACAAAACATGGCTAGACAGTCATGTAAGTAATTCAAAGTCCTGC
TATTTTTACTTACCAAAATCCAGGAGCAGGCAAATGGTATTTGTTCCACTAT
CACAAAATTTTTATAAATGGCAATGGTCCGTTTAAATCAACCCTACACTG
CTACAAAAGCAAAGTGGTTTCCACAAATACAATACCAACAAGAAATTTAT
TAATAGCTTTGTAGAATCAGGACCAATTTGTTCCCAAATATGCAAATCAA
CTGAAAGCAACTGGGAATAAAATATAAATATGTTTTTACATTTAAGTGG
GGTGGACCACAATTCATGAACCAGAAATGCTGACCCTAGCAACAAGA
GCAGTATGATGTCCCCGATACTTTCTACCAACAATACAATGAAGATC
CAGAAGGACAAGACCCAGATCTCTCATCCATGATTGGGACTACAGACGA

```

[0415]

GGCTTTATTAAGAAAGATCTCTTAAAAGAATGTCAACTTACTTCTCAAC  
 TCATACAGATCAGCAAGCAACTTCAGAGGAAGACATTCACAAAAAGAAAA  
 AGAGAAATGGACCCCACTCACAGTCCCACAAACAAAAGAAGAGGAGACA  
 CTGTCTGTCTCTCTCTCTCTGCAAAAAAGATACCTTCCAAGAAACAGA  
 GACACAAGAAGACCTCCAGCAGCTCATCAAGCAGCAGCAGGAGCAGCAGC  
 TCCTCCTCAAGAGAAACATCCTCCAGCTCATCCACAACTAAAAGAGAAT  
 CAACAAATGCTTCAGCTTACACAGGCATGTTACCTTAACCAGATTTAAA  
 CCTGGATTTGAAGAGCAAACAGAGAGAGAATTAGCAATTATATTTTCATAG  
 GCCCCCTAGAACCTACAAAGAGGACCTTCCATTCTATCCCTGGCTACCAC  
 CTGCACCCCTTGTACAATTTAACCTTAACCTTCAAAGGCTAGGCCAACAAAT  
 GTACACTTAGTAAAGCATGTTTATTAAGCACAACCCCAAAATAAATGT  
 AAAAAATAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCGGCGCT  
 CGCGCGCATGTGCGCCTCTGGCGCAAATCACGCAACGCTCGCGCGCCCGC  
 GTATGTCTCTTTACCACGCACCTAGATTGGGGTGCAGCGCTAGCGCGCG  
 CACCCCAATGCGCCCGCCCTCGTTCGACCCGCTTGCAGGGTTCGACC  
 ACTTCGGGCTCGGGGGGCGCGCCTGCGGCGCTTTTTACTAAACAGACT  
 CCGAGCCGCCATTTGGCCCCCTAAGCTCCGCCCCCTCATGAATATTCAT  
 AAAGGAAACACATAATTAGAATTGCGGACCAAACTGCCATATGCTAA  
 TTAGTTCCCTTTTACAAAGTAAAAGGGGAAGTGAACATAGCCCCACACC  
 CGCAGGGGCAAGCCCCGCACCCCTACGTCACTAACCAAGCCCCCGCCGC  
 CATCTTGGGTGCGGCAGGGCGGGGC (SEQ ID NO: 886)

**주석:**

<i>추정되는 도메인</i>	<i>염기 범위</i>
TATA 박스	87- 93
캡 부위	110 - 117
전사 시작 부위	117
5' UTR 보존된 도메인	185 - 254
ORF2	286 - 660
ORF2/2	286 - 656 ; 1998 - 2442
ORF2/3	286 - 656 ; 2209 - 2641
TAIP	385 - 484
ORF1	501 - 2489
ORF1/1	501 - 656 ; 1998 - 2489
ORF1/2	501 - 656 ; 2209 - 2442
3 개의 오픈-리딩 프레임 영역	2209 - 2439
폴리(A) 신호	2672 - 2678
GC-풍부 영역	3076 - 3176

[0416]

[0417] 표 C2. 예시적인 아벨로바이러스 아미노산 서열(감마토르크바이러스)

고리 4 (감마토르크바이러스)	
ORF2	MVSISSDFFKKTGFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLR DYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPAAEEDHTDALFAAAVENFER (SEQ ID NO: 887)
ORF2/2	MVSISSDFFKKTGFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPAAEEDHTDALF AAAVENFESGVDHNSMNQKLLTLANKSSMSPILSTKQYKLLKIQKDKTPDLSSMIG TTDEALLKDLLKECQLTSQLIQISKQLQRKTFPKRKRRELDPNSSQSHNKKRRHCH VSSLSAKKIPSKKQRHKKTSSSSSSSRSSSSSSRETSSSSSTN (SEQ ID NO: 888)
ORF2/3	MVSISSDFFKKTGFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPAAEEDHTDALF AAAVENFERSASNFRGRHSQKEKENWTPHSPTRRRGDTVMSPLSLQKRYLPRNR DTRRPPAAHQAAAGAAAPPQEKHPPAHPQTKRESTNASASHRHVTLTRFKPGFEEQ TERELAIIFHRPPRTYKEDLPFYPLWPPAPLVQFNLNFKG (SEQ ID NO: 889)
TAIP	MRRRRRTKYGCKLLTLMIISAVAGIHLTLFLLPYFLLATKIVILLTKFF (SEQ ID NO: 890)
ORF1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLKRV KRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPKTPYGGG FGVDQYTLKYLVEEYRFAQNIWTQSNVLKDLCRYINVKLIFYRDNKTDVLSYDRN PPFLTFTYPGAHPQIMLQKHHKFIKQMTKPNRGLTKKLLKIKPPKQMLSKWFFS KQFCKYPLLSLKASALDLRHSYLGCCNENPQVFFYYLNHGYYTITNWGAQSSTAYR PNSKVTDTTYRYKNDKRNINIKSHEYEKSYSYENGYFQSSFLQTQCIYTSERGEACI AEKPLGIAIYNPVKNDGNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWGLFGYLD YCRQIKADKTWLDLHVLIQSPAIFTYPNPGAGKWYCPQSQSFINGNPFNPPTLL QKAKWFPQIQYQEIINSFVESGPFVPKYANQTESNWELKYKVFVTFKWGGPQFHE PEIADPSKQEYDVPDFTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSLKRMSYF STHTDQQATSEEDIPKPKKRRIGPQLTVPQQKEEETLSCLLSLCKKDTFQETQEDL QQLIKQQEQQLLLKRNILQLIHKLKENQMLQLHTGMLP (SEQ ID NO: 891)
ORF1/1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLKRW GGPQFHEPEIADPSKQEYDVPDFTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSL KRMSTYFSTHTDQQATSEEDIPKPKKRRIGPQLTVPQQKEEETLSCLLSLCKKDTFQE

[0418]

	TETQEDLQQLIKQQEQQLLLKRNILQLIHKLKENQMLQLHTGMLP (SEQ ID NO: 892)
ORF1/2	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLKIS KQLQRKTFPKRKR ELDPNSQSHNKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSRSSSSSSRETSSSS STN (SEQ ID NO: 893)

[0419]

[0420] 표 D1. 예시적인 아넬로바이러스 핵산 서열(베타토르크바이러스)

명칭 고리 9  
 속/클레이드 베타토르크바이러스  
 수탁 번호 MH649263.1

전체 서열: 2845 bp

```

1          10          20          30          40          50
|          |          |          |          |          |
TTATTAATATTCAACAGGAAAACACCTAATTTAAATTGCCGACCACAAA
CCGTCACTAACTTCCTTATTTAACATTACTTCCTTTTAAACCAATGAATA
TTCATACAACACATCACACTTCCTGGGAGGAGACATAAACTATATAACT
AACTACACAGACGAATGGCTGAGTTTATGCCGCTAGACGGAGGACGCACA
GCTACTGCTGCGACCTGAACTTGGGCGGGTGCCGAAGGTGAGTGTAAACCA
CCGTAGTCAAGGGGCAATTTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAG
ATTATTAATACAACTTATTTTACAGATGAGCAAACAACATAAAACCAAC
TTTATACAAAGACAAATCATTGGAATTACAATGGCTAAACAACATTTTAA
GCTCTCACGACCTGTGCTGCGGCTGCAACGATCCAGTTTACATTACTG
ATTTTAATTAACAAAACCGGAGAAGCACCTAAACCAGAAGAAGACATTAA
AAATATAAAATGCCTCCTTACTGGCGCAAAAATACTACCGAAGAAGATA
TAGACCTTTCTCCTGGAGAACTAGAAGAATTATTCAAAGAAGAAAAGAT
GGAGATACCGCAAACCAAGAAAACATACTGGAGAAGAAAACCTGCGGGTA
AGAAAACGTTTTTATAAAAGAAAGTTAAAAAAATTTGTACTTAAACAGTT
TCAACCAAAAATTTATAGAAGATGTACAATATTTGGAACAATCTGCCTAT
TTC AAGGCTCTCCAGAAAGAGCCAACAATAATTATATCAACAATCTAC
TCCTACGTACCAGATAAAGAACCGAGGAGGAGGGGATGGACTTTAATAAC
TGAAAGCTTAAGTAGTTTATGGGAAGACTGGGAACATTTAAAAAATGTAT
GGACTCAAAGTAAACGCTGGTTTACCCTTGTAAGATACGGGGGAGTAAAC
TTATACTTTTATCAATCTGCCTTACTGACTATATTGCTCAAGTTTTCAA
CTGTTATCCTATGACAGACACAAAATACACACATGCAGACTCAGCACCAA
ACAGAATGTTATTAACAAAACATGTAATAAGAGTACCTAGCAGAGAAACA
CGCAAAAAAAGAAAGCCATACAAAAGAGTTAGAGTAGGACCTCCTTCTCA
AATGCAAAAACAAATGGTACTTTCAAAGAGACATATGTGAAATACCATTAA
TAATGATTGCAGCCACAGCCGTTGACTTTAGATATCCCTTTTGTGCAAGC
GACTGTGCTAGTAACAACCTTAACTCTAACATGTTTAAACCACTATTGTT
TCAAAACCAAGACTTTGACCACCATCCGATACACAAGGCTACTTTCCAA
AACCTGGAGTATATCTATACTCAACAACAAAGAAGTAACAAGCCAAGTTCT
TCAGACTGTATATACTTAGGAAACACAAAAGACAATCAAGAAGGTAATC
TGCAAGTAGTCTAATGACTCTAAAAACACAAAAATAACAGATTGGGGAA
ATCCATTTGGCATTATATATAGACGGTTCTAAAAAAATATTTTCTTAC
    
```

[0421]

TTTAAACCCCATCACAATTAGACAGCAGCGACTTTGAACACATGACAGA  
 ATTAGCAGAACCAATGTTTATACAAGTTAGATACAACCCAGAAAAGAGACA  
 CAGGACAAGGAACTTAATATACGTAACAGAAAACTTTAGAGGACAACAC  
 TGGGACCCCTCCATCTAGTGACAACCTAAAAATAGATGGATTTCCTTATA  
 TGACATGTGCTGGGGTTTCATAGACTGGATAGAAAAAGTTCATGAAACAG  
 AAAACTTACTTACCAACTACTGCTTCTGTATTAGAAGCAGCGCTTCAAT  
 GAAAAAAAAACAGTTTTTATACCTGTAGATCATTCTTTTAAACAGGTTT  
 TAGCCCATATGAAACTCCAGTTAAATCATCAGACCAAGCTCACTGGCACC  
 CACAATAAGATTTCAAACAAAATCAATAAATGACATTTGTTTAAACAGGC  
 CCCGGTTGTGCTAGGTCCCATATGGCAATTACATGCAGGCAAAAATGAG  
 TTATAAATTTTCATGTAAATGGGGAGGATGTCCAAAACCTTATGAAAAAC  
 CATATGATCCTTGTTCACAGCCCAATTGGACTATTTCCCAATAACCTCAAT  
 GAAACAATACAAATCCAGAATCCAAAACACATGCCCAAAAAGAACTCCA  
 AGAATGGGACTGGCGACGTGATATTGTTACAAAAAAGCTATCGAAAGAA  
 TTAGACAACACACGGAACCTCATGAACTTTGCAAACTCTTACAGGTTCC  
 AAACACAACCCACCGTACACAGACAAACATCACCGTGGACGGACTCAGA  
 AACGGACTCGGAAGAGGAAAAAGACCAACACAAAGAGATCCAGATCCAGC  
 TCAACAAGCTCAGAAAGCATCAACAGCATCTCAAGCAGCAGCTCAAGCAG  
 TACCTGAAACCCCAAAATATAGAATAAGTTGCAAGCAACATAAAAAGTTGAA  
 CTTTTTCTACTAAAAACCTTTTAAAAACAGACGCTTTACTCCTTCTGA  
 AAGAGAAACAGAAAGACAATGTGCTAAAGCTTTTTGTAGACCAGAAAGAC  
 ATTTCTTTTATGATCCTCCTTTTTACCTTACTGTGTACCTGAACCTATT  
 GTAACCTTTGCTTTGGGATATAAAATTTAAGGCCAACAAATTTCACTTAG  
 TGGTGTCTGTTTATTAAGTTTAACTTAATAAGCATACTCCGCCTCCCT  
 ACATTAAGGCGCCAAAAGGGGGCTCCGCCCTTAAACCCCAAGGGGGCT  
 CCGCCCCTTAAACCCCAAGGGGGCTCCGCCCTTACACCCCC (SEQ ID NO:  
 1001)

**주석:**

<i>추정되는 도메인</i>	<i>염기 범위</i>
TATA 박스	142 – 148
개시 요소	162 – 177
전사 시작 부위	172
5' UTR 보존된 도메인	226 – 296
ORF2	328 – 651
ORF2/2	328 – 647; 2121 – 2457
ORF2/3	328 – 647; 2296 – 2680
ORF1	510 – 2477
ORF1/1	510 – 647; 2121 – 2477
ORF1/2	510 – 647; 2296 – 2457
3 개의 오픈-리딩 프레임 영역	2296 – 2454
GC-풍부 영역	2734 – 2845

[0422]

[0423] 표 D2. 예시적인 아델로바이러스 아미노산 서열(베타토르크바이러스)

고리 9 (베타토르크바이러스)	
ORF2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLC CGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGD TANQE KHTGEEN CG (SEQ ID NO: 1002)
ORF2/2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLC CGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGD TANQE KHTGEEN CGPIGLFPITSMKQYKSRIQTHAHKQNSKNGTGDVILLQK KLSKELDNTRNL MKLCKSLQVPNTTHQYTDKHHRGRTQKRTRKRK KTKHKRSRSSTSSSESIN SISSSSSSST (SEQ ID NO: 1003)
ORF2/3	MSKQLKPTLYKDKSLELQWLNNIFSSHDLC CGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGD TANQE KHTGEEN CGFQTQPTSTQTNITVDGLRNLGRGKRPNTRDPDPAQQAQKASTASQAAA QAVPETPKYRIVASNIKVELFP TKKPFKNRRFTPSETEREQCAKAF CRPERH FFYDPPFPYPYCVPEPIVNFALGYKI (SEQ ID NO: 1004)
ORF1	MPPYWRQKY YRRRYRPFSWRTRRIIQR RKRWR YRKPRKTYWRRKLRVRKR FYKRK LKKIVLKQFQPKIIRRCTIFGTICLFQGS PERANN NYIQTIYSYVPDKE PGGGGWLITELSSSLWEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSA YTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPSRETRK RKRKYK RVRVGPSPQMKNWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNLT LT CLNPLLFQNDQDFDHPSTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDN QEGKSASSMLTKTQKITDWGNPFWHYYIDGSKKIFS YFKPPS QLDSSDFEH MTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFP LYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEKKT V FIPVDHSFLTGF S PYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNMQAKMSYK FHVKGWGGCPKTYEKPYPDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQ EWD WRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEE KDQTQEIQIQLNKL RKHQQHLKQQLKQYLKPQIE (SEQ ID NO: 1005)
ORF1/1	MPPYWRQKY YRRRYRPFSWRTRRIIQR RKRWR YRKPRKTYWRRKLRPNW TIPHNLNETIQIQNPNTCPQTELQ EWDWRRDIVTKKAIERIRQHTEPHETLQIS TGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKL RKHQQHLKQQL KQYLKPQIE (SEQ ID NO: 1006)
ORF1/2	MPPYWRQKY YRRRYRPFSWRTRRIIQR RKRWR YRKPRKTYWRRKLRVPNT THQYTDKHHRGRTQKRTRKRK KTKHKRSRSSTSSSESINSISSSSSSST (SEQ ID NO: 1007)

[0424]

[0425]

[0426] 표 E1. 예시적인 아넬로바이러스 핵산 서열 (베타토르크바이러스)

명칭	고리 10
속/클레이드	베타토르크바이러스
수탁 번호	JX134044.1

전체 서열: 2912 bp

```

1         10        20        30        40        50
|         |         |         |         |         |
TAATAAATATTCAACAGGAAAACACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCTCTTTTTCCACAACCTCCTCTTTTACTAATGAATA
TTCATGTAATTAATTAATAATCACCGTAATTCGGGGGAGGAGCCTTTAAA
CTATAAAACTAACTACACATTCGAATGGCTGAGTTTATGCCGCCAGACGG
AGACGGGATCACTTCAGTGACTCCAGGCTGATCAAGGGCGGGTGCCGAAG
GTGAGTGAAACCACCGTAGTCAAGGGGCAATTGGGCTAGATCAGTCTGG
CGAAACGGGCAAGAACTTAAATGTACTTTATTTACAGAAATGTCAA
ATCTCCAACATACTTAACAACATAAGGCAAAAACAATGCCTTAATCAACT
GCTTCGTTGGAGACCACGATCTTCTGTGCAGCTGTAACAATCCTGCCTAC
CATTGCCTCCAAACTACTTGCAACTACCTTAGCACCTCAACTAAAAAAGA
AGAAAAACAACAATAATAACAATGCCTTGGTGGTACAGACGCCGTAGCTA
CAACCCGTGGAGACGAAGAAATTGGTTTAGAAGACCTAGAAAACTATTT
ACAGAAGATACAGAAGAAGACGCCGCTGGGTAAGAAGAAAACCTTTTAC
AAACGTAATAATTAAGAGACTAAATATAGTAGAATGGCAACCTAAATCAAT
TAGAAAATGTAGAATAAAGGAATGCTATGCTTGTTCAAAACGACAGAAG
ACAGACTGTCATATAACTTTGATATGTATGAAGAGTCTATTATACCAGAA
AACTGCCGGGAGGGGGGGATTAGCATTAGAATATAAGCTTATATGC
CTTATACCAGAACACATACATGCACACAACATTTTACACACACAAACA
CAGACAGACCACAGCAAGATACACAGGCTGTCTTTAAATTTCTACCAA
AGCAAAGACATAGACTACGTAGTAACATATTCTACATCACTCCCACTAAG
AAGCTCAATGGGAATGTACAACCTCCATGCAACCATCCATACATCTAATGC
AACAAAACAACCTAATTTGTACCAAGCAAAACAAAACAAAAAAGAGAAAA
CCATATATTAAAAAACATATATCACCCACAACAATGAATTTCTCAATG
GTACTTTCAACATAACATTTGCAAAACATACCGCTACTAATGATAAGAACA
CAGCATTAAACATTAGATAATTACTATATAGGAAGCAGACAATTAAGTACA
AATGTCACATATACATACCTTAAACACAACATACATCCAAAACAGAGACTG
GGGAGACAGAAATAAACTTACTACTGCCAAAACATTAGGAACACAAAGAT
ACTTCTATATGGAACACATTTCACTGCACAAAATATTAATGACATAAAG
CTACAAGAATAATACCTTTAACAAACACACAAGACTATGTACAAGGCTT
TGATTTGGACAGAAAAAGACAACATAACATAACAACCTACAAGAATTTCT
TAACTAAAGGAGCAGGAAATCCATTTTACGCGAAGTGGATAACAGCACAA
AACCCAGTAATACACACAGCAACAGTCTACACAAATAGAACAAATATA
CACCGCTTCAACAACAACATTTCAAACAACAAACTAACAGACCTACCAA
CGCCAGGATATATATTTATACTCAACAGTAAGCTTAAGATACAACCCA
TACAAGACCTAGCAGAAAAGAAACAAATGCTACTTTTGTAAAGAAGCAAAAT
AAATGCACACGGGTGGGACCCAGAACAACCAAGAATTAATAAACAGTG

```

[0427]

ACCTACCACAATGGTTACTATTATTTGGCTACCCAGACTACATAAAAAGA  
 ACACAAAACTTTGCCATTAGTAGACACAAAATTACATACTAGTAGACCACCTG  
 CCCATACACAAAATCCAGAAAAAACACCATTATATACCTTTAAGCACATCAT  
 TTATAGAAGGTAGAAGCCCATACAGTCCTTCAGACACACATGAACCAGAT  
 GAAGAAGACCAAAACAGGTGGTACCCATGCTACCAATATCAACAAGAATC  
 AATAAATTCAAATATGTCTTAGCGGTCCAGGCACACAAAAATACCAAAG  
 GAATAACAGCAGAAGCAAAAGTAAAATATTCCTTTAATTTAAGTGGGGT  
 GGTGACCTACCACCAATGTCTACAATTACAACCCGACAGACCAGCCAAC  
 ATATGTTGTTCCCAATAACTTCAATGAAACAACCTTCGTTACAGAATCCAA  
 CCACCAGACCAGAGCACTTCTGTACTCCTTTGACGAAAGGAGGGGACAA  
 CTTACAGAAAAAGCTACAAAACGCTTGCTTAAAGACTGGGAAACTAAAGA  
 AACTTCTTTATTGTTCTACAGAATACAGATTCGCGGAGCCAACACAAACAC  
 AAGCCCCACAAGAGGACCCGCTCGGAAGAAGAAGAAGAGACCAACCTC  
 TTCGAGCGACTCCTCCGACAGCGAACCAAGCAGCTCCAGCTCAAGCGCAG  
 AATAATACAAACATGAAAGACCTACAAAAATTAGAATAACTAACAGCAA  
 AAACACCGTTTACCTATTTCCACCTGAACAAAAGAACAGAAGACTAACAC  
 CATGGGAAATACAAGAAGCAAAGAAATAGCCAATTTATTTGGCAGACCA  
 CATAGATACTTTTAAAAGACATTCCTTTCTATTGGGATATACCCCCAGA  
 GCCTAAAGTAAACTTTGATTTAAATTTCAATAAAGAAATAAAGGCAAG  
 GCCCATTAACCTCAAAGTCGGTGTCTACCTCTTTAAGTTTAACTTTACTA  
 AACGGACTCCGCCCTCCCTAAATTTGGGCGCCAAAAGGGGGCTCCGCCCCC  
 TTAACCCAGGGGGCTCCGCCCTAAAACCCCAAGGGGGCTACGCC  
 CCTTACACCCCC (SEQ ID NO: 1008)

**주석:**

<i>추정되는 도메인</i>	<i>염기 범위</i>
TATA 박스	152 – 158
개시 요소	172 – 187
전사 시작 부위	182
5' UTR 보존된 도메인	239 – 309
ORF2	343 – 633
ORF2/2	343 – 629; 2196 – 2505
ORF2/3	343 – 629; 2371 – 2734
ORF1	522 – 2540
ORF1/1	522 – 629; 2196 – 2540
ORF1/2	522 – 629; 2371 – 2505
3 개의 오픈-리딩 프레임 영역	2276 – 2502
GC-풍부 영역	2803 – 2912

[0428]

[0429] 표 E2. 예시적인 아넬로바이러스 아미노산 서열(베타토르크바이러스)

고리 10 (베타토르크바이러스)	
ORF2	MFKSPTYLTTKGKNNALINCFVGDHDLSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAG (SEQ ID NO: 1009)
ORF2/2	MFKSPTYLTTKGKNNALINCFVGDHDLSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPI TSMKQLRYRIQPPDQSTCTPLTKGGDNLQKQLQNAQLKTGKLLKLLYCLQ NTDSRSQHKHKPHKTRPRKRRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1010)
ORF2/3	MFKSPTYLTTKGKNNALINCFVGDHDLSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGIQRGANT NTSPTRGPVLRRRREQPLRATPPTANQAAPAQAQNNNTNIEPTKIRITNSKN TVYLFPEQKNRRLTPWEIQEDKEIANLFGRPHRYFLKDIPFYWDIPPEPKVN FDLNFQ (SEQ ID NO: 1011)
ORF1	MPWWYRRRSYNPWRRRNWFRPRKTIYRRYRRRRRWVRRKPFYKRKIKR LNIVEWQPKSIRKCRKIGMLCLFQTEDRLSYNFDMYEESIPEKLPGGGGFSI KNISLYALYQEHIAHNIFHTNTDRPLARYTGCSLKFYQSKDIDYVVTYSTS LPLRSSMGMYNSMQPSIHLMQQNKLVPSKQTQKRRKPYIKKHISPTQMKS QWYFQHNIANIPLLMIRTTALTDNYIGSRQLSTNVTIHTLNTTYIQNRDW GDRNKTYYCQTLGTQRYFLYGTHSTAQNINDIKLQELIPLTNTQDYVQGF WTEKDKHNITTYKEFLTKGAGNPFHAEWITAQNPVIHTANSPTQIEQIYAS TTTTQNKKLTDLPTPGYIFITPTVSLRYNPYKDLAERNKCYFVRSKINAHGW DPEQHQLINSDLQPWLLLFQYDYKRTQNFALVDNYILVDHCPYTNPEK TPFIPLSTSFIEGRSPSPSDTHEPDEEDQNRWYPCYQYQESINSICLSGPGTP KIPKGITAEAKVKYSFNFKWGGDLPPMSTITNPTDQPTYVVPNNFNETTSLQ NPTRPEHFLYSFDERRGQLTEKATKRLKDWETKETSLLSTEYRFAEPTQT QAPQEDPSSEEEEEENLFRLLRQRTKQLQLKRRRIIQLKDLQKLE (SEQ ID NO: 1012)
ORF1/1	MPWWYRRRSYNPWRRRNWFRPRKTIYRRYRRRRRWPTYVVPNNFNETT SLQNPTTRPEHFLYSFDERRGQLTEKATKRLKDWETKETSLLSTEYRFAEP TQTQAPQEDPSSEEEEEENLFRLLRQRTKQLQLKRRRIIQLKDLQKLE
ORF1/2	MPWWYRRRSYNPWRRRNWFRPRKTIYRRYRRRRRWNTDSRSQHKHKPH KTRTRPRKRRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1013)

[0430]

[0431]

[0432]

일부 실시형태에서, 아넬로바이러스는 그의 전문이 본원에 참조로 포함되는 PCT 출원 제PCT/US2018/037379호에 열거된 서열을 포함하는 핵산을 포함한다. 일부 실시형태에서, 아넬로바이러스는 그의 전문이 본원에 참조로 포함되는 PCT 출원 제PCT/US2018/037379호에 열거된 서열을 포함하는 폴리펩티드를 포함한다. 일부 실시형태에서, 아넬로바이러스는 그의 전문이 본원에 참조로 포함되는 PCT 출원 PCT/US19/65995호에 열거된 서열을 포함하는 핵산을 포함한다. 일부 실시형태에서, 아넬로바이러스는 그의 전문이 본원에 참조로 포함되는 PCT 출원 PCT/US19/65995호에 열거된 서열을 포함하는 폴리펩티드를 포함한다.

[0433]

**ORF1 분자**

[0434]

일부 실시형태에서, 아넬로바이러스는 ORF1 분자 및/또는 ORF1 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)의 구조적 특징 및/또는 활성을 갖는 폴리펩티드를 포함한다. 일부 실시형태에서, ORF1 분자는 아넬로바이러스

ORF1 단백질(예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)에 비하여 절단을 포함한다. ORF1 분자는 다른 ORF1 분자에 결합하여, 예를 들어, (예를 들어, 본원에 기재된 바와 같은) 단백질성 외부, 예를 들어, 캡시드를 형성하는 능력이 있을 수 있다. 일부 실시형태에서, 단백질성 외부는 핵산 분자(예를 들어, 본원에 기재된 바와 같은 유전 요소)를 봉입할 수 있다. 일부 실시형태에서, 복수의 ORF1 분자는 예를 들어, 단백질성 외부를 형성하기 위해 다합체를 형성할 수 있다. 일부 실시형태에서, 다합체는 동중다합체일 수 있다. 다른 실시형태에서, 다합체는 이종다합체일 수 있다.

- [0435] ORF1 분자는 일부 실시형태에서, 다음 중 하나 이상을 포함할 수 있다: 아르기닌 풍부 영역, 예를 들어, 적어도 60%의 염기성 잔기(예를 들어, 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 또는 100%의 염기성 잔기; 예를 들어, 60% 내지 90%, 60% 내지 80%, 70% 내지 90% 또는 70 내지 80%의 염기성 잔기)를 갖는 영역을 포함하는 제1 영역, 및 젤리-롤 도메인, 예를 들어, 적어도 6개의 베타 가닥(예를 들어, 4, 5, 6, 7, 8, 9, 10, 11 또는 12개의 베타 가닥)을 포함하는 제2 영역.
- [0436] 일부 실시형태에서, 본원에 기재된 바와 같은 ORF1 분자는 야생형 ORF1 단백질 서열(예를 들어, 본원에 기재된 바와 같은)에 비하여 하나 이상의 라이신에서 히스티딘으로의 돌연변이를 포함한다. 특정 실시형태에서, ORF1 분자는 아르기닌-풍부 영역 및/또는 제1 베타 가닥에서 하나 이상의 라이신에서 히스티딘으로의 돌연변이를 포함한다.
- [0437] *아르기닌-풍부 영역*
- [0438] 아르기닌 풍부 영역은 본원에 기재된 아르기닌-풍부 영역 서열에 대하여 적어도 70%(예를 들어, 적어도 약 70, 80, 90, 95, 96, 97, 98, 99 또는 100%) 서열 동일성을 갖거나, 적어도 60%, 70% 또는 80%의 염기성 잔기(예를 들어, 아르기닌, 라이신 또는 그의 조합)를 포함하는 적어도 약 40개의 아미노산의 서열을 갖는다.
- [0439] *젤리 롤 도메인*
- [0440] 젤리-롤 도메인 또는 영역은 다음의 특징 중 하나 이상(예를 들어, 1, 2 또는 3개)을 포함하는 폴리펩티드(예를 들어, 더 큰 폴리펩티드에 포함된 도메인 또는 영역)를 포함한다(예를 들어, 이로 이루어진다):
- [0441] (i) 젤리-롤 도메인의 적어도 30%(예를 들어, 적어도 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90% 이상)의 아미노산이 하나 이상의 β-시트의 부분임;
- [0442] (ii) 젤리-롤 도메인의 2차 구조는 적어도 4개(예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10, 11 또는 12개)의 β-가닥을 포함함; 및/또는
- [0443] (iii) 젤리-롤 도메인의 3차 구조는 적어도 2개(예를 들어, 적어도 2, 3 또는 4개)의 β-시트를 포함함; 및/또는
- [0444] (iv) 젤리-롤 도메인은 적어도 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1 또는 10:1의 β-시트 대 α-나선의 비율 포함함.
- [0445] 특정 실시형태에서, 젤리-롤 도메인은 2개의 β-시트를 포함한다.
- [0446] 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 약 8개(예를 들어, 4, 5, 6, 7, 8, 9, 10, 11 또는 12개)의 β-가닥을 포함한다. 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 8개의 β-가닥을 포함한다. 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 7개의 β-가닥을 포함한다. 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 6개의 β-가닥을 포함한다. 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 5개의 β-가닥을 포함한다. 특정 실시형태에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 β-시트는 4개의 β-가닥을 포함한다.
- [0447] 일부 실시형태에서, 젤리-롤 도메인은 제2 β-시트에 대한 역평행 배향의 제1 β-시트를 포함한다. 특정 실시형태에서, 제1 β-시트는 약 4개(예를 들어, 3, 4, 5 또는 6개)의 β-가닥을 포함한다. 특정 실시형태에서, 제2 β-시트는 약 4개(예를 들어, 3, 4, 5 또는 6개)의 β-가닥을 포함한다. 실시형태들에서, 제1 및 제2 β-시트는 총 약 8개(예를 들어, 6, 7, 8, 9, 10, 11 또는 12개)의 β-가닥을 포함한다.
- [0448] 특정 실시형태에서, 젤리-롤 도메인은 캡시드 단백질(예를 들어, 본원에 기재된 바와 같은 ORF1 분자)의 구성성분이다. 특정 실시형태에서, 젤리-롤 도메인은 자가-어셈블리 활성을 갖는다. 일부 실시형태에서, 젤리-롤 도메인

인을 포함하는 폴리펩티드는 젤리-롤 도메인을 포함하는 또 다른 카피의 폴리펩티드에 결합한다. 일부 실시형태에서, 제1 폴리펩티드의 젤리-롤 도메인은 제2 카피의 폴리펩티드의 젤리-롤 도메인에 결합한다.

[0449] *N22 도메인*

[0450] ORF1 분자는 또한, 아넬로바이러스 N22 도메인(예를 들어, 본원에 기재된 바와 같음, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 N22 도메인)의 구조 또는 활성을 포함하는 제3 영역, 및/또는 아넬로바이러스 C-말단 도메인(CTD)(예를 들어, 본원에 기재된 바와 같음, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 CTD)의 구조 또는 활성을 포함하는 제4 영역을 포함할 수 있다. 일부 실시형태에서, ORF1 분자는 N-말단에서 C-말단 순서로, 제1, 제2, 제3 및 제4 영역을 포함한다.

[0451] *초가변 영역(HVR)*

[0452] ORF1 분자는 일부 실시형태에서, 초가변 영역(HVR), 예를 들어, 본원에 기재된 바와 같음, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 HVR을 추가로 포함할 수 있다. 일부 실시형태에서, HVR은 제2 영역과 제3 영역 사이에 위치한다. 일부 실시형태에서, HVR은 적어도 약 55개(예를 들어, 적어도 약 45, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 또는 65개)의 아미노산(예를 들어, 약 45 내지 160, 50 내지 160, 55 내지 160, 60 내지 160, 45 내지 150, 50 내지 150, 55 내지 150, 60 내지 150, 45 내지 140, 50 내지 140, 55 내지 140 또는 60 내지 140개의 아미노산)을 포함한다.

[0453] *예시적인 ORF1 서열*

[0454] 예시적인 아넬로바이러스 ORF1 아미노산 서열 및 예시적인 ORF1 도메인의 서열이 하기 표에 제공된다. 일부 실시형태에서, 본원에 기재된 폴리펩티드(예를 들어, ORF1 분자)는 예를 들어, 표 N 내지 Z 중 임의의 것에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF1 하위서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, 본원에 기재된 아넬로바이러스는 예를 들어, 표 N 내지 Z 중 임의의 것에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF1 하위서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 ORF1 분자를 포함한다. 일부 실시형태에서, 본원에 기재된 아넬로바이러스는 예를 들어, 표 N 내지 Z 중 임의의 것에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF1 하위서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 ORF1 분자를 인코딩하는 핵산 분자(예를 들어, 유전 요소)를 포함한다.

[0455] 일부 실시형태에서, 하나 이상의 아넬로바이러스 ORF1 하위서열은 아르기닌(Arg)-풍부 도메인, 젤리-롤 도메인, 초가변 영역(HVR), N22 도메인 또는 C-말단 도메인(CTD)(예를 들어, 표 N 내지 Z 중 임의의 것에 기재된 바와 같음), 또는 이에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열 중 하나 이상을 포함한다. 일부 실시형태에서, ORF1 분자는 상이한 아넬로바이러스로부터의 복수의 하위서열(예를 들어, 표 N 내지 Z에 열거된 알파토르크바이러스 클레이드 1 내지 7 하위서열로부터 선택된 ORF1 하위서열의 임의의 조합)을 포함한다. 실시형태들에서, ORF1 분자는 하나의 아넬로바이러스로부터의 Arg-풍부 도메인, 젤리-롤 도메인, N22 도메인, 및 CTD 그리고 다른 아넬로바이러스로부터의 HVR 중 하나 이상을 포함한다. 실시형태들에서, ORF1 분자는 하나의 아넬로바이러스로부터의 젤리-롤 도메인, HVR, N22 도메인 및 CTD, 및 다른 아넬로바이러스로부터의 Arg-풍부 도메인 중 하나 이상을 포함한다. 실시형태들에서, ORF1 분자는 하나의 아넬로바이러스로부터의 Arg-풍부 도메인, HVR, N22 도메인 및 CTD, 및 다른 아넬로바이러스로부터의 젤리-롤 도메인 중 하나 이상을 포함한다. 실시형태들에서, ORF1 분자는 하나의 아넬로바이러스로부터의 Arg-풍부 도메인, 젤리-롤 도메인, HVR 및 CTD, 및 다른 아넬로바이러스로부터의 N22 도메인 중 하나 이상을 포함한다. 실시형태들에서, ORF1 분자는 하나의 아넬로바이러스로부터의 Arg-풍부 도메인, 젤리-롤 도메인, HVR 및 N22 도메인, 및 다른 아넬로바이러스로부터의 CTD 중 하나 이상을 포함한다.

[0456] ORF1 분자, 또는 그의 스플라이스 변이체 또는 기능적 단편이 (예를 들어, 유전 요소를 봉입함으로써, 예를 들어, 아넬로바이러스의 단백질성 외부를 형성하기 위한) 본원에 기재된 조성물 및 방법에 활용될 수 있는 추가의 예시적인 아넬로바이러스는 예를 들어, PCT 출원 PCT/US2018/037379호 및 PCT/US19/65995호(그의 전문이 본원에 참조로 포함됨)에 기재되어 있다.

[0457] 표 N. 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스, 클레이드 3)

명칭	고리 1
속/클레이드	알파토르크바이러스, 클레이드 3
수탁 번호	AJ620231.1
단백질 수탁 번호	CAF05750.1
전체 서열: 743 AA	

```

1      10      20      30      40      50
|      |      |      |      |      |
MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRGR
RKTRTYRRRRRFRRRGRKAKLI IKLWQPAVIKRCRIKGYIPLIISGNGTF
    
```

[0458]

```

ATNFTSHINDRIMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSDNLE
LTRYLGASVKIYRHPDQDFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHK
ILVPSLQTRPKGRKAIRLRIAPPTLFTDKWYFQKDIADLTLFNIMAVEAD
LRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFNNDNSDSKLEFLNKAF
PTTGTKGTSLNALNFRTEGCISHPQLKKPNPQINKPLESQYFAPLDALW
GDPIYYNDLNNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFCHL
TGIYSPPYLNQGRISPEIFGLYTEI IYNPYTDKGTGNKVWMDPLTKENNI
YKEGQSKCLLTDMPWLWTLFGYTDWCKKDTNNWDLPLNYRLVLI CPYTFP
KLYNEKVKDYGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQVM
EDISRSGPFAPKVEKPSTQLVMKYCFNFWGGNPIIEQIVKDPSPQPTYE
IPGTGNI PRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETS
DLVFGPKKPRVDIPKQETQEESHSLQRESRPWETEESETEALSQESQ
EVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCLR
(SEQ ID NO: 185)
    
```

주석:

추정되는 도메인	AA 범위
Arg-풍부 영역	1 - 68
젤리-롤 도메인	69 - 280
초가변 영역	281 - 413
N22	414 - 579
C-말단 도메인	580 - 743

[0459]

[0460] 표 O. 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스, 클레이드 3)

고리 1 ORF1 (알파토르크바이러스 클레이드 3)	
Arg-풍부 영역	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRR GRRKTRTYRRRRFRRRGRK (SEQ ID NO: 186)
젤리-롤 도메인	AKLIKLVQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDRIMKGPFGGGHST MRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQDFIVIYNRR TPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFTDKWY FQKDIADLTLFNIMAVEADLRFPCSPQTDNTCISFQVLSSVYNNYLSI (SEQ ID NO: 187)
초가변 도메인	NTFNNDNSDKLKEFLNKAFTTGTGKGTSLNALNTFRTEGCISHPQLKKPNPQI NKPLESQYFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFP NSYQGNKAFCHLTGIYSPPYLNQGR (SEQ ID NO: 188)
N22	ISPEIFGLYTEIINYNYTDKGTGNKVVMDPLTKENNIYKEGQSKCLLTDMPW TLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDYGYIPYSYK FGAGQMPDGSNYIPFQRAK WYPTVLHQQVMEDISRSGPFAPKVEKPSTQL VMKYCFNFN (SEQ ID NO: 189)
C-말단 도메인	WGGNPIIEQIVKDPFQPTYEIPGTGNIPRIQVIDPRVLGPHYSFRSWDMRRHT FSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQESSHSLQRESRPWETEEE SETEALSQESQEVFPQQQLQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCR R (SEQ ID NO: 190)

[0461]

[0462] 표 P. 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(베타토르크바이러스)

명칭	고리 2
속/클레이드	베타토르크바이러스
수탁 번호	JX134045.1
단백질 수탁 번호	AGG91484.1

전체 서열: 666 AA

```

1      10      20      30      40      50
|      |      |      |      |      |
MPYYYRRRRYNYRRPRWYGRGWIRRPFRFRFRKRRVRPTYTTIPLKQWQ
PPYKRTCYIKGQDCLIIYSNLR LGMNSTMYEKSIVPVHWPGGGSFSVSM
TLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTE
LPANSNKLTYPNTHPLMMMSKYKHIPSRQTRRKKKPYTKIFVKPPPQF
ENKWFATDLYKIPLLQIHCTACNLQNPVKPKLSNNVTLWSLNTISIQ
NRNMSVDQQSQWPFKILGTQSFYFYFTGANLPGDQTQIPVADLLPLTNP
RINRPGQSLNEAKITDHITFTFYKKNFTNYWGNPFNKHIEHLDMILYSL
KSPEAIKNEWTENMKWNQLNAGTMALTPFNEPIFTQIQYNPDRDTGED
TQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGFAFQKNLKKVTNID
TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQLPEDNIKWYPQV
QYQLEAQNKLLQTGPFTPNIQQLSDNISMFYKFYFKWGGSPPKAINVEN
    
```

[0463]

PAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRSISQ  
 DWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQ  
 QQQLYRERIISLLKDQ (SEQ ID NO: 215)

주석:

추정되는 도메인	AA 범위
Arg-풍부 영역	1 - 38
젤리-롤 도메인	39 - 246
초가변 영역	247 - 374
N22	375 - 537
C-말단 도메인	538 - 666

[0464]

[0465]

표 Q. 예시적인 아벨로바이러스 ORF1 아미노산 하위서열 (베타토르크바이러스)

고리 2 ORF1 (베타토르크바이러스)	
Arg-풍부 영역	MPYYYYRRRRYNYRRPRWYGRGWIRRPFRFRFRKRRVR (SEQ ID NO: 216)
젤리-롤 도메인	PTYTTIPLKQWQPPYKRTCYIKGQDCLYYSNLRLGMNSTMYEKSIVPVHWP GGSFVSVM LTLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYI VRIHTELPANSNKLTYPNTHPLMMMSKYKHIIPSRQTRRRKKKPYTKIFVKPP PQFENKWFATDLYKIPLLQIHCTACNLQNPVFKPKLSNNVTLWLSNT (SEQ ID NO: 217)
초가변 도메인	ISIQNRNMSVDQGSWPFKILGTQSFYFYFTGANLPGDTTQIPVADLLPLTNP RINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPE AIKNEWTTENMKWNQLNAG (SEQ ID NO: 218)
N22	TMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPL WLIYWGFAFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDFVEGN SPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKF YFK (SEQ ID NO: 219)
C-말단 도메인	WGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYT TTALSRSISQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLI SNLRQQQLYRERIISLLKDQ (SEQ ID NO: 220)

[0466]

[0467] 표 R. 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

명칭 고리 4  
 속/클레이드 감마토르크바이러스  
 수탁 번호  
 단백질 수탁 번호

전체 서열: 662 AA

```

1          10          20          30          40          50
|          |          |          |          |          |
MPFWRRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRYPYRRPVRRRRRKL
RKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLIN
VPPKTPYGGGFVDQYTLKYLVEEYRFAQNIWTQSNVLKDLCRYINVKLI
FYRDNKTDVFLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKP
NGRLTKKLIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCN
ENPQVFFYYLNHGYYTITNWGAQSSTAYRPNKSVTDTTYRYKNDRNIN
IKSHEYEKSI SYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPV
KDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWGLFGYLDYCRQIK
ADKTWLDShVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNPPTL
LQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKW
GGPQFHPEIADPSKQEYDVPDFTFYQTIQIEDPEGQDPRSLIHDWDYRR
GFIKERSLKRMSYFSTHTDQQATSEEDI PKKKKRIGPQLTVPQQKEEET
LSCLLSLCKKDTFQETETQEDLQQLIKQQEQQLLLKRNILQLIHKLKEN
QQMLQLHTGMLP (SEQ ID NO: 925)
    
```

주석:

추정되는 도메인 AA 범위  
 Arg-풍부 영역 1 - 58  
 젤리-롤 도메인 59 - 260  
 추가변 영역 261 - 339  
 N22 340 - 499  
 C-말단 도메인 500 - 662

[0468]

[0469] 표 S. 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

고리 4 (감마토르크바이러스)	
Arg-풍부 영역	MPFWRRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRYPYRRPVRRRRRKL RKVKRKKK (SEQ ID NO: 926)
젤리-롤 도메인	SLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPKTPYGGGF GVDQYTLKYLVEEYRFAQNIWTQSNVLKDLCRYINVKLI FYRDNKTDVFLSY DRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGR LTKKLIKPPKQM LSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNENPQVFFYYL (SEQ ID NO: 927)
추가변 도메인	NHGYYTITNWGAQSSTAYRPNKSVTDTTYRYKNDRNINIKSHEYEKSI SYE NGYFQSSFLQTQCIYTSERGEACIAE (SEQ ID NO: 928)
N22	KPLGIAIYNPVKDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWGLFGY LDYCRQIKADKTWLDShVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFN QPPTLLQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTF K (SEQ ID NO: 929)
C-말단 도메인	WGGPQFHPEIADPSKQEYDVPDFTFYQTIQIEDPEGQDPRSLIHDWDYRRGFI KERSLKRMSYFSTHTDQQATSEEDI PKKKKRIGPQLTVPQQKEEETLSCLLSL CKKDTFQETETQEDLQQLIKQQEQQLLLKRNILQLIHKLKENQQMLQLHTG MLP (SEQ ID NO: 930)

[0470]

- [0471] 일부 실시형태에서, 제1 영역은 핵산 분자(예를 들어, DNA)에 결합할 수 있다. 일부 실시형태에서, 염기성 잔기는 아르기닌, 히스티딘 또는 라이신 또는 그의 조합으로부터 선택된다. 일부 실시형태에서, 제1 영역은 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 또는 100%의 아르기닌 잔기(예를 들어, 60% 내지 90%, 60% 내지 80%, 70% 내지 90% 또는 70 내지 80%의 아르기닌 잔기)를 포함한다. 일부 실시형태에서, 제1 영역은 약 30 내지 120 개의 아미노산(예를 들어, 약 40 내지 120, 40 내지 100, 40 내지 90, 40 내지 80, 40 내지 70, 50 내지 100, 50 내지 90, 50 내지 80, 50 내지 70, 60 내지 100, 60 내지 90 또는 60 내지 80개의 아미노산)을 포함한다. 일부 실시형태에서, 제1 영역은 바이러스 ORF1 아르기닌-풍부 영역(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 아르기닌-풍부 영역)의 구조 또는 활성을 포함한다. 일부 실시형태에서, 제1 영역은 핵 국소화 신호를 포함한다.
- [0472] 일부 실시형태에서, 제2 영역은 젤리-롤 도메인, 예를 들어, 바이러스 ORF1 젤리-롤 도메인(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 젤리-롤 도메인)의 구조 또는 활성을 포함한다. 일부 실시형태에서, 제2 영역은 또 다른 ORF1 분자의 제2 영역에 결합하여, 예를 들어, 단백질성 외부(예를 들어, 캡시드) 또는 그의 부분을 형성할 수 있다.
- [0473] 일부 실시형태에서, 제4 영역은 단백질성 외부(예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF1 분자의 다합체를 포함하는 단백질성 외부)의 표면 상에 노출된다.
- [0474] 일부 실시형태에서, 제1 영역, 제2 영역, 제3 영역, 제4 영역 및/또는 HVR은 각각 4개 미만(예를 들어, 0, 1, 2 또는 3개)의 베타 시트를 포함한다.
- [0475] 일부 실시형태에서, 제1 영역, 제2 영역, 제3 영역, 제4 영역, 및/또는 HVR 중 하나 이상은 이중 아미노산 서열(예를 들어, 이중 ORF1 분자로부터의 상응하는 영역)에 의해 대체될 수 있다. 일부 실시형태에서, 이중 아미노산 서열은 예를 들어, 본원에 기재된 바와 같은, 요망되는 작용성을 갖는다.
- [0476] 일부 실시형태에서, ORF1 분자는(예를 들어, PCT/US19/65995호의 도 34에 나타낸 바와 같이) 복수의 보존된 모티프(예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100개 이상의 아미노산을 포함하는 모티프)를 포함한다. 일부 실시형태에서, 보존된 모티프는 하나 이상의 야생형 아넬로바이러스 클레이드(예를 들어, 알파토르크바이러스, 클레이드 1; 알파토르크바이러스, 클레이드 2; 알파토르크바이러스, 클레이드 3; 알파토르크바이러스, 클레이드 4; 알파토르크바이러스, 클레이드 5; 알파토르크바이러스, 클레이드 6; 알파토르크바이러스, 클레이드 7; 베타토르크바이러스; 및/또는 감마토르크바이러스)의 ORF1 단백질에 대하여 60, 70, 80, 85, 90, 95 또는 100% 서열 동일성을 보일 수 있다. 실시형태들에서, 보존된 모티프는 각각 1 내지 1000개(예를 들어, 5 내지 10, 5 내지 15, 5 내지 20, 10 내지 15, 10 내지 20, 15 내지 20, 5 내지 50, 5 내지 100, 10 내지 50, 10 내지 100, 10 내지 1000, 50 내지 100, 50 내지 1000 또는 100 내지 1000개)의 아미노산의 길이를 갖는다. 특정 실시형태에서, 보존된 모티프는 약 2 내지 4%(예를 들어, 약 1 내지 8%, 1 내지 6%, 1 내지 5%, 1 내지 4%, 2 내지 8%, 2 내지 6%, 2 내지 5% 또는 2 내지 4%)의 ORF1 분자의 서열로 이루어지며, 각각은 야생형 아넬로바이러스 클레이드의 ORF1 단백질 내의 상응하는 모티프에 대하여 100% 서열 동일성을 보인다. 특정 실시형태에서, 보존된 모티프는 약 5 내지 10%(예를 들어, 약 1 내지 20%, 1 내지 10%, 5 내지 20% 또는 5 내지 10%)의 ORF1 분자의 서열로 이루어지며, 각각은 야생형 아넬로바이러스 클레이드의 ORF1 단백질 내의 상응하는 모티프에 대하여 80% 서열 동일성을 보인다. 특정 실시형태에서, 보존된 모티프는 약 10 내지 50%(예를 들어, 약 10 내지 20%, 10 내지 30%, 10 내지 40%, 10 내지 50%, 20 내지 40%, 20 내지 50% 또는 30 내지 50%)의 ORF1 분자의 서열로 이루어지며, 각각은 야생형 아넬로바이러스 클레이드의 ORF1 단백질 내의 상응하는 모티프에 대하여 60% 서열 동일성을 보인다. 일부 실시형태에서, 보존된 모티프는 표 19에 열거된 바와 같은 하나 이상의 아미노산 서열을 포함한다.
- [0477] 일부 실시형태에서, ORF1 분자 또는 이를 인코딩하는 핵산 분자는 예를 들어, 본원에 기재된 바와 같은 야생형 ORF1 단백질에 비하여 적어도 하나의 차이(예를 들어, 돌연변이, 화학적 변형 또는 후성적 변경)를 포함한다.
- [0478] *N22 도메인 내의 보존된 ORF1 모티프*
- [0479] 일부 실시형태에서, 본원에 기재된 폴리펩티드(예를 들어, ORF1 분자)는 아미노산 서열  $YNPX^2DXGX^2N$ (SEQ ID NO: 829)을 포함하며,  $X^n$ 은 임의의 n개의 아미노산의 인접 서열이다. 예를 들어,  $X^2$ 는 임의의 2개의 아미노산의 인접 서열을 나타낸다. 일부 실시형태에서,  $YNPX^2DXGX^2N$ (SEQ ID NO: 829)은 예를 들어, 본원에 기재된 바와 같은, ORF1 분자의 N22 도메인 내에 포함된다. 일부 실시형태에서, 본원에 기재된 유전 요소는 아미노산 서열

YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829)을 인코딩하는 핵산 서열(예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF1 분자를 인코딩하는 핵산 서열)을 포함하며, X<sup>n</sup>은 임의의 n개의 아미노산의 인접 서열이다.

[0480] 일부 실시형태에서, 폴리펩티드(예를 들어, ORF1 분자)는 예를 들어, N22 도메인 내의, 예를 들어, YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프의 부분을 포함하고/하거나 이에 축적하는 보존된 2차 구조를 포함한다. 일부 실시형태에서, 보존된 2차 구조는 제1 베타 가닥 및/또는 제2 베타 가닥을 포함한다. 일부 실시형태에서, 제1 베타 가닥은 약 5 내지 6개(예를 들어, 3, 4, 5, 6, 7 또는 8개)의 아미노산 길이이다. 일부 실시형태에서, 제1 베타 가닥은 YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프의 N-종말단에 티로신(Y) 잔기를 포함한다. 일부 실시형태에서, YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프는 랜덤 코일(예를 들어, 약 8 내지 9개의 아미노산의 랜덤 코일)을 포함한다. 일부 실시형태에서, 제2 베타 가닥은 약 7 내지 8개(예를 들어, 5, 6, 7, 8, 9 또는 10개) 아미노산 길이이다. 일부 실시형태에서, 제2 베타 가닥은 YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프의 C-종말단에 아스파라긴(N) 잔기를 포함한다.

[0481] 예시적인 YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프-축적 2차 구조는 그의 전문이 본원에 참조로 포함되는 PCT/US19/65995호의 실시예 47 및 도 48에 기재되어 있다. 일부 실시형태에서, ORF1 분자는 PCT/US19/65995호의 도 48에 나타난 2차 구조 요소(예를 들어, 베타 가닥)의 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 전부)을 포함하는 영역을 포함한다. 일부 실시형태에서, ORF1 분자는 (예를 들어, 본원에 기재된 바와 같은) YNPX<sup>2</sup>DXGX<sup>2</sup>N(SEQ ID NO: 829) 모티프에 축적하는, PCT/US19/65995호의 도 48에 나타난 2차 구조 요소(예를 들어, 베타 가닥) 중 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 전부)을 포함하는 영역을 포함한다.

[0482] *ORF1 젤리-롤 도메인 내의 보존된 2차 구조 모티프*

[0483] 일부 실시형태에서, 본원에 기재된 폴리펩티드(예를 들어, ORF1 분자)는 (예를 들어, 본원에 기재된 바와 같은) 아델로바이러스 ORF1 단백질에 의해 포함되는 하나 이상의 2차 구조 요소를 포함한다. 일부 실시형태에서, ORF1 분자는 (예를 들어, 본원에 기재된 바와 같은) 아델로바이러스 ORF1 단백질의 젤리-롤 도메인에 의해 포함되는 하나 이상의 2차 구조 요소를 포함한다. 일반적으로, ORF1 젤리-롤 도메인은 N-말단에서 C-말단 방향의 순서로, 제1 베타 가닥, 제2 베타 가닥, 제1 알파 나선, 제3 베타 가닥, 제4 베타 가닥, 제5 베타 가닥, 제2 알파 나선, 제6 베타 가닥, 제7 베타 가닥, 제8 베타 가닥 및 제9 베타 가닥을 포함하는 2차 구조를 포함한다. 일부 실시형태에서, ORF1 분자는 N-말단에서 C-말단 방향의 순서로, 제1 베타 가닥, 제2 베타 가닥, 제1 알파 나선, 제3 베타 가닥, 제4 베타 가닥, 제5 베타 가닥, 제2 알파 나선, 제6 베타 가닥, 제7 베타 가닥, 제8 베타 가닥 및/또는 제9 베타 가닥을 포함하는 2차 구조를 포함한다.

[0484] 일부 실시형태에서, 보존된 2차 구조 요소(즉, 베타 가닥 및/또는 알파 나선)의 쌍은 예를 들어, 랜덤 코일 서열, 베타 가닥 또는 알파 나선 또는 그의 조합을 포함하는 개재 아미노산 서열에 의해 분리된다. 보존된 2차 구조 요소 사이의 개재 아미노산 서열은 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 이상의 아미노산을 포함할 수 있다. 일부 실시형태에서, ORF1 분자는 하나 이상의 추가의 베타 가닥 및/또는 알파 나선을 (예를 들어, 젤리-롤 도메인 내) 추가로 포함할 수 있다. 일부 실시형태에서, 연속 베타 가닥 또는 연속 알파 나선이 조합될 수 있다. 일부 실시형태에서, 제1 베타 가닥 및 제2 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 실시형태에서, 제3 베타 가닥 및 제4 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 실시형태에서, 제4 베타 가닥 및 제5 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 실시형태에서, 제6 베타 가닥 및 제7 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 실시형태에서, 제7 베타 가닥 및 제8 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 실시형태에서, 제8 베타 가닥 및 제9 베타 가닥은 더 큰 베타 가닥 내에 포함된다.

[0485] 일부 실시형태에서, 제1 베타 가닥은 약 5 내지 7개(예를 들어, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다. 일부 실시형태에서, 제2 베타 가닥은 약 15 내지 16개(예를 들어, 13, 14, 15, 16, 17, 18 또는 19개)의 아미노산 길이이다. 일부 실시형태에서, 제1 알파 나선은 약 15 내지 17개(예를 들어, 13, 14, 15, 16, 17, 18, 19 또는 20개)의 아미노산 길이이다. 일부 실시형태에서, 제3 베타 가닥은 약 3 내지 4개(예를 들어, 1, 2, 3, 4, 5 또는 6개)의 아미노산 길이이다. 일부 실시형태에서, 제4 베타 가닥은 약 10 내지 11개(예를 들어, 8, 9, 10, 11, 12 또는 13개)의 아미노산 길이이다. 일부 실시형태에서, 제5 베타 가닥은 약 6 내지 7개(예

를 들어, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다. 일부 실시형태에서, 제2 알파 나선은 약 8 내지 14개(예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 또는 17개)의 아미노산 길이이다. 일부 실시형태에서, 제2 알파 나선은 (예를 들어, 랜덤 코일 서열에 의해 분리된) 2개의 더 작은 알파 나선으로 파단될 수 있다. 일부 실시형태에서, 2개의 더 작은 알파 나선의 각각은 약 4 내지 6개(예를 들어, 2, 3, 4, 5, 6, 7 또는 8개)의 아미노산 길이이다. 일부 실시형태에서, 제6 베타 가닥은 약 4 내지 5개(예를 들어, 2, 3, 4, 5, 6 또는 7개)의 아미노산 길이이다. 일부 실시형태에서, 제7 베타 가닥은 약 5 내지 6개(예를 들어, 3, 4, 5, 6, 7, 8 또는 9개)의 아미노산 길이이다. 일부 실시형태에서, 제8 베타 가닥은 약 7 내지 9개(예를 들어, 5, 6, 7, 8, 9, 10, 11, 12 또는 13개)의 아미노산 길이이다. 일부 실시형태에서, 제9 베타 가닥은 약 5 내지 7개(예를 들어, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다.

[0486] 예시적인 젤리-롤 도메인 2차 구조는 PCT/US19/65995호의 실시예 47 및 도 47에 기재되어 있다. 일부 실시형태에서, ORF1 분자는 PCT/US19/65995호의 도 47에 나타낸 젤리-롤 도메인 2차 구조 중 임의의 것의 2차 구조 요소(예를 들어, 베타 가닥 및/또는 알파 나선)의 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 전부)을 포함하는 영역을 포함한다.

[0487] 공통 ORF1 도메인 서열

[0488] 일부 실시형태에서, 예를 들어, 본원에 기재된 바와 같은, ORF1 분자는 젤리-롤 도메인, N22 도메인, 및/또는 C-말단 도메인(CTD) 중 하나 이상을 포함한다. 일부 실시형태에서, 젤리-롤 도메인은 본원에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 임의의 것에 열거된 바와 같은) 젤리-롤 도메인 공통 서열을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, N22 도메인은 본원에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 임의의 것에 열거된 바와 같은) N22 도메인 공통 서열을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, CTD 도메인은 본원에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 임의의 것에 열거된 바와 같은) CTD 도메인 공통 서열을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, 표 37A 내지 37C 중 임의의 것에 형식 "(X<sub>a</sub>-b)"로 열거된 아미노산은 인접 아미노산의 연속물을 포함하며, 여기서, 연속물은 적어도 a개 및 최대 b개의 아미노산을 포함한다. 특정 실시형태에서, 연속물에서 모든 아미노산은 동일하다. 다른 실시형태에서, 연속물은 적어도 2개(예를 들어, 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 21개)의 상이한 아미노산을 포함한다.

[0489] 표 37A. 알파토르크바이러스 ORF1 도메인 공통 서열

도메인	서열	SEQ ID NO:
젤리-롤	LVL <sub>0.3</sub> TQWQPNTVRRRCYIRGYLPLIICGEN(X <sub>0.3</sub> )TTSRNYATHS DDTIQKGPFGGGMSTTTFSLRVLYDEYQRFMNRWTYSNED LDLARYLGCKFTFYRHPDXDFIVQYNTNPPFKDTKLTAPSIH P(X <sub>1.5</sub> )GMLMLSKRKILIPSLKTRPKGKHYVKVRIGPPKLFED KWYTQSDLCDVPLVXL <sub>Y</sub> ATAADLQHPFGSPQTDNPCVTFQ VLGSXYNKHLSISP;  상기 식에서, X = 임의의 아미노산.	227
N22	SNFEFPGAYTDITYNPLTDKGVGNMVW <sub>IQ</sub> YLTKPDTIXDKT QS(X <sub>0.3</sub> )KCLIEDLPLWAALYGYVDFCEKETGDSAIIXNXGRV LIRCPYTKPPLYDKT(X <sub>0.4</sub> )NKGFPYSTNFGNGKMPGGSGY VPIYWRARWYPTLFHQKEVLEDIVQSGPFAYKDEK <sub>P</sub> STQLV MKYCFNFN;  상기 식에서, X = 임의의 아미노산.	228
CTD	WGGNPISQVV <sub>R</sub> NPCKDSG(X <sub>0.3</sub> )SGXGRQPRSVQV <sub>VD</sub> PKY MGPEYTFHSWDWRRGLFGEKAIKRMSEQPTDDEIFTGGXPK RPRRD <sub>P</sub> PTXQXPEE(X <sub>1.4</sub> )QKESSFR(X <sub>2-14</sub> )PWESSSQEXESES QEEEE(X <sub>0.30</sub> )EQT <sub>V</sub> QQQLRQQLREQRRLRVQLQLLFQQLLKT (X <sub>0.4</sub> )QAGLHINPLLSQA(X <sub>0.40</sub> )*;  상기 식에서, X = 임의의 아미노산.	229

[0490]

[0491] 표 37B. 베타토르크바이러스 ORF1 도메인 공통 서열

도메인	서열	SEQ ID NO:
젤리-롤	LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESIVPH HEPGGGGWSIQFTLGALYEEHLKLRNWWTKSNDGLPLVR YLGCTIKLYRSEDTDYIVTYQRCYPMTATKLYLSTQPSRM LMNKHKIIVPSKXT(X <sub>1-4</sub> )NKKKKPKYKIFIKPPSQMQNKWYF QQDIANTPLLQLTXTACSLDRMYLSSDSISNNITFTSLNTNFF QNPNFQ;  상기 식에서, X = 임의의 아미노산.	230
N22	(X <sub>4-10</sub> )TPLYFECRYNPFKDKGTGNKVYLVSN(X <sub>1-8</sub> )TGWDPP TDPDLIIEGFPLWLLWGWLDWQKLGKIQNIDTDYILVIQS XYIYPP(X <sub>1-3</sub> )KLPYVPLDXD(X <sub>0-2</sub> )FLHGRSPY(X <sub>3-16</sub> )PSDKQH WHPKVRFXETINNIALTGPSTPKLPNQKSIQAHMKYKFF K;  상기 식에서, X = 임의의 아미노산.	231
CTD	WGGCPAPMETITDPCKQPKYPIPNLLQTSLQXPTPIETYL YKFDERRGLLTKKAAKRIKDXTTETTLFTDTGXSTTLPT XXQTETTQEEXTSEEE(X <sub>0-5</sub> )ETLLQQLQLRRKQQLRXRIL QLLQLLXLL(X <sub>0-26</sub> )*;  상기 식에서, X = 임의의 아미노산.	232

[0492]

[0493] 표 37C. 감마토르크바이러스 ORF1 도메인 공통 서열

도메인	서열	SEQ ID NO:
젤리-롤	TIPLKQWPESIRKCKIKGYGLVLAEGRQFYCYTNEKDE YTPPKAPGGGGFVLEFSLEYLYEQWKARNNIWTKSNXYK DLCRYTGCKITFYRHPTTDFIVXYSRQPPFEIDKXTYMXHP QXLLLRKHKKIILSKATNPKGKLLKKIKIKPPKQMLNKWFF QKQFAXYGLVQLQAAACBLRYPRLGCCNENRLITLYYLN;  상기 식에서, X = 임의의 아미노산.	233
N22	LPIVVARYNPAXDTGKGNKXWLXSTLNGSXWAPPTTDKDL IIEGLPLWLALYGYWSYJKKVKKDKGILQSHMFVVKSPAIQP LXTATTQXTFYPXIDNSFIQGXYPDEPJTXXNQKKLWYPTLE HQQETINAIVESGPYVPKLDNQKNSTWELXYXYTFYFK;  상기 식에서, X = 임의의 아미노산.	234
CTD	WGGPQIPDQPVDPKXQGTYPVPTXQQTIQIXNPLKQKPE TMFHDWDYRRGIITSTALKRMQENLETSSFXSDSEETP(X <sub>0-2</sub> )KKKKRLTXELXPQEETEEIQSCLLSLCEESTCQEE(X <sub>1-6</sub> )ENL QQLIHQQQQQQQLKHNILKLLSDLKZKQRLQLQTGILE(X 1-10)*;  상기 식에서, X = 임의의 아미노산.	235

[0494]

[0495] 일부 실시형태에서, 젤리-롤 도메인은 표 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10 또는 37A 내

지 37C 중 임의의 것에 열거된 바와 같은 젤리-롤 도메인 아미노산 서열 또는 이에 대하여 적어도 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, N22 도메인은 표 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10 또는 37A 내지 37C 중 임의의 것에 열거된 바와 같은 N22 도메인 아미노산 서열 또는 이에 대하여 적어도 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, CTD 도메인은 표 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10 또는 37A 내지 37C 중 임의의 것에 열거된 바와 같은 CTD 도메인 아미노산 서열 또는 이에 대하여 적어도 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다.

[0496] *ORF1 단백질 서열의 식별*

[0497] 일부 실시형태에서, 아넬로바이러스 ORF1 단백질 서열, 또는 ORF1 단백질을 인코딩하는 핵산 서열은 아넬로바이러스의 계놈(예를 들어, 핵산 시퀀싱 기술, 예를 들어, 심층 시퀀싱 기술에 의해 식별된 추정 아넬로바이러스 계놈)으로부터 식별될 수 있다. 일부 실시형태에서, ORF1 단백질 서열은 다음의 선택 기준 중 하나 이상(예를 들어, 1, 2, 또는 3개 모두)에 의해 식별된다:

[0498] (i) *길이 선택*: 단백질 서열(예를 들어, 하기 (ii) 또는 (iii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)은 추정 아넬로바이러스 ORF1 단백질을 식별하기 위해 약 600개 초과 아미노산 잔기에 대해 크기-선택될 수 있다. 일부 실시형태에서, 아넬로바이러스 ORF1 단백질 서열은 적어도 약 600, 650, 700, 750, 800, 850, 900, 950 또는 1000개의 아미노산 잔기 길이이다. 일부 실시형태에서, 알파토르크바이러스 ORF1 단백질 서열은 적어도 약 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 900 또는 1000개의 아미노산 잔기 길이이다. 일부 실시형태에서, 베타토르크바이러스 ORF1 단백질 서열은 적어도 약 650, 660, 670, 680, 690, 700, 750, 800, 900 또는 1000개의 아미노산 잔기 길이이다. 일부 실시형태에서, 감마토르크바이러스 ORF1 단백질 서열은 적어도 약 650, 660, 670, 680, 690, 700, 750, 800, 900 또는 1000개의 아미노산 잔기 길이이다. 일부 실시형태에서, 아넬로바이러스 ORF1 단백질을 인코딩하는 핵산 서열은 적어도 약 1800, 1900, 2000, 2100, 2200, 2300, 2400 또는 2500개의 뉴클레오티드 길이이다. 일부 실시형태에서, 알파토르크바이러스 ORF1 단백질 서열을 인코딩하는 핵산 서열은 적어도 약 2100, 2150, 2200, 2250, 2300, 2400 또는 2500개의 뉴클레오티드 길이이다. 일부 실시형태에서, 베타토르크바이러스 ORF1 단백질을 인코딩하는 핵산 서열은 적어도 약 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400 또는 2500 또는 1000개의 뉴클레오티드 길이이다. 일부 실시형태에서, 감마토르크바이러스 ORF1 단백질을 인코딩하는 핵산 서열은 적어도 약 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400 또는 2500 또는 1000개의 뉴클레오티드 길이이다.

[0499] (ii) *ORF1 모티프의 존재*: 단백질 서열(예를 들어, 위의 (i) 또는 하기 (iii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)은 위에 기재된 N22 도메인 내의 보존된 ORF1 모티프를 함유하는 것들을 식별하기 위해 필터링될 수 있다. 일부 실시형태에서, 추정 아넬로바이러스 ORF1 서열은 서열 YNPXXDXGXN을 포함한다. 일부 실시형태에서, 추정 아넬로바이러스 ORF1 서열은 서열 Y[NCS]PXXDX[GASKR].XX[NTSVAK]를 포함한다.

[0500] (iii) *아르기닌-풍부 영역의 존재*: 단백질 서열(예를 들어, 위의 (i) 및/또는 (ii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)은 아르기닌-풍부 영역(예를 들어, 본원에 기재된 바와 같음)을 포함하는 것들에 대해 필터링될 수 있다. 일부 실시형태에서, 추정 아넬로바이러스 ORF1 서열은 적어도 약 30%(예를 들어, 적어도 약 20%, 25%, 30%, 35%, 40%, 45% 또는 50%) 아르기닌 잔기를 포함하는 적어도 약 30, 35, 40, 45, 50, 55, 60, 65 또는 70개의 아미노산의 인접 서열을 포함한다. 일부 실시형태에서, 추정 아넬로바이러스 ORF1 서열은 적어도 30% (예를 들어, 적어도 약 20%, 25%, 30%, 35%, 40%, 45% 또는 50%) 아르기닌 잔기를 포함하는 약 35 내지 40, 40 내지 45, 45 내지 50, 50 내지 55, 55 내지 60, 60 내지 65 또는 65 내지 70개의 아미노산의 인접 서열을 포함한다. 일부 실시형태에서, 아르기닌-풍부 영역은 추정 아넬로바이러스 ORF1 단백질의 시작 코돈의 적어도 약 30, 40, 50, 60, 70 또는 80개의 아미노산 하류에 위치한다. 일부 실시형태에서, 아르기닌-풍부 영역은 추정 아넬로바이러스 ORF1 단백질의 시작 코돈의 적어도 약 50개의 아미노산 하류에 위치한다.

[0501] *ORF2 분자*

[0502] 일부 실시형태에서, 아넬로바이러스는 ORF2 분자 및/또는 ORF2 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF2 분자는 아넬로바이러스 ORF2 단백질(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 표 A2, A4, A6, A8, A10, A12, C1 내지 C5, 2, 4, 6, 8, 10, 12, 14, 16 또는 18 중 임의의 것에 열거된 바와 같은 아넬로바이러스 ORF2 단백질)의 구조적 특징 및/또는 활성을 갖는 폴리펩티드 또는 그의 기능적 단편을 포함한다. 일부 실시형태에서, ORF2 분자는 표 A2, A4, A6, A8, A10, A12, C1 내지 C5, 2, 4, 6, 8, 10, 12, 14, 16 또는 18 중 임의

의 것에 나타낸 바와 같은 아넬로바이러스 ORF2 단백질 서열에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다.

[0503] 일부 실시형태에서, ORF2 분자는 알파토르크바이러스, 베타토르크바이러스 또는 감마토르크바이러스 ORF2 단백질에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, ORF2 분자(예를 들어, 알파토르크바이러스 ORF2 단백질에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 ORF2 분자)는 250개 이하의 아미노산(예를 들어, 약 150 내지 200개의 아미노산)의 길이를 갖는다. 일부 실시형태에서, ORF2 분자(예를 들어, 베타토르크바이러스 ORF2 단백질에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 ORF2 분자)는 약 50 내지 150개의 아미노산의 길이를 갖는다. 일부 실시형태에서, ORF2 분자(예를 들어, 감마토르크바이러스 ORF2 단백질에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 ORF2 분자)는 약 100 내지 200개의 아미노산(예를 들어, 약 100 내지 150개의 아미노산)의 길이를 갖는다. 일부 실시형태에서, ORF2 분자는 나선-회전-나선 모티프(예를 들어, 회전 영역에 축적하는 2개의 알파 나선을 포함하는 나선-회전-나선 모티프)를 포함한다. 일부 실시형태에서, ORF2 분자는 TTV 단리주 TA278 또는 TTV 단리주 SANBAN의 ORF2 단백질의 아미노산 서열을 포함하지 않는다. 일부 실시형태에서, ORF2 분자는 단백질 포스파타제 활성을 갖는다. 일부 실시형태에서, ORF2 분자 또는 이를 인코딩하는 핵산 분자는 예를 들어, 본원에 기재된 바와 같은(예를 들어, 표 A2, A4, A6, A8, A10, A12, C1 내지 C5, 2, 4, 6, 8, 10, 12, 14, 16 또는 18 중 임의의 것에 나타낸 바와 같은) 야생형 ORF2 단백질에 비하여 적어도 하나의 차이(예를 들어, 돌연변이, 화학적 변형 또는 후성적 변경)를 포함한다.

[0504] 보존된 ORF2 모티프

[0505] 일부 실시형태에서, 본원에 기재된 폴리펩티드(예를 들어, ORF2 분자)는 아미노산 서열  $[W/F]X^7HX^3CX^1CX^5H$ (SEQ ID NO: 949)를 포함하며, 여기서,  $X^n$ 은 임의의 n개의 아미노산의 인접 서열이다. 실시형태들에서,  $X^7$ 은 임의의 7개의 아미노산의 인접 서열을 나타낸다. 실시형태들에서,  $X^3$ 은 임의의 3개의 아미노산의 인접 서열을 나타낸다. 실시형태들에서,  $X^1$ 은 임의의 단일의 아미노산을 나타낸다. 실시형태들에서,  $X^5$ 는 임의의 5개의 아미노산의 인접 서열을 나타낸다. 일부 실시형태에서, [W/F]는 트립토판 또는 페닐알라닌 중 어느 하나일 수 있다. 일부 실시형태에서,  $[W/F]X^7HX^3CX^1CX^5H$ (SEQ ID NO: 949)는 예를 들어, 본원에 기재된 바와 같은 ORF2 분자의 N22 도메인 내에 포함된다. 일부 실시형태에서, 본원에 기재된 유전 요소는 아미노산 서열  $[W/F]X^7HX^3CX^1CX^5H$ (SEQ ID NO: 949)를 인코딩하는 핵산 서열(예를 들어, 본원에 기재된 바와 같은, 예를 들어, ORF2 분자를 인코딩하는 핵산 서열)을 포함하며, 여기서,  $X^n$ 은 임의의 n개의 아미노산의 인접 서열이다.

[0506] 유전 요소

[0507] 일부 실시형태에서, 아넬로바이러스는 유전 요소를 포함한다. 일부 실시형태에서, 유전 요소는 다음의 특징 중 하나 이상을 갖는다: 숙주 세포의 게놈과 실질적으로 비-통합성임, 에피솜 핵산임, 단일 가닥 RNA임, 환형임, 약 1 내지 10 kb임, 세포의 핵 내에 존재함, 내인성 단백질에 의해 결합될 수 있음, 이펙터, 에센스 숙주 또는 표적 세포의 유전자, 활성 또는 기능을 표적화하는 폴리펩티드 또는 핵산(예를 들어, RNA, iRNA, 마이크로RNA)를 생성함. 일 실시형태에서, 유전 요소는 실질적으로 비-통합성이다. 일부 실시형태에서, 유전 요소는 패키징 신호, 예를 들어, 캡시드 단백질에 결합하는 서열을 포함한다. 일부 실시형태에서, 패키징 또는 캡시드-결합 서열의 외부에서, 유전 요소는 야생형 아넬로바이러스 핵산 서열에 대하여 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% 미만의 서열 동일성을 가지며, 예를 들어, 이는 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 핵산 서열에 대하여 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% 미만의 서열 동일성을 갖는다. 일부 실시형태에서, 패키징 또는 캡시드-결합 서열의 외부에서, 유전 요소는 아넬로바이러스 핵산 서열에 대하여 적어도 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 동일한 500, 450, 400, 350, 300, 250, 200, 150 또는 100개 미만의 인접 뉴클레오티드를 갖는다.

[0508] 일부 실시형태에서, 유전 요소는 20 kb 미만(예를 들어, 약 19 kb, 18 kb, 17 kb, 16 kb, 15 kb, 14 kb, 13 kb, 12 kb, 11 kb, 10kb, 9 kb, 8 kb, 7 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb 또는 그 미만)의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 독립적으로 또는 부가적으로 1000b 초과(예를 들어, 적어도 약 1.1 kb, 1.2 kb, 1.3 kb, 1.4 kb, 1.5 kb, 1.6 kb, 1.7 kb, 1.8 kb, 1.9 kb, 2 kb, 2.1 kb, 2.2 kb, 2.3 kb, 2.4 kb, 2.5 kb, 2.6 kb, 2.7 kb, 2.8 kb, 2.9 kb, 3 kb, 3.1 kb, 3.2 kb, 3.3 kb, 3.4 kb, 3.5 kb, 3.6 kb, 3.7 kb,

3.8 kb, 3.9 kb, 4 kb, 4.1 kb, 4.2 kb, 4.3 kb, 4.4 kb, 4.5 kb, 4.6 kb, 4.7 kb, 4.8 kb, 4.9 kb, 5 kb 이상)의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 2.5 내지 4.6, 2.8 내지 4.0, 3.0 내지 3.8 또는 3.2 내지 3.7 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 1.5 내지 2.0, 1.5 내지 2.5, 1.5 내지 3.0, 1.5 내지 3.5, 1.5 내지 3.8, 1.5 내지 3.9, 1.5 내지 4.0, 1.5 내지 4.5 또는 1.5 내지 5.0 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 2.0 내지 2.5, 2.0 내지 3.0, 2.0 내지 3.5, 2.0 내지 3.8, 2.0 내지 3.9, 2.0 내지 4.0, 2.0 내지 4.5 또는 2.0 내지 5.0 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 2.5 내지 3.0, 2.5 내지 3.5, 2.5 내지 3.8, 2.5 내지 3.9, 2.5 내지 4.0, 2.5 내지 4.5 또는 2.5 내지 5.0 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 3.0 내지 5.0, 3.5 내지 5.0, 4.0 내지 5.0 또는 4.5 내지 5.0 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 1.5 내지 2.0, 2.0 내지 2.5, 2.5 내지 3.0, 3.0 내지 3.5, 3.1 내지 3.6, 3.2 내지 3.7, 3.3 내지 3.8, 3.4 내지 3.9, 3.5 내지 4.0, 4.0 내지 4.5 또는 4.5 내지 5.0 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 3.6 내지 3.9 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 2.8 내지 2.9 kb의 길이를 갖는다. 일부 실시형태에서, 유전 요소는 약 2.0 내지 3.2 kb의 길이를 갖는다.

[0509] 일부 실시형태에서, 유전 요소는 본원에 기재된 특징, 예를 들어, 실질적으로 비-병원성인 단백질을 인코딩하는 서열, 단백질 결합 서열, 조절 핵산을 인코딩하는 하나 이상의 서열, 하나 이상의 조절 서열, 복제 단백질을 인코딩하는 하나 이상의 서열 및 기타 서열 중 하나 이상을 포함한다.

[0510] 실시형태에서, 유전 요소는 이중-가닥 환형 DNA로부터 생성되었다(예를 들어, 전사에 의해 생성됨).

[0511] 일부 실시형태에서, 유전 요소는 하나 이상의 박테리아 플라스미드 요소(예를 들어, 박테리아 복제 원점 또는 선택 가능한 마커, 예를 들어, 박테리아 내성 유전자)를 포함하지 않는다. 일부 실시형태에서, 유전 요소는 박테리아 플라스미드 백본을 포함하지 않는다.

[0512] 일 실시형태에서, 본 개시내용은 (i) 실질적으로 비-병원성인 외부 단백질, (ii) 유전 요소를 실질적으로 비-병원성인 외부 단백질에 결합시키는 외부 단백질 결합 서열 및 (iii) 조절 핵산을 인코딩하는 핵산 서열(예를 들어, RNA 서열)을 포함하는 유전 요소를 제공한다. 이러한 실시형태에서, 유전 요소는 고유 바이러스 서열(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 고유 아넬로바이러스 서열)에 대한 뉴클레오티드 서열 중 어느 하나에 대하여 적어도 약 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% 및 99% 뉴클레오티드 서열 동일성을 갖는 하나 이상의 서열을 포함할 수 있다.

[0513] **단백질 결합 서열**

[0514] 일부 실시형태에서, 유전 요소는 실질적으로 비-병원성인 단백질에 결합하는 단백질 결합 서열을 인코딩한다. 일부 실시형태에서, 단백질 결합 서열은 단백질성 외부로의 유전 요소의 패키징을 용이하게 한다. 일부 실시형태에서, 단백질 결합 서열은 실질적으로 비-병원성인 단백질의 아르기닌-풍부 영역에 특이적으로 결합한다. 일부 실시형태에서, 유전 요소는 PCT/US19/65995호의 실시예 8에 기재된 바와 같은 단백질 결합 서열을 포함한다.

[0515] 일부 실시형태에서, 유전 요소는 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 서열의 5' UTR 보존된 도메인 또는 GC-풍부 도메인에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 단백질 결합 서열을 포함한다.

[0516] 실시형태에서, 단백질 결합 서열은 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 5' UTR 보존된 도메인 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는다.

[0517] **5' UTR 영역**

[0518] 일부 실시형태에서, 본원에 기재된 바와 같은 핵산 분자(예를 들어, 유전 요소, 유전 요소 작제물 또는 유전 요소 영역)는 5' UTR 서열, 예를 들어, 본원에(예를 들어, 표 A1, B1 또는 C1 중 임의의 것에) 기재된 바와 같은 5' UTR 보존된 도메인 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다.

[0519] 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGA AACACCACCGAAGTCAAGGGCAATTCGGGCTAGGGX<sub>1</sub>CAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGA AACACCACCGAAGTCAAGGGCAATTCGGGCTAGGGX<sub>1</sub>CAGTCT, 또는 이에 비하여 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실 또는 부가)를

갖는 핵산 서열을 포함한다. 실시형태들에서,  $X_1$ 은 A이다. 실시형태들에서,  $X_1$ 은 부재이다.

[0520] 일부 실시형태에서, 5' UTR 서열은 알파토르크바이러스의 5' UTR의 핵산 서열(예를 들어, 고리1), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 5' UTR 서열은 표 A1에 열거된 5' UTR 보존된 도메인의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95.775% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTTTACACACCGCAGTCAAGGGCAATTCGGGCTCGGGACTGGC, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTTTACACACCGCAGTCAAGGGCAATTCGGGCTCGGGACTGGC, 또는 이에 비하여 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실 또는 부가)를 갖는 핵산 서열을 포함한다.

[0521] 일부 실시형태에서, 5' UTR 서열은 베타토르크바이러스의 5' UTR의 핵산 서열(예를 들어, 고리2), 또는 이에 대하여 적어도 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 5' UTR 서열은 표 B1에 열거된 5' UTR 보존된 도메인의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 85% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 87% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 87.324% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 88% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 88.732% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 91% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 91.549% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 92% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 92.958% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 94% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 94.366% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTAAAACCCGAAGTCAAGGGCAATTCGGGCTAGATCAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTAAAACCCGAAGTCAAGGGCAATTCGGGCTAGATCAGTCT, 또는 이에 비하여 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실 또는 부가)를 갖는 핵산 서열을 포함한다.

[0522] 일부 실시형태에서, 5' UTR 서열은 감마토르크바이러스의 5' UTR의 핵산 서열(예를 들어, 고리4), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 실시형태들에서, 5' UTR 서열은 표 C1에 열거된 5' UTR 보존된 도메인의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 C1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 핵산 분자는 표 C1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열

AGGTGAGTGAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 또는 이에 비하여 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10 개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실 또는 부가)를 갖는 핵산 서열을 포함한다.

[0523]

일부 실시형태에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 아넬로바이러스 5' UTR 서열, 예를 들어, 표 38에 나타난 핵산 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 일부 실시형태에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 공통 5' UTR 서열의 핵산 서열을 포함하며,  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$  및  $X_5$  는 각각 독립적으로, 임의의 뉴클레오티드이며, 예를 들어,  $X_1 = G$  또는  $T$ 이며,  $X_2 = C$  또는  $A$ 이며,  $X_3 = G$  또는  $A$ 이며,  $X_4 = T$  또는  $C$ 이며,  $X_5 = A$ ,  $C$  또는  $T$ 이다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 공통 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 예시적인 TTV 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 TTV-CT30F 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 TTV-HD23a 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 TTV-JA20 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 TTV-TJN02 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 TTV-tth8 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다.

[0524]

실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 공통 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 1 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 2 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 3 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 4 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 5 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 6 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타난 알파토르크바이러스 클레이드 7 5' UTR 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다.

[0525] 표 38. 아넬로바이러스로부터의 예시적인 5' UTR 서열

공급원	서열	SEQ ID NO:
공통	CGGGTGCCGX <sub>1</sub> AGGTGAGTTTACACACCGX <sub>2</sub> AGT CAAGGGGCAATTCGGGCTCX <sub>3</sub> GGACTGGCCGGG CX <sub>4</sub> X <sub>5</sub> TGGG X <sub>1</sub> = G 또는 T X <sub>2</sub> = C 또는 A X <sub>3</sub> = G 또는 A X <sub>4</sub> = T 또는 C X <sub>5</sub> = A, C, 또는 T	105
예시적인 TTV 서열	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT WTGGG	106
TTV-CT30F	CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	107
TTV-HD23a	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CTGGG	108
TTV-JA20	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT TTGGG	109

[0526]

TTV-TJN02	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	110
TTV-tth8	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	111
알파토르크바이러스 공통 5' UTR	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGC X <sub>1</sub> X <sub>2</sub> TGGG; 상기 식에서, X <sub>1</sub> 은 T 또는 C 를 포함하고, 상기 식에서, X <sub>2</sub> 는 A, C, 또는 T 를 포함함.	112
알파토르크바이러스 클레이드 1 5' UTR (예를 들어, TTV-CT30F)	CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	113
알파토르크바이러스 클레이드 2 5' UTR (예를 들어, TTV-P13-1)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CGGG	114
알파토르크바이러스 클레이드 3 5' UTR (예를 들어, TTV-tth8)	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	115
알파토르크바이러스 클레이드 4 5' UTR (예를 들어, TTV-HD20a)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGAGGCCGGGCCAT GGG	116
알파토르크바이러스 클레이드 5 5' UTR (예를 들어, TTV-16)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CCGGG	117
알파토르크바이러스 클레이드 6 5' UTR (예를 들어, TTV-TJN02)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	118
알파토르크바이러스 클레이드 7 5' UTR (예를 들어, TTV-HD16d)	CGGGTGCCGAAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	119

[0527]

[0528]

5' UTR 서열의 식별

[0529]

일부 실시형태에서, 아넬로바이러스 5' UTR 서열은 아넬로바이러스의 게놈(예를 들어, 핵산 시퀀싱 기술, 예를 들어, 심층 시퀀싱 기술에 의해 식별된 추정 아넬로바이러스 게놈) 내에서 식별될 수 있다. 일부 실시형태에서, 아넬로바이러스 5' UTR 서열은 다음의 단계 중 하나 또는 둘 모두에 의해 식별된다:

[0530]

(i) **환화 접합점의 식별:** 일부 실시형태에서, 5' UTR은 전장의 환화된 아넬로바이러스 게놈의 환화 접합점 근처에 위치할 것이다. 환화 접합점은 예를 들어, 서열의 중첩 영역을 식별함으로써 식별될 수 있다. 일부 실시형태에서, 서열의 중첩 영역은 환화된 전장 아넬로바이러스 게놈 서열을 생성하기 위해 서열로부터 트리밍될 수 있다. 일부 실시형태에서, 게놈 서열은 소프트웨어를 사용하여 이러한 방식으로 환화된다. 이론에 결부시키지 않고, 게놈을 진산적으로 환화하면 서열의 시작 위치가 비-생물학적으로 배향되는 것을 초래할 수 있다. 서열 내의 랜드마크는 서열을 적절한 방향으로 재-배향하기 위해 사용될 수 있다. 예를 들어, 랜드마크 서열은 본원에 기재된 바와 같은 아넬로바이러스 게놈 내의 하나 이상의 요소(예를 들어, 본원에 기재된 바와 같은, 예를 들어, 아넬로바이러스의 TATA 박스, 캡 부위, 개시자 요소, 전사 시작 부위, 5' UTR 보존된 도메인, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, 3개의 오픈-리딩 프레임 영역, 폴리(A) 신호 또는 GC-풍부 영역 중 하나 이상)에 대하여 실질적으로 상동성을 갖는 서열을 포함할 수 있다.

[0531]

(ii) **5' UTR 서열의 식별:** 추정 아넬로바이러스 게놈 서열이 수득되면, 서열(또는 예를 들어, 약 40 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90 또는 90 내지 100개의 뉴클레오타이드의 길이를 갖는 그의

일부)은 이에 실질적으로 상동성을 갖는 서열을 식별하기 위해 하나 이상의 아넬로바이러스 5' UTR 서열(예를 들어, 본원에 기재된 바와 같음)과 비교될 수 있다. 일부 실시형태에서, 추정 아넬로바이러스 5' UTR 영역은 본원에 기재된 바와 같은 아넬로바이러스 5' UTR 서열에 대하여 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는다.

[0532] **GC-풍부 영역**

[0533] 일부 실시형태에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 핵산 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 GC-풍부 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다.

[0534] 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 바와 같은 36-뉴클레오티드 GC-풍부 서열(예를 들어, 36-뉴클레오티드 공통 GC-풍부 영역 서열 1, 36-뉴클레오티드 공통 GC-풍부 영역 서열 2, TTV 클레이드 1 36-뉴클레오티드 영역, TTV 클레이드 3 36-뉴클레오티드 영역, TTV 클레이드 3 단리주 GH1 36-뉴클레오티드 영역, TTV 클레이드 3 sle1932 36-뉴클레오티드 영역, TTV 클레이드 4 ctdc002 36-뉴클레오티드 영역, TTV 클레이드 5 36-뉴클레오티드 영역, TTV 클레이드 6 36-뉴클레오티드 영역 또는 TTV 클레이드 7 36-뉴클레오티드 영역)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 바와 같은 36-뉴클레오티드 GC-풍부 서열(예를 들어, 36-뉴클레오티드 공통 GC-풍부 영역 서열 1, 36-뉴클레오티드 공통 GC-풍부 영역 서열 2, TTV 클레이드 1 36-뉴클레오티드 영역, TTV 클레이드 3 36-뉴클레오티드 영역, TTV 클레이드 3 단리주 GH1 36-뉴클레오티드 영역, TTV 클레이드 3 sle1932 36-뉴클레오티드 영역, TTV 클레이드 4 ctdc002 36-뉴클레오티드 영역, TTV 클레이드 5 36-뉴클레오티드 영역, TTV 클레이드 6 36-뉴클레오티드 영역 또는 TTV 클레이드 7 36-뉴클레오티드 영역)의 적어도 10, 15, 20, 25, 30, 31, 32, 33, 34, 35 또는 36개의 연속 뉴클레오티드를 포함하는 핵산 서열을 포함한다.

[0535] 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 예를 들어, 표 39에 열거된 바와 같은, 예를 들어, TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02 또는 TTV-HD16d로부터 선택되는 알파토르크바이러스 GC-풍부 영역 서열에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 예를 들어, 표 39에 열거된 바와 같은, 예를 들어, TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02 또는 TTV-HD16d로부터 선택되는 알파토르크바이러스 GC-풍부 영역 서열의 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 104, 105, 108, 110, 111, 115, 120, 122, 130, 140, 145, 150, 155 또는 156개의 연속 뉴클레오티드를 포함하는 핵산 서열을 포함한다.

[0536] 실시형태들에서, 36-뉴클레오티드 GC-풍부 서열은 하기로부터 선택된다:

[0537] (i) CGCGCTGCGCGCGCCAGTAGGGGAGCCATGC(SEQ ID NO: 160),

[0538] (ii) GCGCTX<sub>1</sub>CGCGCGCGCGCCGGGGGCTGCGCCCCC(SEQ ID NO: 164), 여기서, X<sub>1</sub>은 T, G 또는 A로부터 선택됨;

[0539] (iii) GCGTTCGCGCGCGCCACTAGGGGGCGTTCGCG(SEQ ID NO: 165);

[0540] (iv) GCGTTCGCGCGCGCCAGTAGGGGGCGCAATGCG(SEQ ID NO: 166);

[0541] (v) GCGTTCGCGCGCGCCCGGGGGAGGCATTGCCT(SEQ ID NO: 167);

[0542] (vi) GCGTTCGCGCGCGCGCCGGGGGGCGCCAGCGCC(SEQ ID NO: 168);

[0543] (vii) GCGTTCGCGCGCGCGCGGGGGTTCGCCCCC(SEQ ID NO: 169);

[0544] (viii) GCGTTCGCGCGCGCGCGGGGGGCTGCGCCCCC(SEQ ID NO: 170);

[0545] (ix) GCGTACGCGCGCGCGCGGGGGGCTGCGCCCCC(SEQ ID NO: 171); 또는

[0546] (x) GCGTACGCGCGCGCGCGGGGGGCTGCGCCCCC(SEQ ID NO: 172).

[0547] 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 핵산 서열 CGCGCTGCGCGCGCGCCAGTAGGGGAGCCATGC(SEQ ID NO: 160)를 포함한다.

[0548]

실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 공통 GC-풍부 서열의 핵산 서열을 포함하며,  $X_1$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ ,  $X_{12}$ ,  $X_{13}$ ,  $X_{14}$ ,  $X_{15}$ ,  $X_{20}$ ,  $X_{21}$ ,  $X_{22}$ ,  $X_{26}$ ,  $X_{29}$ ,  $X_{30}$  및  $X_{33}$ 은 각각 독립적으로 임의의 뉴클레오티드이며,  $X_2$ ,  $X_3$ ,  $X_8$ ,  $X_9$ ,  $X_{10}$ ,  $X_{11}$ ,  $X_{16}$ ,  $X_{17}$ ,  $X_{18}$ ,  $X_{19}$ ,  $X_{23}$ ,  $X_{24}$ ,  $X_{25}$ ,  $X_{27}$ ,  $X_{28}$ ,  $X_{31}$ ,  $X_{32}$  및  $X_{34}$ 는 각각 독립적으로 부재이거나, 임의의 뉴클레오티드이다. 일부 실시형태에서,  $X_1$  내지  $X_{34}$  중 하나 이상(예를 들어, 이들 전부)는 각각 독립적으로 표 39에 특정된 뉴클레오티드(또는 부재)이다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 예시적인 TTV GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 내지 3)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 TTV-CT30F GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3, 단편 4, 단편 5, 단편 6, 단편 7, 단편 8 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 내지 7)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 TTV-HD23a GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3, 단편 4, 단편 5, 단편 6 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 내지 6)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 TTV-JA20 GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 및 2)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 TTV-TJN02 GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3, 단편 4, 단편 5, 단편 6, 단편 7, 단편 8 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 내지 8)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 TTV-tth8 GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3, 단편 4, 단편 5, 단편 6, 단편 7, 단편 8, 단편 9 또는 그의 임의의 조합, 예를 들어, 순서대로 단편 1 내지 6)에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 단편 7에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 단편 8에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다. 실시형태들에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타난 단편 9에 대하여 적어도 약 75%(예를 들어, 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100%) 동일성을 갖는 핵산 서열을 포함한다.

[0549] 표 39. 아넬로바이러스로부터의 예시적인 GC-풍부 서열

공급원	서열	SEQ ID NO:
공통	<p>CGGCGGX<sub>1</sub>GGX<sub>2</sub>GX<sub>3</sub>X<sub>4</sub>X<sub>5</sub>CGCGCTX<sub>6</sub>CGCGC                      GCX<sub>7</sub>X<sub>8</sub>X<sub>9</sub>X<sub>10</sub>CX<sub>11</sub>X<sub>12</sub>X<sub>13</sub>X<sub>14</sub>GGGGX<sub>15</sub>X<sub>16</sub>X<sub>17</sub>X<sub>18</sub>                      X<sub>19</sub>X<sub>20</sub>X<sub>21</sub>GCX<sub>22</sub>X<sub>23</sub>X<sub>24</sub>X<sub>25</sub>CCCCCCX<sub>26</sub>CGCGC                      ATX<sub>27</sub>X<sub>28</sub>GCX<sub>29</sub>CGGGX<sub>30</sub>CCCCCCCCX<sub>31</sub>X<sub>32</sub>X<sub>33</sub>GGGGGGCTCCGX<sub>34</sub>CCCCCGGCCCCCC</p> <p>X<sub>1</sub> = G 또는 C                      X<sub>2</sub> = G, C, 또는 부재                      X<sub>3</sub> = C 또는 부재                      X<sub>4</sub> = G 또는 C                      X<sub>5</sub> = G 또는 C                      X<sub>6</sub> = T, G, 또는 A                      X<sub>7</sub> = G 또는 C                      X<sub>8</sub> = G 또는 부재                      X<sub>9</sub> = C 또는 부재                      X<sub>10</sub> = C 또는 부재                      X<sub>11</sub> = G, A, 또는 부재                      X<sub>12</sub> = G 또는 C                      X<sub>13</sub> = C 또는 T</p>	120

[0550]

		<p>X<sub>14</sub> = G 또는 A                  X<sub>15</sub> = G 또는 A                  X<sub>16</sub> = A, G, T, 또는 부재                  X<sub>17</sub> = G, C, 또는 부재                  X<sub>18</sub> = G, C, 또는 부재                  X<sub>19</sub> = C, A, 또는 부재                  X<sub>20</sub> = C 또는 A                  X<sub>21</sub> = T 또는 A                  X<sub>22</sub> = G 또는 C                  X<sub>23</sub> = G, T, 또는 부재                  X<sub>24</sub> = C 또는 부재                  X<sub>25</sub> = G, C, 또는 부재                  X<sub>26</sub> = G 또는 C                  X<sub>27</sub> = G 또는 부재                  X<sub>28</sub> = C 또는 부재                  X<sub>29</sub> = G 또는 A                  X<sub>30</sub> = G 또는 T                  X<sub>31</sub> = C, T, 또는 부재                  X<sub>32</sub> = G, C, A, 또는 부재                  X<sub>33</sub> = G 또는 C                  X<sub>34</sub> = C 또는 부재</p>	
예시적인 TTV 서열	전체 서열	<p>GCCGCCGCGGGCGGCGSGGNGNSGCGCGCT                  DCGCGCGCSNNNCRCCRGGGGGNNNNCWG                  CSNCNCCCCCCCCCGCGCATGCGGGGKCC                  CCCCCCNCGGGGGGCTCCGCCCCCGGC                  CCCCCCGTGCTAAACCCACCGCGCATGC                  GCGACCACGCCCCGCGGCC</p>	121
	단편 1	<p>GCCGCCGCGGGCGGCGSGGNGNSGCGCGCT                  DCGCGCGCSNNNCRCCRGGGGGNNNNCWG                  CSNCNCCCCCCCCCGCGCAT</p>	122
	단편 2	<p>GCGCGGGKCCCCCCCCNCGGGGGGCTC                  CG</p>	123

[0551]

	단편 3	CCCCCGGCCCGCCCGTGCTAAACCCAC CGCGCATGCGCGACCACGCCCCGCGCC	124
TTV-CT30F	전체 서열	GCGGCGG-GGGGGCG-GCCGCG- TTCGCGCGCCGCCACCAGGGGGTG-- CTGCG-CGCCCCCCCCGCGCAT GCGCGGGGCCCCCCCC-- GGGGGGGCTCCGCCCCCGGCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCGCGCC	125
	단편 1	GCGGCGG	126
	단편 2	GGGGGCG	127
	단편 3	GCCGCG	128
	단편 4	TTCGCGCGCCGCCACCAGGGGGTG	129
	단편 5	CTGCG	130
	단편 6	CGCCCCCCCCGCGCAT	131
	단편 7	GCGCGGGGCCCCCCCC	132
	단편 8	GGGGGGGCTCCGCCCCCGGCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCGCGCC	133
TTV-HD23a	전체 서열	CGGCGGCGGCGGCG- CGCGCGCTGCGCGCGCG--- CGCCGGGGGGGCGCCAGCG- CCCCCCCCCGCGCAT GCACGGGTCCCCCCCCACGGGGGGTCC G CCCCCGGCCCCCCCC	134
	단편 1	CGGCGGCGGCGGCG	135
	단편 2	CGCGCGCTGCGCGCGCG	136
	단편 3	CGCCGGGGGGGCGCCAGCG	137
	단편 4	CCCCCCCCCGCGCAT	138
	단편 5	GCACGGGTCCCCCCCCACGGGGGGTCC G	139
	단편 6	CCCCCGGCCCGCCCG	140

[0552]

TTV-JA20	전체 서열	CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCC- CCGGGGGAGGCACAGCCTCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCTCCGGG GGGCTCCGCCCCCGCCCCCCC	141
	단편 1	CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCC	142
	단편 2	CCGGGGGAGGCACAGCCTCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCTCCGGG GGGCTCCGCCCCCGCCCCCCC	143
TTV-TJN02	전체 서열	CGGCGGCGGCG-CGCGCGCTACGCGCGCG-- -CGCCGGGGGG---CTGCCGC- CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCC- GCGGGGGGCTCCG CCCCCGGCCCCCCC	144
	단편 1	CGGCGGCGGCG	145
	단편 2	CGCGCGCTACGCGCGCG	146
	단편 3	CGCCGGGGGG	147
	단편 4	CTGCCGC	148
	단편 5	CCCCCCCCCGCGCAT	149
	단편 6	GCGCGGGGCCCCCCCC	150
	단편 7	GCGGGGGGCTCCG	151
단편 8	CCCCCGGCCCCCCC	152	
TTV-tth8	전체 서열	GCCGCCGCGGCGGGGGG- GCGGCGCGCTGCGCGCGCCAGTAGG GGGAGCCATGCG---CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCC- GCGGGGGGCTCCG CCCCCGGCCCCCCCCG	153
	단편 1	GCCGCCGCGGCGGGGGG	154
	단편 2	GCGGCGCGCTGCGCGCGCCAGTAGG GGGAGCCATGCG	155
	단편 3	CCCCCCCCCGCGCAT	156

[0553]

	단편 4	GCGCGGGGCCCCCCCC	157
	단편 5	GCGGGGGGCTCCG	158
	단편 6	CCCCCGGCCCCCCCG	159
	단편 7	CGCGCTGCGCGCGCCGCCAGTAGGGGA GCCATGC	160
	단편 8	CCGCCATCTTAAGTAGTTGAGGCGGACGGT GGCGTGAGTTCAAAGGTCACCATCAGCCAC ACCTACTCAAATGGTGG	161
	단편 9	CTTAAGTAGTTGAGGCGGACGGTGGCGTGA GTTCAAAGGTCACCATCAGCCACACCTACT CAAATGGTGGACAATTTCTCCGGGTCAA AGGTTACAGCCGCCATGTTAAAACACGTGA CGTATGACGTCACGGCCGCCATTTGTGAC ACAAGATGGCCGACTTCCTCC	162
추가 GC-풍부 서열	36- 뉴클레오티드 공통 GC-풍부 영역 서열 1	CGCGCTGCGCGCGCCGCCAGTAGGGGA GCCATGC	163
	36- 뉴클레오티드 영역 공통 서열 2	GCGCTX <sub>1</sub> CGCGCGCGCGGGGGGCTGCG CCCCCC, 상기 식에서, X <sub>1</sub> 은 T, G, 또는 A 로부터 선택됨	164
	TTV 클레이드 1 36- 뉴클레오티드 영역	GCGCTTCGCGCGCCGCCACTAGGGGGCGT TGCGCG	165
	TTV 클레이드 3 36- 뉴클레오티드 영역	GCGCTGCGCGCGCCGCCAGTAGGGGGCG CAATGCG	166
	TTV 클레이드 3 단리 GH1 36- 뉴클레오티드 영역	GCGCTGCGCGCGGGCCCCGGGGAGGC ATTGCCT	167

[0554]

	TTV 클레이드 3 sle1932 36- 뉴클레오티드 영역	GCGCTGCGCGCGCGCGCCGGGGGGGCGCC AGCGCCC	168
	TTV 클레이드 4 ctdc002 36- 뉴클레오티드 영역	GCGCTTCGCGCGCGCGCCGGGGGGCTCCGC CCCCCC	169
	TTV 클레이드 5 36-뉴클레오티드 영역	GCGCTTCGCGCGCGCGCCGGGGGGCTGCGC CCCCCC	170
	TTV 클레이드 6 36-뉴클레오티드 영역	GCGCTACGCGCGCGCGCCGGGGGGCTGCG CCCCCC	171
	TTV 클레이드 7 36-뉴클레오티드 영역	GCGCTACGCGCGCGCGCCGGGGGGCTCTGC CCCCCC	172
추가 의 알파토르크바이러스 GC-풍부 영역 서열	TTV-CT30F	GCGGCGGGGGGCGGCCGCGTTCGCGCGC CGCCCACCAGGGGGTGTGCGCGCCCCCCC CCGCGCATGCGCGGGGCCCCCCCGGGG GGGCTCCGCCCCCGGCCCCCCCGTGC TAAACCCACCGCGCATGCGGACACGCC CCGCCGCC	801
	TTV-P13-1	CCGAGCGTTAGCGAGGAGTGCACCTACC CCCTGGGCCCCTTCTTCGGAGCCGCGCGC TACGCCTTCGGCTGCGCGGGCACCTCAGA CCCCGCTCGTGCTGACACGCTTGC GCGTG TCAGACCACTTCGGGCTCGGGGGTTCGGG	802
	TTV-tth8	GCCGCCGCGGGCGGGGGGCGGCGCGCT GCGCGCGCCGCCAGTAGGGGAGCCATG CGCCCCCCCCCGCGCATGCGGGGGCCCC CCCCGCGGGGGCTCCGCCCGGCCCCC CCCC	803

[0555]

	TTV-HD20a	CGGCCAGCGGGCGGCGCGCGCTTCGCGC GCGCGCCGGGGGCTCCGCCCCCCCCGCG CATGCGCGGGGCCCCCCCCCGGGGGGCT CCGCCCCCGGTCCCCCCCCG	804
	TTV-16	CGGCCGTGCGGGCGGCGCGCGCTTCGCGC GCGCGCCGGGGGCTGCCGCCCCCCCCGCG CATGCGCGGGGCCCCCCCCCGGGGGG GCTCCGCCCCCGGCCCCCCCCCCG	805
	TTV-TJN02	CGGCGGCGGCGCGCGCTACGCGCGCGC GCCGGGGGCTGCCGCCCCCCCCCGCGCA TGCGGGGGCCCCCCCCCGGGGGGCTCC GCCCCCCGGCCCCC	806
	TTV-HD16d	GGCGGCGGCGCGCGCTACGCGCGCGC CCGGGAGCTCTGCCCCCCCGCGCATGC GCGCGGGTCCCCCCCCCGGGGGGCTCCG CCCCCGGTCCCCCCCCG	807

[0556]

[0557]

**이펙터**

[0558]

일부 실시형태에서, 유전 요소는 이펙터, 예를 들어, 기능적 이펙터, 예를 들어, 내인성 이펙터 또는 외인성 이펙터, 예를 들어, 치료적 폴리펩티드 또는 핵산, 예를 들어, 세포독성 또는 세포용해 RNA 또는 단백질이거나, 이를 인코딩하는 하나 이상의 서열을 포함할 수 있다. 일부 실시형태에서, 기능적 핵산은 비-코딩 RNA이다. 일부 실시형태에서, 기능적 핵산은 코딩 RNA이다. 이펙터는 생물학적 활성을 조절할 수 있으며, 예를 들어, 효소 활성, 유전자 발현, 세포 신호전달 및 세포 또는 기관 기능을 증가시키거나 감소시킬 수 있다. 이펙터 활성은 또한 조절 단백질에 결합하여, 조절제의 활성, 예컨대 전사 또는 번역을 조절하는 것을 포함할 수 있다. 이펙터 활성은 또한, 활성화제 또는 저해제 기능을 포함할 수 있다. 예를 들어, 이펙터는 효소에서 증가된 기질 친화성을 촉발시킴으로써 효소 활성을 유도할 수 있으며, 예를 들어, 프록토스 2,6-비스포스페이트는 포스포프록토키나제 1을 활성화시키고, 인슐린에 반응하여 당분해 속도를 증가시킨다. 또 다른 예에서, 이펙터는 수용체로의 기질 결합을 저해하고, 그의 활성화를 저해할 수 있으며, 예를 들어, 날트렉손(naltrexone) 및 날록손(naloxone)은 활성화시키지 않고 오피오이드 수용체에 결합하며, 오피오이드에 결합하는 수용체의 능력을 차단한다. 이펙터 활성은 또한, 단백질 안정성/분해 및/또는 전사물 안정성/분해를 조절하는 것을 포함할 수 있다. 예를 들어, 단백질은 분해를 위해 폴리펩티드 보조-인자, 유비퀴틴에 의해 단백질 상에 표적화되어, 분해를 위해 그들을 표시할 수 있다. 또 다른 예에서, 이펙터는 효소의 활성 부위를 차단함으로써 효소 활성을 저해하며, 예를 들어, 메토틱세이트는 천연 기질보다 1000배 더 단단하게 다이하이드로폴레이트 환원효소에 결합하고, 뉴클레오티드 염기 합성을 저해하는, 효소 다이하이드로폴레이트 환원효소에 대한 보조효소인 테트라하이드로폴레이트의 구조적 유사체이다.

[0559]

일부 실시형태에서, 이펙터를 인코딩하는 서열은 100 내지 2000개, 100 내지 1000개, 100 내지 500개, 100 내지 200개, 200 내지 2000개, 200 내지 1000개, 200 내지 500개, 500 내지 1000, 500 내지 2000 또는 1000 내지 2000개의 뉴클레오티드를 포함한다. 일부 실시형태에서, 이펙터는 예를 들어, 본원에 기재된 바와 같이 핵산 또는 단백질 페이로드이다.

[0560]

**조절 핵산**

[0561]

일부 실시형태에서, 이펙터는 조절 핵산이다. 조절 핵산은 내인성 유전자 및/또는 외인성 유전자의 발현을 변경시킨다. 일 실시형태에서, 조절 핵산은 숙주 유전자를 표적화한다. 조절 핵산은 내인성 유전자(예를 들어, 본원의 다른 곳에 기재된 바와 같은 miRNA, siRNA, mRNA, lncRNA, RNA, DNA, 안티센스 RNA, gRNA)에 혼성화되는 핵산, 외인성 핵산, 예컨대 바이러스 DNA 또는 RNA에 혼성화하는 핵산, RNA에 혼성화하는 핵산, 유전자 전사를 간섭하는 핵산, RNA 번역을 간섭하는 핵산, RNA를 안정화시키거나, 예컨대 분해에 대한 표적화를 통하여 RNA를 불안정화시키는 핵산, 및 DNA 또는 RNA 결합 인자를 조절하는 핵산을 포함할 수 있지만, 이에 제한되지 않는다. 실시형태들에서, 조절 핵산은 miRNA를 인코딩한다. 일부 실시형태에서, 조절 핵산은 야생형 아벨로바이러스에 대하여 내인성이다. 일부 실시형태에서, 조절 핵산은 야생형 아벨로바이러스에 대하여 외인성이다.

- [0562] 일부 실시형태에서, 조절 핵산은 전형적으로 5 내지 500개의 염기쌍(특정 RNA 구조에 따라, 예를 들어, miRNA 5 내지 30 bp, lncRNA 200 내지 500 bp)을 함유하는 RNA 또는 RNA-유사 구조를 포함하며, 세포 내의 발현된 표적 유전자의 코딩 서열 또는 세포 내의 발현된 표적 유전자를 인코딩하는 서열과 동일하거나(또는 상보성이거나) 거의 동일한(또는 실질적으로 상보성인) 핵염기 서열을 가질 수 있다.
- [0563] 일부 실시형태에서, 조절 핵산은 핵산 서열, 예를 들어, 가이드 RNA(gRNA)를 포함한다. 일부 실시형태에서, DNA 표적화 모이어티는 가이드 RNA 또는 가이드 RNA를 인코딩하는 핵산을 포함한다. gRNA 짧은 합성 RNA는 불완전한 이펙터 모이어티로의 결합에 필요한 "스캐폴드" 서열 및 게놈 표적에 대한 사용자-정의된 약 20개의 뉴클레오티드 표적화 서열로 구성될 수 있다. 실시에서, 가이드 RNA 서열은 일반적으로 17 내지 24개의 뉴클레오티드(예를 들어, 19, 20 또는 21개의 뉴클레오티드)의 길이를 갖도록 설계되고 표적화된 핵산 서열에 대하여 상보성이다. 커스텀 gRNA 생성자 및 알고리즘은 효율적인 가이드 RNA의 설계에 사용하기 위하여 상업적으로 이용 가능하다. 유전자 편집은 또한, 자연 발생 crRNA-tracrRNA 복합체를 모방하고 tracrRNA(뉴클레아제에 결합) 및 적어도 하나의 crRNA(뉴클레아제를 편집에 대해 표적화된 서열로 가이드함) 둘 모두를 함유하는 엔지니어링된(합성) 단일 RNA 분자인, 키메라 "단일 가이드 RNA"("sgRNA")를 사용하여 달성되었다. 화학적으로 변형된 sgRNA는 또한, 게놈 편집에서 효율적인 것으로 입증되었으며; 예를 들어, 문헌[Hendel et al. (2015) Nature Biotechnol., 985 - 991]을 참조한다.
- [0564] 조절 핵산은 특정 DNA 서열(예를 들어, 프로모터, 유전자의 인핸서, 침묵자 또는 억제자에 인접하거나 그 내의 서열)을 인식하는 gRNA를 포함한다.
- [0565] 특정 조절 핵산은 RNA 간섭(RNAi)의 생물학적 프로세스를 통하여 유전자 발현을 저해할 수 있다. RNAi 분자는 전형적으로 15 내지 50개의 염기쌍(예컨대 약 18 내지 25개의 염기쌍)을 함유하고, 세포 내의 발현된 표적 유전자의 코딩 서열과 동일하거나(상보성이거나) 또는 거의 동일한(실질적으로 상보성인) 핵염기 서열을 갖는 RNA 또는 RNA-유사 구조를 포함한다. RNAi 분자는 짧은 간섭 RNA(siRNA), 이중-가닥 RNA(dsRNA), 마이크로 RNA(miRNA), 짧은 헤어핀 RNA(shRNA), 메로듀플렉스(meroduplex) 및 다이서(dicer) 기질(미국 특허 제 8,084,599호, 제8,349,809호 및 제8,513,207호)을 포함하나 이들에 제한되지 않는다.
- [0566] 긴 비-코딩 RNA(lncRNA)는 100개 초과 뉴클레오티드의 비-단백질 코딩 전사물로서 정의된다. 이러한 다소 임의의 제한은 lncRNA를 소형 조절 RNA, 예컨대 마이크로RNA(miRNA), 짧은 간섭 RNA(siRNA) 및 다른 짧은 RNA로부터 구별한다. 일반적으로, 대다수(약 78%)의 lncRNA는 조직-특이적인 것으로 특성화된다. 근처의 단백질-코딩 유전자에 대하여 반대 방향으로 전사되는(포유동물 게놈 내에 총 lncRNA의 유의미한 비율 약 20%를 포함하는) 분지 lncRNA는 아마도 근처의 유전자의 전사를 조절할 수 있다.
- [0567] 유전 요소는 내인성 유전자 또는 유전자 생성물(예를 들어, mRNA)의 전부 또는 그의 단편에 실질적으로 상보성이거나, 완전히 상보성인 서열을 갖는 조절 핵산을 인코딩할 수 있다. 조절 핵산은 인트론과 엑손 사이의 경계에서의 서열에 상보성이어서, 특정 유전자의 새로 생성된 핵 RNA 전사물이 전사를 위해 mRNA로 성숙되는 것을 방지할 수 있다. 특정 유전자에 상보성인 조절 핵산은 당해 유전자에 대한 mRNA와 혼성화하고, 그의 번역을 방지할 수 있다. 안티센스 조절 핵산은 DNA, RNA 또는 그의 유도체 또는 혼성물일 수 있다.
- [0568] 관심 전사물에 혼성화하는 조절 핵산의 길이는 5 내지 30개의 뉴클레오티드, 약 10 내지 30개의 뉴클레오티드 또는 약 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 이상의 뉴클레오티드일 수 있다. 표적화된 전사물에 대한 조절 핵산의 동일성의 정도는 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 또는 적어도 95%여야 한다.
- [0569] 유전 요소는 조절 핵산, 예를 들어, 표적 유전자의 약 5 내지 약 25개의 인접 뉴클레오티드와 동일한 마이크로 RNA(miRNA) 분자를 인코딩할 수 있다. 일부 실시형태에서, miRNA 서열은 mRNA를 표적화하며, 다이뉴클레오티드 AA로 시작하며, 약 30 내지 70%(약 30 내지 60%, 약 40 내지 60% 또는 약 45% 내지 55%)의 GC-함량을 포함하며, 예를 들어, 표준 BLAST 검색에 의해 결정시, 그것이 도입된 포유동물의 게놈 내의 표적 이외의 임의의 뉴클레오티드 서열에 대하여 높은 동일성 백분율을 갖지 않는다.
- [0570] siRNA 및 shRNA는 내인성 마이크로RNA(miRNA) 유전자의 가공 경로 내의 중간체와 유사하다(문헌[Bartel, Cell 116:281-297, 2004]). 일부 실시형태에서, siRNA는 miRNA로서 기능할 수 있으며, 그 역도 그러하다(문헌[Zeng et al., Mol Cell 9:1327-1333, 2002]; 문헌[Doench et al., Genes Dev 17:438-442, 2003]). siRNA와 같이 마이크로RNA는 RISC를 사용하여, 표적 유전자를 하향조절하지만, siRNA와 달리, 대부분의 동물 miRNA는 mRNA를 절단하지 않는다. 대신에, miRNA는 번역 억제 또는 폴리A 제거 및 mRNA 분해를 통해 단백질 출력물을 감소시킨다

(문헌[Wu et al., Proc Natl Acad Sci USA 103:4034-4039, 2006]). 알려진 miRNA 결합 부위는 mRNA 3' UTR 내에 존재하며; miRNA는 miRNA의 5' 말단으로부터 뉴클레오티드 2 내지 8에 대하여 거의 완벽하게 상보성을 갖는 부위를 표적화하는 것으로 보인다(문헌[Rajewsky, Nat Genet 38 Suppl:S8-13, 2006]; 문헌[Lim et al., Nature 433:769-773, 2005]). 이러한 영역은 씨드 영역으로 공지되어 있다. siRNA 및 miRNA가 상호 교환 가능하기 때문에, 외인성 siRNA는 siRNA에 대하여 씨드 상보성을 갖는 mRNA를 하향조절한다(문헌[Birmingham et al., Nat Methods 3:199-204, 2006]). 3' UTR 내의 다수의 표적 부위는 더 강한 하향조절을 제공한다(문헌[Doench et al., Genes Dev 17:438-442, 2003]).

[0571] 알려진 miRNA 서열의 목록은 특히 Wellcome Trust Sanger Institute, Penn Center for Bioinformatics, Memorial Sloan Kettering Cancer Center, 및 European Molecule Biology Laboratory와 같은 연구 기관에 의해 유지되는 데이터베이스에서 확인될 수 있다. 또한, 알려진 효과적인 siRNA 서열 및 동족(cognate) 결합 부위는 관련 문헌에서 잘 나타나 있다. RNAi 분자는 해당 분야에 알려진 기술에 의해 쉽게 설계되고 생성된다. 또한, 효과적이고 특이적인 서열 모티프를 찾는 기회를 증가시키는 컴퓨터 기반 도구가 존재한다(문헌[Lagana et al., Methods Mol. Bio., 2015, 1269:393-412]).

[0572] 조절 핵산은 유전자에 의해 인코딩된 RNA의 발현을 조절할 수 있다. 다수의 유전자가 서로 어느 정도의 서열 상동성을 공유할 수 있기 때문에, 일부 실시형태에서, 조절 핵산은 충분한 서열 상동성을 갖는 유전자의 부류를 표적화하도록 설계될 수 있다. 일부 실시형태에서, 조절 핵산은 상이한 유전자 표적 간에 공유되는 또는 특정 유전자 표적에 특유한 서열에 상보성을 갖는 서열을 함유할 수 있다. 일부 실시형태에서, 조절 핵산은 몇몇의 유전자 간에 상동성을 갖는 RNA 서열의 보존된 영역을 표적화하여, 그에 의해, 유전자 과 내의 몇몇의 유전자 (예를 들어, 상이한 유전자 아이소폼, 슬라이스 변이체, 돌연변이 유전자 등)를 표적화하도록 설계될 수 있다. 일부 실시형태에서, 조절 핵산은 단일 유전자의 특정 RNA 서열에 특유한 서열을 표적화하도록 설계될 수 있다.

[0573] 일부 실시형태에서, 유전 요소는 하나 이상의 유전자의 발현을 조절하는 조절 핵산을 인코딩하는 하나 이상의 서열을 포함할 수 있다.

[0574] 일 실시형태에서, 본원의 다른 곳에 기재된 gRNA는 유전자 편집을 위한 CRISPR 시스템의 부분으로서 사용된다. 유전자 편집의 목적을 위하여, 아넬로벡터는 요망되는 표적 DNA 서열에 상응하는 하나의 또는 다수의 가이드 RNA 서열을 포함하도록 설계될 수 있으며; 예를 들어, 문헌[Cong et al. (2013) Science, 339:819-823]; 문헌[Ran et al. (2013) Nature Protocols, 8:2281 - 2308]을 참조한다. 적어도 약 16 또는 17개의 뉴클레오티드의 gRNA 서열은 일반적으로 Cas9-매개된 DNA 절단이 발생하게 하며; Cpf1에 있어서, 적어도 약 16개의 뉴클레오티드의 gRNA 서열이 검출 가능한 DNA 절단을 달성하는 데 필요하다.

[0575] *치료적 이펙터(예를 들어, 펩티드 또는 폴리펩티드)*

[0576] 일부 실시형태에서, 유전 요소는 치료적 발현 서열, 예를 들어, 치료적 펩티드 또는 폴리펩티드, 예를 들어, 세포내 펩티드 또는 세포내 폴리펩티드, 분비 폴리펩티드 또는 단백질 대체 치료제를 인코딩하는 서열을 포함한다. 일부 실시형태에서, 유전 요소는 단백질, 예를 들어, 치료적 단백질을 인코딩하는 서열을 포함한다. 치료적 단백질의 일부 예는 호르몬, 사이토카인, 효소, 항체(예를 들어, 적어도 중쇄 또는 경쇄를 인코딩하는 하나 또는 복수의 폴리펩티드), 전사 인자, 수용체(예를 들어, 막 수용체), 리간드, 막 수송체, 분비 단백질, 펩티드, 담체 단백질, 구조 단백질, 뉴클레아제 또는 그의 구성성분을 포함할 수 있으나, 이에 제한되지 않는다.

[0577] 일부 실시형태에서, 유전 요소는 펩티드, 예를 들어, 치료적 펩티드를 인코딩하는 서열을 포함한다. 펩티드는 선형 또는 분지형일 수 있다. 펩티드는 약 5 내지 약 500개의 아미노산, 약 15 내지 약 400개의 아미노산, 약 20 내지 약 325개의 아미노산, 약 25 내지 약 250개의 아미노산, 약 50 내지 약 200개의 아미노산 또는 그 사이의 임의의 범위의 길이를 갖는다.

[0578] 일부 실시형태에서, 치료적 발현 서열에 의해 인코딩된 폴리펩티드는 위의 것들 중 임의의 기능적 변이체 또는 그의 단편, 예를 들어, 그의 UniProt ID를 참조하여 본원의 표에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질일 수 있다.

[0579] 일부 실시형태에서, 치료적 발현 서열은 위의 것들 중 임의의 것에 결합하는 항체 또는 항체 단편, 예를 들어, 그의 UniProt ID를 참조하여 본원의 표에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질에 대한 항체를 인코딩할 수 있다. 본원에서 용어 "항체"는 가장 넓은 의미로 사용되며, 이들이 요망되는 항원-결합 활성을 나타내는 한, 비제한적으로 단클론 항체, 다클론 항체, 다중특이적

항체(예를 들어, 이중특이적 항체) 및 항체 단편을 포함하는 다양한 항체 구조를 포괄한다. "항체 단편"은 적어도 하나의 중쇄 또는 경쇄를 포함하고 항원에 결합하는 분자를 지칭한다. 항체 단편의 예는 Fv, Fab, Fab', Fab'-SH, F(ab')<sub>2</sub>; 디아바디; 선형 항체; 단일-쇄 항체 분자(예를 들어, scFv); 및 항체 단편으로부터 형성된 다중특이적 항체를 포함하나, 이에 제한되지 않는다.

[0580] **예시적인 세포내 폴리펩티드 이펙터**

[0581] 일부 실시형태에서, 이펙터는 사이토졸 폴리펩티드 또는 사이토졸 펩티드를 포함한다. 일부 실시형태에서, 이펙터는 DPP-4 억제제, GLP-1 신호전달의 활성화제 또는 호중구 엘라스타제의 억제제인 사이토졸 펩티드를 포함한다. 일부 실시형태에서, 이펙터는 성장 인자 또는 그의 수용체(예를 들어, FGF 수용체, 예를 들어, FGFR3)의 수준 또는 활성을 증가시킨다. 일부 실시형태에서, 이펙터는 n-myc 상호작용 단백질 활성화의 억제제(예를 들어, n-myc 상호작용 단백질 억제제); EGFR 활성화의 억제제(예를 들어, EGFR 억제제); IDH1 및/또는 IDH2 활성화의 억제제(예를 들어, IDH1 억제제 및/또는 IDH2 억제제); LRP5 및/또는 DKK2 활성화의 억제제(예를 들어, LRP5 및/또는 DKK2 억제제); KRAS 활성화의 억제제; HTT 활성화의 활성화제; 또는 DPP-4 활성화의 억제제(예를 들어, DPP-4 억제제)를 포함한다.

[0582] 일부 실시형태에서, 이펙터는 조절 세포내 폴리펩티드를 포함한다. 일부 실시형태에서, 조절 세포내 폴리펩티드는 표적 세포에 대하여 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)에 결합한다. 일부 실시형태에서, 조절 세포내 폴리펩티드는 표적 세포에 대하여 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)의 수준 또는 활성을 증가시킨다. 일부 실시형태에서, 조절 세포내 폴리펩티드는 표적 세포에 대하여 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)의 수준 또는 활성을 감소시킨다.

[0583] **예시적인 분비 폴리펩티드 이펙터**

[0584] 예시적인 분비 치료제는 본원에서, 예를 들어, 하기 표에 기재되어 있다.

[0585] **표 50. 예시적인 사이토카인 및 사이토카인 수용체**

사이토카인	사이토카인 수용체(들)	Entrez 유전자 ID	UniProt ID
IL-1 $\alpha$ , IL-1 $\beta$ , 또는 그의 이종이합체	IL-1 유형 1 수용체, IL-1 유형 2 수용체	3552, 3553	P01583, P01584
IL-1Ra	IL-1 유형 1 수용체, IL-1 유형 2 수용체	3454, 3455	P17181, P48551
IL-2	IL-2R	3558	P60568
IL-3	IL-3 수용체 $\alpha + \beta c$ (CD131)	3562	P08700
IL-4	IL-4R 유형 I, IL-4R 유형 II	3565	P05112
IL-5	IL-5R	3567	P05113
IL-6	IL-6R (sIL-6R) gp130	3569	P05231
IL-7	IL-7R 및 sIL-7R	3574	P13232

[0586]

IL-8	CXCR1 및 CXCR2	3576	P10145
IL-9	IL-9R	3578	P15248
IL-10	IL-10R1/IL-10R2 복합체	3586	P22301
IL-11	IL-11R $\alpha$ 1 gp130	3589	P20809
IL-12 (예를 들어, p35, p40, 또는 그의 이종이합체)	IL-12R $\beta$ 1 및 IL-12R $\beta$ 2	3593, 3592	P29459, P29460
IL-13	IL-13R1 $\alpha$ 1 및 IL-13R1 $\alpha$ 2	3596	P35225
IL-14	IL-14R	30685	P40222
IL-15	IL-15R	3600	P40933
IL-16	CD4	3603	Q14005
IL-17A	IL-17RA	3605	Q16552
IL-17B	IL-17RB	27190	Q9UHF5
IL-17C	IL-17RA 내지 IL-17RE	27189	Q9P0M4
e	SEF	53342	Q8TAD2
IL-17F	IL-17RA, IL-17RC	112744	Q96PD4
IL-18	IL-18 수용체	3606	Q14116
IL-19	IL-20R1/IL-20R2	29949	Q9UHD0
IL-20	L-20R1/IL-20R2 및 IL-22R1/ IL-20R2	50604	Q9NYY1
IL-21	IL-21R	59067	Q9HBE4
IL-22	IL-22R	50616	Q9GZX6
IL-23 (예를 들어, p19, p40, 또는 그의 이종이합체)	IL-23R	51561	Q9NPF7
IL-24	IL-20R1/IL-20R2 및 IL-22R1/ IL-20R2	11009	Q13007
IL-25	IL-17RA 및 IL-17RB	64806	Q9H293
IL-26	IL-10R2쇄 및 IL-20R1쇄	55801	Q9NPH9
IL-27 (예를 들어, p28, EB13, 또는 그의 이종이합체)	WSX-1 및 gp130	246778	Q8NEV9
IL-28A, IL-28B, 및 IL29	IL-28R1/IL-10R2	282617, 282618	Q8IZI9, Q8IU54
IL-30	IL6R/gp130	246778	Q8NEV9
IL-31	IL-31RA/OSMR $\beta$	386653	Q6EBC2

[0587]

IL-32		9235	P24001
IL-33	ST2	90865	O95760
IL-34	코로나-자극 인자 1 수용체	146433	Q6ZMJ4
IL-35 (예를 들어, p35, EB13, 또는 그의 이종이합체)	IL-12R $\beta$ 2/gp130; IL- 12R $\beta$ 2/IL-12R $\beta$ 2; gp130/gp130	10148	Q14213
IL-36	IL-36Ra	27179	Q9UHA7
IL-37	IL-18R $\alpha$ 및 IL-18BP	27178	Q9NZH6
IL-38	IL-1R1, IL-36R	84639	Q8WWZ1
IFN- $\alpha$	IFNAR	3454	P17181
IFN- $\beta$	IFNAR	3454	P17181
IFN- $\gamma$	IFNGR1/IFNGR2	3459	P15260
TGF- $\beta$	T $\beta$ R-I 및 T $\beta$ R-II	7046, 7048	P36897, P37173
TNF- $\alpha$	TNFR1, TNFR2	7132, 7133	P19438, P20333

[0588]

[0589] 일부 실시형태에서, 본원에 기재된 이펙터는 표 50의 사이토카인, 또는 그의 기능적 변이체, 예를 들어, 그의 상동체(예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 50에 열거된 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 96%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대해 상응하는 야생형 사이토카인의 Kd보다 10%, 20%, 30%, 40% 또는 50% 이하로 더 높거나 더 낮은 Kd로 상응하는 사이토카인 수용체에 결합한다. 일부 실시형태에서, 이펙터는 제1 영역(예를 들어, 표 50의 사이토카인 폴리펩티드 또는 그의 기능적 변이체 또는 단편) 및 제2, 이종 영역을 포함하는 융합 단백질을 포함한다. 일부 실시형태에서, 제1 영역은 표 50의 제1 사이토카인 폴리펩티드이다. 일부 실시형태에서, 제2 영역은 표 50의 제2 사이토카인 폴리펩티드이며, 여기서 제1 및 제2 사이토카인 폴리펩티드는 야생형 세포에서 서로 사이토카인 이종이합체를 형성한다. 일부 실시형태에서, 표 50의 폴리펩티드 또는 그의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대하여 내인성인 신호 서열 또는 이종 신호 서열을 포함한다. 일부 실시형태에서, 표 50의 사이토카인을 인코딩하는 아넬로백터 또는 그의 기능적 변이체는 본원에 기재된 질병 또는 장애의 치료에 사용된다.

[0590] 일부 실시형태에서, 본원에 기재된 이펙터는 표 50의 사이토카인에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 표 50의 사이토카인 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 실시형태에서, 항체 분자는 신호 서열을 포함한다.

[0591] 예시적인 사이토카인 및 사이토카인 수용체는 예를 들어, 그 안의 표 I를 포함하여 그의 전문이 본원에 참조로 포함되는 문헌[Akdis et al., "Interleukins (from IL-1 to IL-38), interferons, transforming growth factor  $\beta$ , and TNF- $\alpha$ : Receptors, functions, and roles in diseases" October 2016 Volume 138, Issue 4, Pages 984-1010]에 기재되어 있다.

[0592] 표 51. 예시적인 폴리펩티드 호르몬 및 수용체

호르몬	수용체	Entrez 유전자 ID	UniProt ID
나트륨이노 펩티드, 예를 들어, 심방 나트륨이노 펩티드(ANP)	NPRA, NPRB, NPRC	4878	P01160
뇌 나트륨이노 펩티드(BNP)	NPRA, NPRB	4879	P16860
C-유형 나트륨이노 펩티드 (CNP)	NPRB	4880	P23582
성장 호르몬(GH)	GHR	2690	P10912
인간 성장 호르몬(hGH)	hGH 수용체 (인간 GHR)	2690	P10912
프로락틴 (PRL)	PRLR	5617	P01236
갑상선-자극 호르몬 (TSH)	TSH 수용체	7253	P16473
부신피질자극호르몬 (ACTH)	ACTH 수용체	5443	P01189
난포 자극 호르몬 (FSH)	FSHR	2492	P23945
황체형성 호르몬(LH)	LHR	3973	P22888
항이노 호르몬(ADH)	바소프레신 수용체, 예를 들어, V2; AVPR1A; AVPR1B; AVPR3; AVPR2	554	P30518
옥시토신	OXTR	5020	P01178
칼시토닌	칼시토닌 수용체(CT)	796	P01258
부갑상선 호르몬(PTH)	PTH1R 및 PTH2R	5741	P01270
인슐린	인슐린 수용체(IR)	3630	P01308
글루카곤	글루카곤 수용체	2641	P01275

[0593]

[0594] 일부 실시형태에서, 본원에 기재된 이펙터는 표 51의 호르몬, 또는 그의 기능적 변이체, 예를 들어, 그의 상동체(예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 51에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대해 상응하는 야생형 호르몬의 Kd보다 10%, 20%, 30%, 40% 또는 50% 이하 더 높은 Kd로 상응하는 수용체에 결합한다. 일부 실시형태에서, 표 51의 폴리펩티드 또는 그의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대하여 내인성인 신호 서열 또는 이중 신호 서열을 포함한다. 일부 실시형태에서, 표 51의 호르몬을 인코딩하는 아넬로벡터 또는 그의 기능적 변이체는 본원에 기재된 질병 또는 장애의 치료에 사용된다.

[0595] 일부 실시형태에서, 본원에 기재된 이펙터는 표 51의 호르몬에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 표 51의 호르몬 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 실시형태에서, 항체 분자는 신호 서열을 포함한다.

[0596] 표 52. 예시적인 성장 인자

성장 인자		Entrez 유전자 ID	UniProt ID
PDGF 패밀리			
PDGF (예를 들어, PDGF-1, PDGF-2, 또는 그의 이종이합체)	PDGF 수용체, 예를 들어, PDGFR $\alpha$ , PDGFR $\beta$	5156	P16234
CSF-1	CSF1R	1435	P09603
SCF	CD117	3815	P10721
VEGF 패밀리			
VEGF (예를 들어, 아이소폼 VEGF 121,	VEGFR-1, VEGFR-2	2321	P17948

[0597]

VEGF 165, VEGF 189, 및 VEGF 206)			
VEGF-B	VEGFR-1	2321	P17949
VEGF-C	VEGFR-2 및 VEGFR -3	2324	P35916
PlGF	VEGFR-1	5281	Q07326
EGF 패밀리			
EGF	EGFR	1950	P01133
TGF- $\alpha$	EGFR	7039	P01135
암피레굴린	EGFR	374	P15514
HB-EGF	EGFR	1839	Q99075
베타셀룰린	EGFR, ErbB-4	685	P35070
에피레굴린	EGFR, ErbB-4	2069	O14944
헤레굴린	EGFR, ErbB-4	3084	Q02297
FGF 패밀리			
FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7, FGF-8, FGF-9	FGFR1, FGFR2, FGFR3, 및 FGFR4	2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254	P05230, P09038, P11487, P08620, P12034, P10767, P21781, P55075, P31371
인슐린 패밀리			
인슐린	IR	3630	P01308
IGF-I	IGF-I 수용체, IGF-II 수용체	3479	P05019
IGF-II	IGF-II 수용체	3481	P01344
HGF 패밀리			
HGF	MET 수용체	3082	P14210
MSP	RON	4485	P26927
뉴로트로핀 패밀리			
NGF	LNGFR, trkA	4803	P01138
BDNF	trkB	627	P23560
NT-3	trkA, trkB, trkC	4908	P20783
NT-4	trkA, trkB	4909	P34130
NT-5	trkA, trkB	4909	P34130

[0598]

안지오펜티에틴 패밀리			
ANGPT1	HPK-6/TEK	284	Q15389
ANGPT2	HPK-6/TEK	285	O15123
ANGPT3	HPK-6/TEK	9068	O95841
ANGPT4	HPK-6/TEK	51378	Q9Y264

[0599]

[0600]

일부 실시형태에서, 본원에 기재된 이펙터는 표 52의 성장 인자, 또는 그의 기능적 변이체, 예를 들어, 그의 상동체(예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 52에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대해 반응하는 야생형 성장 인자의 Kd보다 10%, 20%, 30%, 40% 또는 50% 이하 더 높은 Kd로 반응하는 수용체에 결합한다. 일부 실시형태에서, 표 52의 폴리펩티드 또는 그의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대하여 내인성인 신호 서열 또는 이중 신호 서열을 포함한다. 일부 실시형태에서, 표 52의 성장 인자를 인코딩하는 아빌로벡터 또는 그의 기능적 변이체는 본원에 기재된 질병 또는 장애의 치료에 사용된다.

[0601]

일부 실시형태에서, 본원에 기재된 이펙터는 표 52의 성장 인자에 결합하는 항체 분자(예를 들어, scFv)를 포함

한다. 일부 실시형태에서, 본원에 기재된 이펙터는 표 52의 성장 인자 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 실시형태에서, 항체 분자는 신호 서열을 포함한다.

[0602] 예시적인 성장 인자 및 성장 인자 수용체는 예를 들어, 그의 전문이 본원에 참조로 포함되는 문헌[Bafico et al., "Classification of Growth Factors and Their Receptors" Holland-Frei Cancer Medicine. 6th edition]에 기재되어 있다.

[0603] 표 53. 응고-연관된 인자

이펙터	적응증	Entrez 유전자 ID	UniProt ID
인자 I (피브리노겐)	무피브리노겐혈증	2243, 2266, 2244	P02671, P02679, P02675
인자 II	인자 II 결핍증	2147	P00734
인자 IX	혈우병 B	2158	P00740
인자 V	오우렌병	2153	P12259
인자 VIII	혈우병 A	2157	P00451
인자 X	스튜어트-프록워 인자 결핍증	2159	P00742
인자 XI	혈우병 C	2160	P03951
인자 XIII	피브리린 안정화 인자 결핍증	2162, 2165	P00488, P05160
vWF	폰빌레브란트병	7450	P04275

[0604]

[0605] 일부 실시형태에서, 본원에 기재된 이펙터는 표 53의 폴리펩티드, 또는 그의 기능적 변이체, 예를 들어, 그의 상동체(예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 53에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 기능적 변이체는 상응하는 야생형 단백질과 동일한 반응을, 예를 들어, 야생형 단백질보다 10%, 20%, 30%, 40% 또는 50% 이하의 더 낮은 속도로 촉매한다. 일부 실시형태에서, 표 53의 폴리펩티드 또는 그의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대하여 내인성인 신호 서열 또는 이중 신호 서열을 포함한다. 일부 실시형태에서, 표 53의 폴리펩티드 또는 그의 기능적 변이체를 인코딩하는 아넬로벡터는 표 53의 질병 또는 장애의 치료에 사용된다.

[0606] 예시적인 단백질 대체 치료제

[0607] 예시적인 단백질 대체 치료제는 본원에, 예를 들어, 하기 표에 기재되어 있다.

[0608] 표 54. 예시적인 효소 이펙터 및 상응하는 적응증

이펙터	결핍증	Entrez 유전자 ID	UniProt ID
3-메틸크로토닐-CoA 카복실라제	3-메틸크로토닐-CoA 카복실라제 결핍증	56922, 64087	Q96RQ3, Q9HCC0
아세틸-CoA-글루코사미나이드 N-아세틸트랜스퍼라제	점액다당류증 MPS III (산필리포증후군) 유형 III-C	138050	Q68CP4
ADAMTS13	혈전성 혈소판감소성 자반증	11093	Q76LX8
아데닌 포스포리보실트랜스퍼라제	아데닌 포스포리보실트랜스퍼라제 결핍증	353	P07741

[0609]

아데노신 데아미나제	아데노신 데아미나제 결핍증	100	P00813
ADP-리보스 단백질 가수분해 효소	글루타밀 리보스-5-포스페이트 축적병	26119, 54936	Q5SW96, Q9NX46
알파 글루코시다제	글리코겐 축적병 유형 2 (폼페병)	2548	P10253
아르기나제	가족성 아르기닌혈증	383, 384	P05089, P78540
아릴설파타제 A	이염성 백질이영양증	410	P15289
카텝신 K	피크노디소토시스	1513	P43235
세라미다제	파버병(지방육아종증)	125981, 340485, 55331	Q8TDN7, Q5QJU3, Q9NUN7
시스타티오닌 B 신타제	호모시스틴뇨증	875	P35520
돌리콜-P-만노스 신타제	N-글리코실화 CDG Ic의 선천성 장애	8813, 54344	O60762, Q9P2X0
돌리코-P-Glc:Man9GlcNAc2-PP-돌리콜 글루코실트랜스퍼라제	N-글리코실화 CDG Ic의 선천성 장애	84920	Q5BKT4
돌리코-P-Man:Man5GlcNAc2-PP-돌리콜 만노실트랜스퍼라제	N-글리코실화 CDG Id의 선천성 장애	10195	Q92685
돌리칠-P-글루코스:Glc-1-Man-9-GlcNAc-2-PP-돌리칠- $\alpha$ -3-글루코실트랜스퍼라제	N-글리코실화 CDG Ih의 선천성 장애	79053	Q9BVK2
돌리칠-P-만노스:Man-7-GlcNAc-2-PP-돌리칠- $\alpha$ -6-만노실트랜스퍼라제	N-글리코실화 CDG Ig의 선천성 장애	79087	Q9BV10
인자 II	인자 II 결핍증	2147	P00734

[0610]

인자 IX	혈우병 B	2158	P00740
인자 V	오우렌병	2153	P12259
인자 VIII	혈우병 A	2157	P00451
인자 X	스튜어트-프로워 인자 결핍증	2159	P00742
인자 XI	혈우병 C	2160	P03951
인자 XIII	피브리린 안정화 인자 결핍증	2162, 2165	P00488, P05160
갈락토사민-6-설페이트 설파타제	점액다당류증 MPS IV (모르퀴오 증후군) 유형 IV-A	2588	P34059
갈락토실세라미드 β- 갈락토시다제	크라베병	2581	P54803
강글리오시드 β- 갈락토시다제	GM1 강글리오시드증, 전신	2720	P16278
강글리오시드 β- 갈락토시다제	GM2 강글리오시드증	2720	P16278
강글리오시드 β- 갈락토시다제	스핑고지질증 유형 I	2720	P16278
강글리오시드 β- 갈락토시다제	스핑고지질증 유형 II (소아 유형)	2720	P16278
강글리오시드 β- 갈락토시다제	스핑고지질증 유형 III (성인 유형)	2720	P16278
글루코시다제 I	N-글리코실화 CDG IIb 의 선천성 장애	2548	P10253
글루코실세라미드 β- 글루코시다제	고세병	2629	P04062
해파란-S-설페이트 설파미다제	점액다당류증 MPS III (산필리포 증후군) 유형 III-A	6448	P51688
호모겐티세이트 산화효소	알칼톤뇨증	3081	Q93099
히알루로니다제	점액다당류증 MPS IX(히알루로니다제 결핍증)	3373, 8692, 8372, 23553	Q12794, Q12891, O43820, Q2M3T9

[0611]

이두로네이트 설페이트 설페타제	점액다당류증 MPS II(헌터 증후군)	3423	P22304
레시틴-콜레스테롤 아실트랜스퍼라제(LCAT)	완전 LCAT 결핍증, 물고기-눈 질병 (Fish-eye disease), 죽상경화증, 고콜레스테롤혈증	3931	606967
리신 산화효소	글루타르산혈증 유형 I	4015	P28300
리소좀산 리파제	콜레스테롤 에스테르 축적병(CESD)	3988	P38571
리소좀산 리파제	리소좀산 리파제 결핍증	3988	P38571
리소좀산 리파제	월만병	3988	P38571
리소좀 펩티다제 펩티다제	세로이드 리포푸신증 후기 영아형(CLN2, 안스키- 빌쇼스키 (Jansky- Bielschowsky) 질병)	1200	O14773
만노스(Man) 포스페이트(P) 이소머라제	N-글리코실화 CDG Ib 의 선천성 장애	4351	P34949
만노실- $\alpha$ -1,6-당단백질- $\beta$ - 1,2-N- 아세틸글루코시미닐트랜스 퍼라제	N-글리코실화 CDG IIa 의 선천성 장애	4247	Q10469
메탈로프로테이나제-2	원체스터 증후군	4313	P08253
메틸말로닐 -CoA 뮤타제	메틸말론산혈증(비타민 b12 비- 반응성)	4594	P22033
N-아세틸 갈락토사민 $\alpha$ -4- 설페이트 설페타제(아릴설페타제 B)	점액다당류증 MPS VI(마로토- 라미 증후군)	411	P15848
N-아세틸-D- 글루코사미니다제	점액다당류증 MPS III(산필리포 증후군) 유형 III-B	4669	P54802

[0612]

N-아세틸-갈락토사미니다제	신들러병 유형 I (영아 중증형)	4668	P17050
N-아세틸-갈락토사미니다제	신들러병 유형 II (칸자키병, 성인-발병형)	4668	P17050
N-아세틸-갈락토사미니다제	신들러병 유형 III (중간형)	4668	P17050
N-아세틸-글루코사미닌-6-설페이트 설파타제	점액다당류증 MPS III(산필리포 증후군) 유형 III-D	2799	P15586
N-아세틸글루코사미닐-1-포스포트랜스퍼라제	점액지질증 ML III (슈도-혈러 다발이영양증)	79158	Q3T906
N-아세틸글루코사미닐-1-포스포트랜스퍼라제 촉매 서브유닛	점액지질증 ML II (I-세포 질병)	79158	Q3T906
N-아세틸글루코사미닐-1-포스포트랜스퍼라제, 기질-인식 서브유닛	점액지질증 ML III (슈도-혈러 다발이영양증) 유형 III-C	84572	Q9UJJ9
N-아스파르틸글루코사미니다제	아스파르틸글루코사민뇨증	175	P20933
뉴라미니다제 I(시알리다제)	시알산증	4758	Q99519
팔미토일-단백질 티오에스테라제-1	세로이드 리포푸신증 성인형(CLN4, 쿠프스병)	5538	P50897
팔미토일-단백질 티오에스테라제-1	세로이드 리포푸신증 영아형(CLN1, 산타부오리-할티아병)	5538	P50897
페닐알라닌 하이드록실라제	페닐케톤뇨증	5053	P00439

[0613]

포스포만노뉴타제-2	N-글리코실화 CDG Ia의 선천성 장애 (전적으로 신경학적 및 신경학적-다기관 형태)	5373	O15305
포르포빌리노겐 데아미나제	급성 간혈성 포르피린증	3145	P08397
퓨린 뉴클레오시드 포스포릴라제 포스포릴라제	퓨린 뉴클레오시드 포스포릴라제 결핍증	4860	P00491
피리미딘 5' 뉴클레오티다제	용혈성 빈혈 및/또는 피리미딘 5' 뉴클레오티다제 결핍증	51251	Q9H0P0
스핑고미엘리나제	니만-픽병 유형 A	6609	P17405
스핑고미엘리나제	니만-픽병 유형 B	6609	P17405
스테롤 27-하이드록실라제	뇌침줄 황색종증(콜레스타롤 지질증)	1593	Q02318
티미딘 포스포릴라제	미토콘드리아 신경위장 뇌근병증(MNGIE)	1890	P19971
트리헥소실세라미드 $\alpha$ -갈락토시다제	파브리병	2717	P06280
티로시나제, 예를 들어, OCA1	백색증, 예를 들어, 안구 백색증	7299	P14679
UDP-GlcNAc:돌리칠-P NAcGlc 포스포트랜스퍼라제	N-글리코실화 CDG Ij의 선천성 장애	1798	Q9H3H5
UDP-N-아세틸글루코사민-2-에피머라제 /N-아세틸만노사민 키나제, 시알린	시알루리아 프랑스 유형	10020	Q9Y223
우리카제	레슈-니한 증후군, 통풍	391051	단백질 없음

[0614]

우리딘 디포스페이트 글루쿠로닐- 트랜스퍼라제(예를 들어, UGT1A1)	크리글러-나자르 증후군	54658	P22309
$\alpha$ -1,2-만노실트랜스퍼라제	N-글리코실화 CDG II (608776)의 선천성 장애	79796	Q9H6U8
$\alpha$ -1,2-만노실트랜스퍼라제	N-글리코실화의 선천성 장애, 유형 I (사전-골지 글리코실화 결함)	79796	Q9H6U8
$\alpha$ -1,3-만노실트랜스퍼라제	N-글리코실화 CDG II의 선천성 장애	440138	Q2TAA5
$\alpha$ -D-만노시다제	$\alpha$ -만노시드증, 유형 I(중증) 또는 II(경증)	10195	Q92685
$\alpha$ -L-푸코시다제	푸코시드축적증	4123	Q9NTJ4
$\alpha$ -I-이두로니다제	점액다당류증 MPS I H/S (혈러- 사이에 증후군)	2517	P04066
$\alpha$ -I-이두로니다제	점액다당류증 MPS I-H (혈러 증후군)	3425	P35475
$\alpha$ -I-이두로니다제	점액다당류증 MPS I-S (사이에 증후군)	3425	P35475
$\beta$ -1,4- 갈락토실트랜스퍼라제	N-글리코실화 CDG IIId의 선천성 장애	3425	P35475
$\beta$ -1,4-만노실트랜스퍼라제	N-글리코실화 CDG IIk의 선천성 장애	2683	P15291
$\beta$ -D-만노시다제	$\beta$ -만노시드증	56052	Q9BT22
$\beta$ -갈락토시다제	점액다당류증 MPS IV (모르퀴오 증후군) 유형 IV-B	4126	O00462
$\beta$ -글루쿠로니다제	점액다당류증 MPS VII (슬라이 증후군)	2720	P16278
$\beta$ -헥소사미니다제 A	테이-삭스병	2990	P08236
$\beta$ -헥소사미니다제 B	샌드호프병	3073	P06865

[0615]

[0616]

일부 실시형태에서, 본원에 기재된 이펙터는 표 54의 효소 또는 그의 기능적 변이체, 예를 들어, 그의 상동체 (예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 54에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 96%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 기능적 변이체는 예를 들어, 야생형 단백질보다 10%, 20%, 30%, 40% 또는 50% 이상 더 낮은 속도로, 상응하는 야생형 단백질과 동일한 반응을 촉매시킨다. 일부 실시형태에서, 표 54의 효소 또는 그의 기능적 변이체를 인코딩하는 아넬로백터는 표 54의 질병 또는 장애의 치료를 위해 사용된다. 일부 실시형태에서, 아넬로백터는 우리딘 다이포스페이트 글루쿠로닐-트랜스퍼라제 또는 그의 기능적 변이체를 표적 세포, 예를 들어, 간 세포로 운반하기 위해 사용된다. 일부 실시형태에서, 아넬로백터는 OCA1 또는 그의 기능적 변이체를 표적 세포, 예를 들어, 망막 세포로 운반하기 위해 사용된다.

[0617] 표 55. 예시적인 비-효소적 이펙터 및 상응하는 적응증

이펙터	적응증	Entrez 유전자 ID	UniProt ID
생존 운동 뉴런 단백질(SMN)	척수성 근위축증	6606	Q16637
디스트로핀 또는 마이크로-디스트로핀	근이영양증 (예를 들어, 뒤시엔느 근이영양증 또는 베커 근이영양증)	1756	P11532
보체 단백질, 예를 들어, 보체 인자 C1	보체 인자 I 결핍증	3426	P05156
보체 인자 H	비정형 용혈성 요독증	3075	P08603
시스틴증(리소좀 시스틴 수송체)	시스틴증	1497	O60931
부고환 분비 단백질 I(HE1; NPC2 단백질)	니만-픽병 유형 C2	10577	P61916
GDP-푸코스 수송체-1	N-글리코실화 CDG IIc 의 선천성	55343	Q96A29

[0618]

	장애(람밤-하사론 증후군)		
GM2 활성화제 단백질	GM2 활성화제 단백질 결핍증(테이-삭스병 AB 변이형, GM2A)	2760	Q17900
리소좀 막횡단 CLN3 단백질	세로이드 리포푸신증 소아형(CLN3, 바텐병, 보그트-스필마이어병)	1207	Q13286
리소좀 막횡단 CLN5 단백질	세로이드 리포푸신증 변이형 후기 영아형, 핀란드 유형(CLN5)	1203	O75503
Na 인산염 공동수송체, 시알린	영아 시알산 축적 장애	26503	Q9NRA2
Na 인산염 공동수송체, 시알린	시알루리아 핀란드 유형 (살라병)	26503	Q9NRA2
NPC1 단백질	니만-픽병 유형 C1/유형 D	4864	O15118
올리고머 글리 복합체-7	N-글리코실화 CDG IIc 의 선천성 장애	91949	P83436
프로사포신	프로사포신 결핍증	5660	P07602
보호 단백질/카텝신 A (PPCA)	갈락토시알산증(골드버그 증후군, 복합 뉴라미니다제 및 β-갈락토시다제 결핍증)	5476	P10619
만노스-P-돌리콜 이용에 수반되는 단백질	N-글리코실화 CDG If 의 선천성 장애	9526	O75352

[0619]

사포신 B	사포신 B 결핍증(설파티드 활성화제 결핍증)	5660	P07602
사포신 C	사포신 C 결핍증(고세 활성화제 결핍증)	5660	P07602
설파타제-변형 인자-1	뮤코설파티도시스(다발성 설파타제 결핍증)	285362	Q8NBK3
막횡단 CLN6 단백질	세로이드 리포푸신증 변이형 후기 영아형 (CLN6)	54982	Q9NWW5
막횡단 CLN8 단백질	세로이드 리포푸신증 지적 장애를 동반한 진행성 간질	2055	Q9UBY8
vWF	폰 빌레브란트병	7450	P04275
인자 I(피브리노겐)	무피브리노겐혈증	2243, 2244, 2266	P02671, P02675, P02679
에리트로포이에틴 (hEPO)			

[0620]

[0621]

일부 실시형태에서, 본원에 기재된 이펙터는 에리트로포이에틴(EPO), 예를 들어, 인간 에리트로포이에틴(hEPO) 또는 그의 기능적 변이체를 포함한다. 일부 실시형태에서, 에리트로포이에틴 또는 그의 기능적 변이체를 인코딩하는 아넬로백터는 적혈구형성을 자극하기 위해 사용된다. 일부 실시형태에서, 에리트로포이에틴 또는 그의 기능적 변이체를 인코딩하는 아넬로백터는 질병 또는 장애, 예를 들어, 빈혈의 치료를 위해 사용된다. 일부 실시형태에서, 아넬로백터는 EPO 또는 그의 기능적 변이체를 표적 세포, 예를 들어, 적혈구로 운반하기 위해 사용된다.

[0622]

일부 실시형태에서, 본원에 기재된 이펙터는 표 55의 폴리펩티드 또는 그의 기능적 변이체, 예를 들어, 그의 상동체(예를 들어, 오르토로그 또는 파라로그) 또는 단편을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 그의 UniProt ID를 참조하여 표 55에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 실시형태에서, 표 55의 폴리펩티드 또는 그의 기능적 변이체를 인코딩하는 아넬로백터는 표 55의 질병 또는 장애의 치료에 사용된다. 일부 실시형태에서, 아넬로백터는 SMN 또는 그의 기능적 변이체를 표적 세포, 예를 들어, 척수 및/또는 운동 뉴런의 세포로 운반하기 위해 사용된다. 일부 실시형태에서, 아넬로백터는 마이크로-디스트로핀을 표적 세포, 예를 들어, 근세포로 운반하기 위해 사용된다.

[0623]

예시적인 마이크로-디스트로핀은 문헌[Duan, "Systemic AAV Micro-dystrophin Gene Therapy for Duchenne Muscular Dystrophy." Mol Ther. 2018 Oct 3;26(10):2337-2356. doi: 10.1016/j.ymthe.2018.07.011. Epub 2018 Jul 17]에 기재되어 있다.

[0624]

일부 실시형태에서, 본원에 기재된 이펙터는 응고 인자, 예를 들어, 본원에서 표 54 또는 표 55에 열거된 응고 인자를 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 돌연변이되는 경우 리소좀 축적 장애를 야기하는 단백질, 예를 들어, 본원에서 표 54 또는 표 55에 열거된 단백질을 포함한다. 일부 실시형태에서, 본원에 기재된 이펙터는 수송체 단백질, 예를 들어, 본원에서 표 55에 열거된 수송체 단백질을 포함한다.

[0625]

일부 실시형태에서, 야생형 단백질의 기능적 변이체는 야생형 단백질의 하나 이상의 활성을 갖는 단백질을 포함하며, 예를 들어, 기능적 변이체는 예를 들어, 야생형 단백질보다 10%, 20%, 30%, 40% 또는 50% 이상 더 낮은 속도로, 상응하는 야생형 단백질과 동일한 반응을 촉매시킨다. 일부 실시형태에서, 기능적 변이체는 야생형 단백질에 의해 결합되는 동일한 결합 파트너에, 예를 들어, 동일한 조건 하의 동일한 결합 파트너에 대해 상응하는 야생형 단백질의 Kd보다 10%, 20%, 30%, 40% 또는 50% 이하로 더 큰 Kd로 결합한다. 일부 실시형태에서, 기능적 변이체는 야생형 폴리펩티드의 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일한 폴리펩티드 서열을 갖는다. 일부 실시형태에서, 기능적 변이체는 상응하는 야생형 단백질의 상동체(예를 들어, 오르토로그 또는 파라로그)를 포함한다. 일부 실시형태에서, 기능적 변이체는 융합 단백질이다. 일부 실시형태에서, 융합물은 상응하는 야생형 단백질에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일성을 갖는 제1 영역 및 제2 이중 영역을 포함한다. 일부 실시형태에서, 기능적 변이체는 상응하는 야생형 단백질의 단편을 포함하거나, 이로 이루어진다.

[0626] **재생, 복구 및 섬유증 인자**

[0627] 본원에 기재된 치료적 폴리펩티드는 예를 들어, 표 56에 개시된 바와 같은 성장 인자 또는 그의 기능적 변이체, 예를 들어, 그의 UniProt ID를 참조하여 표 56에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질도 포함한다. 이러한 성장 인자에 대한 항체 또는 그의 단편, 또는 재생 및 복구를 촉진하는 miRNA도 포함된다.

[0628] **표 56. 예시적인 재생, 복구 및 섬유증 인자**

표적	유전자 수탁 번호	단백질 수탁 번호
VEGF-A	NG_008732	NP_001165094
NRG-1	NG_012005	NP_001153471
FGF2	NG_029067	NP_001348594
FGF1	유전자 ID: 2246	NP_001341882
miR-199-3p	MIMAT0000232	
miR-590-3p	MIMAT0004801	
mi-17-92	MI0000071	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/</a>
miR-222	MI0000299	
miR-302-367	MIR302A 및 MIR367	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/</a>

[0629]

[0630] **형질전환 인자**

[0631] 본원에 기재된 치료적 폴리펩티드는 형질전환 인자, 예를 들어, 섬유아세포를 분화 세포로 형질전환시키는 단백질 인자, 예를 들어, 표 57에 개시된 인자 또는 그의 기능적 변이체, 예를 들어, 그의 UniProt ID를 참조하여 표 57에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질도 포함한다.

[0632] 표 57. 예시적인 형질전환 인자

표적	적응증	유전자 수탁 번호	단백질 수탁 번호
MESP1	섬유아세포 형질전환에 의한 기관 복구	유전자 ID: 55897	EAX02066
ETS2	섬유아세포 형질전환에 의한 기관 복구	유전자ID: 2114	NP_005230
HAND2	섬유아세포 형질전환에 의한 기관 복구	유전자ID: 9464	NP_068808
MYOCARDIN	섬유아세포 형질전환에 의한 기관 복구	유전자ID: 93649	NP_001139784
ESRRA	섬유아세포 형질전환에 의한 기관 복구	유전자 ID: 2101	AAH92470
miR-1	섬유아세포 형질전환에 의한 기관 복구	MI0000651	n/a
miR-133	섬유아세포 형질전환에 의한 기관 복구	MI0000450	n/a
TGFb	섬유아세포 형질전환에 의한 기관 복구	유전자ID: 7040	NP_000651.3
WNT	섬유아세포 형질전환에 의한 기관 복구	유전자 ID: 7471	NP_005421
JAK	섬유아세포 형질전환에 의한 기관 복구	유전자 ID: 3716	NP_001308784
NOTCH	섬유아세포 형질전환에 의한 기관 복구	유전자ID: 4851	XP_011517019

[0633]

[0634] 세포 재생을 촉진하는 단백질

[0635]

본원에 기재된 치료적 폴리펩티드는 세포 재생을 자극하는 단백질, 예를 들어, 표 58에 개시된 단백질 또는 그의 기능적 변이체, 예를 들어, 그의 UniProt ID를 참조하여 표 58에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질도 포함한다.

[0636] 표 58. 세포 재생을 자극하는 예시적인 단백질

표적	유전자 수탁 번호	단백질 수탁 번호
MST1	NG_016454	NP_066278
STK30	유전자 ID: 26448	NP_036103
MST2	유전자 ID: 6788	NP_006272
SAV1	유전자 ID: 60485	NP_068590
LATS1	유전자 ID: 9113	NP_004681
LATS2	유전자 ID: 26524	NP_055387
YAP1	NG_029530	NP_001123617
CDKN2b	NG_023297	NP_004927
CDKN2a	NG_007485	NP_478102

[0637]

[0638]

**STING 조절자 이펙터**

[0639]

일부 실시형태에서, 본원에 기재된 분비 이펙터는 STING/cGAS 신호전달을 조절한다. 일부 실시형태에서, STING 조절자는 폴리펩티드, 예를 들어, 바이러스 폴리펩티드 또는 그의 기능적 변이체이다. 예를 들어, 이펙터는 그의 전문이 본원에 참조로 포함되는 문헌[Maringer et al. "Message in a bottle: lessons learned from antagonism of STING signalling during RNA virus infection" Cytokine & Growth Factor Reviews Volume 25, Issue 6, December 2014, Pages 669-679]에 기재된 STING 조절자(예를 들어, 억제제)를 포함할 수 있다. 추가의 STING 조절자(예를 들어, 활성화제)는 예를 들어, 이들 각각의 전문이 본원에 참조로 포함되는 문헌[Wang et al. "STING activator c-di-GMP enhances the anti-tumor effects of peptide vaccines in melanoma-bearing mice." Cancer Immunol Immunother. 2015 Aug;64(8):1057-66. doi: 10.1007/s00262-015-1713-5. Epub 2015 May 19; Bose "cGAS/STING Pathway in Cancer: Jekyll and Hyde Story of Cancer Immune Response" Int J Mol Sci. 2017 Nov; 18(11): 2456; 및 Fu et al. "STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade" Sci Transl Med. 2015 Apr 15; 7(283): 283ra52]에 기재되어 있다.

[0640]

펩티드의 일부 예는 형광 태그 또는 마커, 항원, 펩티드 치료제, 천연적인 생물활성 펩티드로부터의 합성 또는 유사체 펩티드, 효능제 또는 길항제 펩티드, 항-미생물 펩티드, 표적화 또는 세포독성 펩티드, 분해 또는 자가-파괴 펩티드, 및 분해 또는 자가-파괴 펩티드들을 포함하지만, 이들에 제한되지 않는다. 본원에 기재된 본 발명에 유용한 펩티드는 또한, 항원-결합 펩티드, 예를 들어, 항원 결합 항체 또는 항체-유사 단편, 예컨대 단일쇄 항체, 나노바디(예를 들어, 문헌[Steeland et al. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. Drug Discov Today: 21(7):1076-113] 참조)를 포함한다. 이러한 항원 결합 펩티드는 시토졸 항원, 핵 항원 또는 소기관-내 항원에 결합할 수 있다.

[0641]

일부 실시형태에서, 유전 요소는 작은 펩티드, 펩티도미메틱(예를 들어, 펩토이드), 아미노산 및 아미노산 유사체를 인코딩하는 서열을 포함한다. 이러한 치료제는 일반적으로 몰당 약 5,000그램 미만의 분자량, 몰당 약 2,000그램 미만의 분자량, 몰당 약 1,000그램 미만의 분자량, 몰당 약 500그램 미만의 분자량, 및 이러한 화합물의 염, 에스테르 및 기타 약제학적으로 허용 가능한 형태를 갖는다. 이러한 치료제는 신경전달물질, 호르몬, 약물, 독소, 바이러스 또는 미생물 입자, 합성 분자, 및 그의 효능제 또는 길항제를 포함할 수 있으나, 이에 제한되지 않는다.

[0642]

일부 실시형태에서, 본원에 기재된 조성물 또는 아넬로백터는 특이적 위치, 조직 또는 세포를 표적화할 수 있는

리간드에 연결된 폴리펩티드를 포함한다.

[0643]

유전자 편집 구성성분

[0644]

아넬로박터의 유전 요소는 유전자 편집 시스템의 구성성분을 인코딩하는 하나 이상의 유전자를 포함할 수 있다. 예시적인 유전자 편집 시스템은 클러스터링된 조절적으로 산재된 짧은 회문 반복부(CRISPR) 시스템, 아연 핑거 뉴클레아제(ZFN) 및 전사 활성화제-유사 이펙터-기반의 뉴클레아제(TALEN)를 포함한다. ZFN, TALEN 및 CRISPR-기반의 방법은 예를 들어, 문헌[Gaj et al. Trends Biotechnol. 31.7(2013):397-405]에 기재되어 있으며; 유전자 편집의 CRISPR 방법은 예를 들어, 문헌[Guan et al., Application of CRISPR-Cas system in gene therapy: Pre-clinical progress in animal model. DNA Repair 2016 Oct;46:1-8. doi: 10.1016/j.dnarep.2016.07.004]; 문헌[Zheng et al., Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. BioTechniques, Vol. 57, No. 3, September 2014, pp. 115-124]에 기재되어 있다.

[0645]

CRISPR 시스템은 원래 박테리아 및 조류에서 관찰되는 적응 방어 시스템이다. CRISPR 시스템은 CRISPR-회합 또는 "Cas" 엔도뉴클레아제(예를 들어, Cas9 또는 Cpf1)로 명명되는 RNA-가이드된 뉴클레아제를 사용하여, 외래 DNA를 절단한다. 전형적인 CRISPR/Cas 시스템에서, 엔도뉴클레아제는 단일- 또는 이중-가닥 DNA 서열을 표적화하는 서열-특이적 비-코딩 "가이드 RNA"에 의해 표적 뉴클레오티드 서열(예를 들어, 서열-편집될 게놈 내의 부위)에 지향된다. CRISPR 시스템의 3가지 부류(I 내지 III)가 확인된 바 있다. 부류 II CRISPR 시스템은 (다수의 Cas 단백질보다는) 단일의 Cas 엔도뉴클레아제를 사용한다. 하나의 부류 II CRISPR 시스템은 II형 Cas 엔도뉴클레아제, 예컨대 Cas9, CRISPR RNA("crRNA") 및 트랜스-활성화 crRNA("tracrRNA")를 포함한다. crRNA는 "가이드 RNA", 전형적으로 표적 DNA 서열에 상응하는 약 20-뉴클레오티드 RNA 서열을 함유한다. crRNA는 또한, tracrRNA에 결합하여, RNase III에 의해 절단되는 부분 이중-가닥 구조를 형성하여, crRNA/tracrRNA 혼성물을 초래하는 영역을 함유한다. 이어서, crRNA/tracrRNA 혼성물은 Cas9 엔도뉴클레아제가 표적 DNA 서열을 인식하고 절단하게 한다. 표적 DNA 서열은 일반적으로 주어진 Cas 엔도뉴클레아제에 특이적인 "프로토스페이서 인접 모티프("PAM")에 인접해야 하지만; PAM 서열은 주어진 게놈의 도처에서 나타난다.

[0646]

일부 실시형태에서, 아넬로박터는 CRISPR 엔도뉴클레아제에 대한 유전자를 포함한다. 예를 들어, 다양한 원핵생물 종으로부터 확인되는 일부 CRISPR 엔도뉴클레아제는 독특한 PAM 서열 요건을 가지며; PAM 서열의 예는 5'-NGG(스트렙토코커스 피오게네스(*Streptococcus pyogenes*)), 5'-NNAGAA(스트렙토코커스 썬모필러스(*Streptococcus thermophilus*) CRISPR1), 5'-NGGNG(스트렙토코커스 썬모필러스 CRISPR3) 및 5'-NNNGATT(네이세리아 메닌지디티스(*Neisseria meningitidis*))를 포함한다. 일부 엔도뉴클레아제, 예를 들어, Cas9 엔도뉴클레아제는 G-풍부 PAM 부위, 예를 들어, 5'-NGG와 회합되며, PAM 부위(이로부터 5')로부터 3개 뉴클레오티드 상류의 위치에서 표적 DNA의 블린트-말단 절단을 수행한다. 또 다른 부류 II CRISPR 시스템은 Cas9보다 더 작은 V형 엔도뉴클레아제 Cpf1을 포함하며; 예에는 AsCpf1(아시드아미노코커스(*Acidaminococcus*) 종 유래) 및 LbCpf1(라크노스피라세아에(*Lachnospiraceae*) 종 유래)이 포함된다. Cpf1 엔도뉴클레아제는 T-풍부 PAM 부위, 예를 들어, 5'-TTN과 회합된다. 또한, Cpf1은 5'-CTA PAM 모티프를 인식할 수 있다. Cpf1은 4- 또는 5-뉴클레오티드 5' 오버행 갖는 오프셋 또는 스테거형 이중-가닥 파단부를 도입함으로써, 예를 들어, 코딩 가닥 상의 PAM 부위(그로부터 3')로부터 18개 뉴클레오티드 하류, 및 상보성 가닥 상의 PAM 부위로부터 23개 뉴클레오티드 하류에 위치한 5-뉴클레오티드 오프셋 또는 스테거형 절단으로 표적 DNA를 절단함으로써 표적 DNA를 절단하며; 이러한 오프셋 절단으로부터 초래되는 5-뉴클레오티드 오버행은 상동성 재조합에 의한 DNA 삽입에 의해 블린트-말단 절단된 DNA에서의 삽입에 의한 것보다 더욱 정밀한 게놈 편집을 가능하게 한다. 예를 들어, 문헌[Zetsche et al. (2015) Cell, 163:759 - 771]을 참조한다.

[0647]

다양한 CRISPR 회합(Cas) 유전자가 아넬로박터 내에 포함될 수 있다. 유전자의 구체적인 예는 Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1 또는 C2C3을 포함하는 부류 II 시스템으로부터의 Cas 단백질을 인코딩하는 것들이다. 일부 실시형태에서, 아넬로박터는 Cas 단백질, 예를 들어, 다양한 원핵생물 종 중 임의의 것으로부터일 수 있는 Cas9 단백질을 인코딩하는 유전자를 포함한다. 일부 실시형태에서, 아넬로박터는 특정 Cas 단백질, 예를 들어, 특정 프로토스페이서-인접 모티프(PAM) 서열을 인식하도록 선택되는 특정 Cas9 단백질을 인코딩하는 유전자를 포함한다. 일부 실시형태에서, 아넬로박터는 세포, 접합체, 배아 또는 동물 내로 도입되어, 예를 들어, 동일한, 유사한 또는 상이한 PAM 모티프를 포함하는 부위의 인식 및 변형을 가능하게 할 수 있는 2가지 이상의 상이한 Cas 단백질 또는 2가지 이상의 Cas 단백질을 인코딩하는 핵산을 포함한다. 일부 실시형태에서, 아넬로박터는 불활성화된 뉴클레아제를 갖는 변형된 Cas 단백질, 예를 들어, 뉴클레아제-결합 Cas9를 인코딩하는 유전자를 포함한다.

- [0648] 야생형 Cas9 단백질이 gRNA에 의해 표적화된 특정 DNA 서열에서 이중-가닥 파단부(DSB)를 생성하는 한편, 변형된 기능을 갖는 수많은 CRISPR 엔도뉴클레아제가 알려져 있으며, 예를 들어, "닉카제" 버전의 Cas 엔도뉴클레아제(예를 들어, Cas9)는 오직 단일-가닥 파단만을 생성하며; 촉매적으로 비활성인 Cas 엔도뉴클레아제, 예를 들어, Cas9("dCas9")는 표적 DNA를 절단하지 않는다. dCas9를 인코딩하는 유전자를 이펙터 도메인을 인코딩하는 유전자와 융합시켜, 표적 유전자의 발현을 억제(CRISPRi)하거나 활성화(CRISPRa)시킬 수 있다. 예를 들어, 유전자는 전사 침묵화제(예를 들어, KRAB 도메인) 또는 전사 활성화제(예를 들어, dCas9-VP64 융합물)를 갖는 Cas9 융합물을 인코딩할 수 있다. 2개의 gRNA에 상동성인 표적 서열에 DSB를 생성하기 위하여, FokI 뉴클레아제에 융합된 촉매적 비활성 Cas9(dCas9)("dCas9-FokI")를 인코딩하는 유전자가 포함될 수 있다. 예를 들어, 애드진 레포지터리(Addgene repository)(Addgene, 75 Sidney St., Suite 550A, Cambridge, MA 02139; addgene.org/crispr/)에 게시되고, 그로부터 공개적으로 이용 가능한 수많은 CRISPR/Cas9 플라스미드를 참조한다. 각각 개별 가이드 RNA에 의해 유도되는, 2개의 개별 이중-가닥 파단을 도입하는 "이중 닉카제" Cas9는 더욱 정확한 게놈 편집을 달성하는 것으로 문헌[Ran et al. (2013) Cell, 154:1380 - 1389]에 기재되어 있다.
- [0649] 진핵생물의 유전자를 편집하기 위한 CRISPR 기술은 미국 특허 출원 공개 제2016/0138008A1호 및 제 US2015/0344912A1호에, 그리고 미국 특허 제8,697,359호, 제8,771,945호, 제8,945,839호, 제8,999,641호, 제8,993,233호, 제8,895,308호, 제8,865,406호, 제8,889,418호, 제8,871,445호, 제8,889,356호, 제8,932,814호, 제8,795,965호 및 제8,906,616호에 게시되어 있다. Cpf1 엔도뉴클레아제 및 상응하는 가이드 RNA 및 PAM 부위는 미국 특허 출원 공개 제2016/0208243 A1호에 게시되어 있다.
- [0650] 일부 실시형태에서, 아넬로박터는 본원에 기재된 폴리펩티드, 예를 들어, 표적화된 뉴클레아제, 예를 들어, Cas9, 예를 들어, 야생형 Cas9, 닉카제 Cas9(예를 들어, Cas9 D10A), 데드(dead) Cas9(dCas9), eSpCas9, Cpf1, C2C1 또는 C2C3, 및 gRNA를 인코딩하는 유전자를 포함한다. 뉴클레아제 및 gRNA(들)를 인코딩하는 유전자의 선택은 표적화된 돌연변이가 뉴클레오티드의 결실인지, 치환인지 또는 부가인지, 예를 들어, 표적화된 서열에 대한 뉴클레오티드의 결실인지, 치환인지 또는 부가인지에 의해 결정된다. (하나 이상의) 이펙터 도메인(예를 들어, VP64)의 전부 또는 그의 일부(예를 들어, 그의 생물학적 활성 부분)와 테더링된 촉매적 비활성 엔도뉴클레아제, 예를 들어, 데드 Cas9(dCas9, 예를 들어, D10A; H840A)를 인코딩하는 유전자는 하나 이상의 표적 핵산 서열의 활성 및/또는 발현을 조절할 수 있는 키메라 단백질을 생성한다.
- [0651] 일부 실시형태에서, 아넬로박터는 본원에 기재된 방법에서 유용한 키메라 단백질을 생성하기 위하여 하나 이상의 이펙터 도메인의 전부 또는 부분(예를 들어, 전장 야생형 이펙터 도메인, 또는 그의 단편 또는 변이체, 예를 들어, 그의 생물학적 활성 부분)와 dCas9의 융합물을 인코딩하는 유전자를 포함한다. 따라서, 일부 실시형태에서, 아넬로박터는 dCas9-메틸라제 융합물을 인코딩하는 유전자를 포함한다. 다른 일부 실시형태에서, 아넬로박터는 내인성 유전자를 표적화하기 위하여 부위-특이적 gRNA와 함께 dCas9-효소 융합물을 인코딩하는 유전자를 포함한다.
- [0652] 다른 양태에서, 아넬로박터는 dCas9와 융합된 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 이펙터 도메인(모든 또는 생물학적 활성 부분)을 인코딩하는 유전자를 포함한다.
- [0653] **조절 서열**
- [0654] 일부 실시형태에서, 유전 요소는 이펙터를 인코딩하는 서열에 작동 가능하게 연결된 조절 서열, 예를 들어, 프로모터 또는 인핸서를 포함한다. 일부 실시형태에서, 예를 들어, 유전 요소가 mRNA인 경우, 프로모터는 유전 요소에 부재할 수 있다. 일부 실시형태에서, 유전 요소 자체물은 RNA 유전 요소의 생성을 구동하기 위해 사용되는 프로모터를 포함한다.
- [0655] 일부 실시형태에서, 프로모터는 발현 생성물을 인코딩하는 DNA 서열에 인접하여 위치한 DNA 서열을 포함한다. 프로모터는 인접 DNA 서열에 작동 가능하게 연결될 수 있다. 프로모터는 전형적으로 프로모터가 존재하지 않는 경우 발현되는 생성물의 양에 비하여 DNA 서열로부터 발현되는 생성물의 양을 증가시킨다. 하나의 유기체로부터의 프로모터를 사용하여 또 다른 유기체로부터 기원한 DNA 서열로부터의 생성물 발현을 향상시킬 수 있다. 예를 들어, 척추동물 프로모터는 척추동물에서 해파리 GFP의 발현을 위해 사용될 수 있다. 그러므로, 하나의 프로모터 요소는 하나 이상의 생성물의 발현을 향상시킬 수 있다. 다수의 프로모터 요소는 당업자에게 널리 알려져 있다.
- [0656] 일 실시형태에서, 높은 수준의 구성적 발현이 요망된다. 이러한 프로모터의 예는 제한 없이, 레트로바이러스 라우스 육종 바이러스(RSV) 긴 말단 반복부(LTR) 프로모터/인핸서, 사이토메갈로바이러스(CMV) 즉시 초기 프로모

터/인헨서(예를 들어, 문헌[Boshart et al, Cell, 41:521-530 (1985)] 참고), SV40 프로모터, 다이하이드로폴레이트 환원효소 프로모터, 세포질.베타.-액틴 프로모터 및 포스포글리세롤 키나제(PGK) 프로모터를 포함한다.

[0657] 또 다른 실시형태에서, 유도성 프로모터가 요망될 수 있다. 유도성 프로모터는 제한 없이, 아연-유도성 양 메탈로티오닌(MT) 프로모터; 텍사메타손(Dex)-유도성 마우스 유방 종양 바이러스(MMTV) 프로모터; T7 중합효소 프로모터 시스템(WO 98/10088호); 테트라사이클린-억제성 시스템(문헌[Gossen et al, Proc. Natl. Acad. Sci. USA, 89:5547-5551 (1992)]); 테트라사이클린-유도성 시스템(문헌[Gossen et al., Science, 268:1766-1769 (1995)]); 또한, 문헌[Harvey et al., Curr. Opin. Chem. Biol., 2:512-518 (1998)] 참조); RU486-유도성 시스템(문헌[Wang et al., Nat. Biotech., 15:239-243 (1997)] 및 문헌[Wang et al., Gene Ther., 4:432-441 (1997)]); 및 라파마이신-유도성 시스템(문헌[Magari et al., J. Clin. Invest., 100:2865-2872 (1997)]); 문헌[Rivera et al., Nat. Medicine. 2:1028-1032 (1996)])을 포함하여, 예를 들어, 시스로 또는 트랜스로 제공되는, 외인적으로 공급되는 화합물에 의해 조절되는 것들이다. 이러한 맥락에서 유용할 수 있는 다른 유형의 유도성 프로모터는 특정 생리학적 상태, 예를 들어, 온도, 급성 단계에 의해 또는 오직 복제하는 세포에서만 조절되는 것들이다.

[0658] 일부 실시형태에서, 관심 유전자 또는 핵산 서열에 대한 고유 프로모터가 사용된다. 유전자 또는 핵산 서열의 발현이 고유 발현을 모방하는 것이 요망되는 경우에 고유 프로모터가 사용될 수 있다. 유전자 또는 다른 핵산 서열의 발현이 일시적으로 또는 발달적으로, 또는 조직-특이적 방식으로 또는 특정 전사 자극에 반응하여 조절되어야 하는 경우에, 고유 프로모터가 사용될 수 있다. 추가의 실시형태에서, 다른 고유 발현 제어 요소, 예컨대, 인헨서 요소, 폴리아데닐화 부위 또는 코작(Kozak) 공통 서열도 또한 고유 발현을 모방하기 위해 사용될 수 있다.

[0659] 일부 실시형태에서, 유전 요소는 조직-특이적 프로모터에 작동 가능하게 연결된 유전자를 포함한다. 예를 들어, 골격근에서의 발현이 요망되는 경우, 근육에서 활성인 프로모터가 사용될 수 있다. 이들은 골격 α-액틴, 미오신 경쇄 2A, 디스트로빈, 근육 크레아틴 키나제를 인코딩하는 유전자로부터의 프로모터 및 자연-발생 프로모터보다 더 높은 활성을 갖는 합성 근육 프로모터를 포함한다. 문헌[Li et al., Nat. Biotech., 17:241-245 (1999)]을 참조한다. 특히 간 알부민, 문헌[Miyatake et al. J. Virol., 71:5124-32 (1997)]; B형 간염 바이러스 코어 프로모터, 문헌[Sandig et al., Gene Ther. 3:1002-9 (1996)]; 알파-태아단백질(AFP), 문헌[Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)], 골(오스테오칼신(osteocalcin), 문헌[Stein et al., Mol. Biol. Rep., 24:185-96 (1997)]; 골 시알로단백질(sialoprotein), 문헌[Chen et al., J. Bone Miner. Res. 11:654-64 (1996)], 림프구(CD2, 문헌[Hansal et al., J. Immunol., 161:1063-8 (1998)]; 면역글로불린 중쇄; T 세포 수용체 α 체), 뉴런(뉴런-특이적 에놀라제(NSE) 프로모터, 문헌[Andersen et al. Cell. Mol. Neurobiol., 13:503-15 (1993)]; 신경필라멘트 경쇄 유전자, 문헌[Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)]; 뉴런-특이적 vgf 유전자, 문헌[Piccioli et al., Neuron, 15:373-84 (1995)]을 위한 조직-특이적인 프로모터의 예가 알려져 있다.

[0660] 유전 요소 작제물은 인헨서, 예를 들어, 유전자를 인코딩하는 DNA 서열에 인접하여 위치한 DNA 서열을 포함할 수 있다. 인헨서 요소는 전형적으로 프로모터 요소의 상류에 위치하거나, 코딩 DNA 서열(예를 들어, 생성물 또는 생성물들로 전사 또는 번역되는 DNA 서열)의 하류에 또는 그 내에 위치할 수 있다. 그러므로, 인헨서 요소는 생성물을 인코딩하는 DNA 서열의 100개 염기쌍, 200개 염기쌍 또는 300개 이상의 염기쌍 상류 또는 하류에 위치할 수 있다. 인헨서 요소는 DNA 서열로부터 발현되는 재조합 생성물의 양을 프로모터 요소에 의해 제공되는 증가된 발현을 초과하여 증가시킬 수 있다. 다중의 인헨서 요소는 당업자에게 용이하게 이용 가능하다.

[0661] 일부 실시형태에서, 유전 요소는 본원에 기재된 발현 생성물을 인코딩하는 서열을 측정하는 하나 이상의 역위 말단 반복부(ITR)를 포함한다. 일부 실시형태에서, 유전 요소는 본원에 기재된 발현 생성물을 인코딩하는 서열을 측정하는 하나 이상의 긴 말단 반복부(LTR)를 포함한다. 사용될 수 있는 프로모터 서열의 예는 시미안 바이러스 40(SV40) 초기 프로모터, 마우스 유방 종양 바이러스(MMTV), 인간 면역결핍 바이러스(HIV) 긴 말단 반복부(LTR) 프로모터, MoMuLV 프로모터, 조류 백혈병 바이러스 프로모터, 엡스테인-바 바이러스 즉시 초기 프로모터 및 라우스 육종 바이러스 프로모터를 포함하나 이들에 제한되지 않는다.

[0662] **기타 서열**

[0663] 일부 실시형태에서, 유전 요소는 생성물(예를 들어, 리보자임, 단백질을 인코딩하는 치료적 mRNA, 외인성 유전자)을 인코딩하는 핵산을 추가로 포함한다.

- [0664] 일부 실시형태에서, 유전 요소는 숙주 또는 숙주 세포에서의 아넬로박테리의 종 및/또는 조직 및/또는 세포 형성 (예를 들어, 캡시드 단백질 서열), 감염성(예를 들어, 캡시드 단백질 서열), 면역억제/활성화(예를 들어, 조절 핵산), 바이러스 게놈 결합 및/또는 패키징, 면역 회피(비-면역원성 및/또는 관용), 약동학, 엔도시토시스 및/또는 세포 부착, 핵 유입, 세포내 조절 및 국소화, 엑소시토시스 조절, 증식 및 핵산 보호에 영향을 미치는 하나 이상의 서열을 포함한다.
- [0665] 일부 실시형태에서, 유전 요소는 DNA, RNA 또는 인공 핵산을 포함하는 다른 서열을 포함할 수 있다. 다른 서열은 게놈 DNA, cDNA, 또는 tRNA, mRNA, rRNA, miRNA, gRNA, siRNA 또는 다른 RNAi 분자를 인코딩하는 서열을 포함할 수 있지만, 이들에 제한되지 않는다. 일 실시형태에서, 유전 요소는 조절 핵산과 동일한 유전자 발현 생성물의 상이한 유전자좌를 표적화하기 위하여 siRNA를 인코딩하는 서열을 포함한다. 일 실시형태에서, 유전 요소는 조절 핵산과 상이한 유전자 발현 생성물을 표적화하기 위하여 siRNA를 인코딩하는 서열을 포함한다.
- [0666] 일부 실시형태에서, 유전 요소는 다음의 서열 중 하나 이상을 추가로 포함한다: 하나 이상의 miRNA를 인코딩하는 서열, 하나 이상의 복제 단백질을 인코딩하는 서열, 외인성 유전자를 인코딩하는 서열, 치료제를 인코딩하는 서열, 조절 서열(예를 들어, 프로모터, 인핸서), 내인성 유전자를 표적화하는 하나 이상의 조절 서열(siRNA, lncRNA, shRNA)을 인코딩하는 서열, 및 치료적 mRNA 또는 단백질을 인코딩하는 서열.
- [0667] 다른 서열은 약 2 내지 약 5000개의 뉴클레오티드, 약 10 내지 약 100개의 뉴클레오티드, 약 50 내지 약 150개의 뉴클레오티드, 약 100 내지 약 200개의 뉴클레오티드, 약 150 내지 약 250개의 뉴클레오티드, 약 200 내지 약 300개의 뉴클레오티드, 약 250 내지 약 350개의 뉴클레오티드, 약 300 내지 약 500개의 뉴클레오티드, 약 10 내지 약 1000개의 뉴클레오티드, 약 50 내지 약 1000개의 뉴클레오티드, 약 100 내지 약 1000개의 뉴클레오티드, 약 1000 내지 약 2000개의 뉴클레오티드, 약 2000 내지 약 3000개의 뉴클레오티드, 약 3000 내지 약 4000개의 뉴클레오티드, 약 4000 내지 약 5000개의 뉴클레오티드 또는 그 사이의 임의의 범위의 길이를 가질 수 있다.
- [0668] *인코딩된 유전자*
- [0669] 예를 들어, 유전 요소는 신호전달 생화학 경로와 연관된 유전자, 예를 들어, 신호전달 생화학 경로-연관 유전자 또는 폴리뉴클레오티드를 포함할 수 있다. 예에는 질병 연관 유전자 또는 폴리뉴클레오티드가 포함된다. "질병-연관" 유전자 또는 폴리뉴클레오티드는 비 질병 대조군의 조직 또는 세포와 비교하여, 질병-이환된 조직으로부터 유래된 세포에서 전사 또는 번역 생성물을 비정상적인 수준 또는 비정상적인 형태로 제공하는 임의의 유전자 또는 폴리뉴클레오티드를 지칭한다. 그것은 비정상적으로 높은 수준으로 발현되는 유전자일 수 있으며; 비정상적으로 낮은 수준으로 발현되는 유전자일 수 있으며, 여기서, 변경된 발현은 질병의 발생 및/또는 진행과 상호 연관된다. 질병-연관 유전자는 또한, 질병의 병인의 직접적인 원인이 되거나, 질병의 병인을 담당하는 유전자(들)와 연관 불균형인 돌연변이(들) 또는 유전적 변이를 지니는 유전자를 지칭한다.
- [0670] 질병-연관 유전자 및 폴리뉴클레오티드의 예는 맥쿠식-나탄스(McKusick-Nathans) 유전 의학 기관, 존스 홉킨스 대학(미국 메릴랜드주 볼티모어) 및 국립 생명과학 정보 센터, 국립 의학 도서관(미국 메릴랜드주 베데스다)로부터 이용 가능하다. 질병-연관 유전자 및 폴리뉴클레오티드의 예는 본원에 그들 전문이 참조로 포함되는 미국 특허 제8,697,359호의 표 A 및 B에 열거되어 있다. 질병 특이적 정보는 맥쿠식-나탄스 유전 의학 기관, 존스 홉킨스 대학(미국 메릴랜드주 볼티모어) 및 국립 생명과학 정보 센터, 국립 의학 도서관(미국 메릴랜드주 베데스다)로부터 이용 가능하다. 신호전달 생화학 경로-연관 유전자 및 폴리뉴클레오티드의 예는 본원에 그들 전문이 참조로 포함되는 미국 특허 제8,697,359호의 표 A 내지 C에 열거되어 있다.
- [0671] 또한, 유전 요소는 본원의 다른 곳에 기재된 바와 같이 표적화 모이어티를 인코딩할 수 있다. 이는 예를 들어, 당, 당지질 또는 단백질, 예컨대 항체를 인코딩하는 폴리뉴클레오티드를 삽입함으로써 달성될 수 있다. 당업자는 표적화 모이어티의 추가의 생성 방법을 안다.
- [0672] *바이러스 서열*
- [0673] 일부 실시형태에서, 유전 요소는 적어도 하나의 바이러스 서열을 포함한다. 일부 실시형태에서, 서열은 단일 가닥 DNA 바이러스, 예를 들어, 아넬로바이러스, 비드나바이러스, 썬코바이러스, 제미니바이러스, 게노모바이러스, 이노바이러스, 마이크로바이러스, 나노바이러스, 파보바이러스 및 스피라바이러스로부터의 하나 이상의 서열에 대하여 상동성 또는 동일성을 갖는다. 일부 실시형태에서, 서열은 이중 가닥 DNA 바이러스, 예를 들어, 아데노바이러스, 암폴라바이러스, 아스코바이러스, 아스파바이러스, 배칼로바이러스, 푸셀로바이러스, 글로볼로바이러스, 구타바이러스, 하이트로사바이러스, 헤르페스바이러스, 이리도바이러스, 리포트릭스바이

러스, 니마바이러스 및 폭스바이러스로부터의 하나 이상의 서열에 대하여 상동성 또는 동일성을 갖는다. 일부 실시형태에서, 서열은 RNA 바이러스, 예를 들어, 알파바이러스, 푸로바이러스, 간염 바이러스, 호르데이바이러스, 토마모바이러스, 토브라바이러스, 트리코르나바이러스, 루비바이러스, 비르나바이러스, 시스토바이러스, 파르티티바이러스 및 레오바이러스로부터의 하나 이상의 서열에 대하여 상동성 또는 동일성을 갖는다.

[0674] 일부 실시형태에서, 유전 요소는 비-병원성 바이러스, 예를 들어, 상리공생 바이러스, 예를 들어, 편리공생 바이러스, 예를 들어, 고유 바이러스, 예를 들어, 아넬로바이러스로부터의 하나 이상의 서열을 포함할 수 있다. 명명법의 최근의 변화는 인간 세포를 감염시킬 수 있는 3가지 아넬로바이러스를 알파토르크바이러스(TT), 베타토르크바이러스(TTM) 및 감마토르크바이러스(TTMD) 속의 아넬로바이러스과의 바이러스로 분류하였다. 일부 실시형태에서, 유전 요소는 토르크 테노 바이러스(TT), 환형 음성-센스 게놈을 갖는 비-외피, 단일-가닥 DNA 바이러스에 대하여 상동성 또는 동일성을 갖는 서열을 포함할 수 있다. 일부 실시형태에서, 유전 요소는 SEN 바이러스, 센티넬(Sentinel) 바이러스, TTV-유사 미니 바이러스 및 TT 바이러스에 대하여 상동성 또는 동일성을 갖는 서열을 포함할 수 있다. TT 바이러스 유전형 6, TT 바이러스 군, TTV-유사 바이러스 DXL1 및 TTV-유사 바이러스 DXL2를 포함하는 상이한 유형의 TT 바이러스가 기재된 바 있다. 일부 실시형태에서, 유전 요소는 더 작은 바이러스, 토르크 테노-유사 미니 바이러스(TTM) 또는 토르크 테노-유사 미디 바이러스(TTMD)로 지칭되는 TTV와 TTMV 사이의 게놈 크기를 갖는 제3 바이러스에 대하여 상동성 또는 동일성을 갖는 서열을 포함할 수 있다. 일부 실시형태에서, 유전 요소는 본원에, 예를 들어, 표 19에 기재된 뉴클레오티드 서열 중 어느 하나에 대하여 적어도 약 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% 및 99% 뉴클레오티드 서열 동일성을 갖는 비-병원성 바이러스로부터의 하나 이상의 서열 또는 서열의 단편을 포함할 수 있다.

[0675] 일부 실시형태에서, 유전 요소는 본원에, 예를 들어, 표 41에 기재된 뉴클레오티드 서열 중 어느 하나에 대하여 적어도 약 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% 및 99% 뉴클레오티드 서열 동일성을 갖는 실질적으로 비-병원성인 바이러스로부터의 하나 이상의 서열 또는 서열의 단편을 포함할 수 있다.

[0676] **표 41. 아넬로바이러스 및 그들의 서열의 예.** 수탁 번호 및 관련 서열 정보는 2018년 12월 11일자에 언급된 바와 같이 [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)에서 수득될 수 있다.

수탁 번호	설명
AB017613.1	토르크 테노 바이러스 16 DNA, 완전한 게놈, 단리주: TUS01
AB026345.1	ORF1 및 ORF2 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TRM1
AB026346.1	ORF1 및 ORF2 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TK16
AB026347.1	ORF1 및 ORF2 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TP1-3
AB028669.1	ORF1 및 ORF2 에 대한 TT 바이러스 유전자, 완전한 게놈, 단리주:TJN02
AB030487.1	pORF2a, pORF2b, pORF1 에 대한 TT 바이러스 유전자, 완전한 cds, 클론:JaCHCTC19
AB030488.1	pORF2a, pORF2b, pORF1 에 대한 TT 바이러스 유전자, 완전한 cds, 클론:JaBD89
AB030489.1	pORF2a, pORF2b, pORF1 에 대한 TT 바이러스 유전자, 완전한 cds, 클론:JaBD98
AB038340.1	ORF2s, ORF1, ORF3 에 대한 TT 바이러스 유전자, 완전한 cds
AB038622.1	ORF2, ORF1, ORF3 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TTVyon-LC011
AB038623.1	ORF2, ORF1, ORF3 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TTVyon-KC186
AB038624.1	ORF2, ORF1, ORF3 에 대한 TT 바이러스 유전자, 완전한 cds, 단리주:TTVyon-KC197
AB041821.1	VP1 에 대한 TT 바이러스 mRNA, 완전한 cds
AB050448.1	ORF1, ORF2, ORF3, ORF4 에 대한 토르크 테노 바이러스 유전자, 완전한 cds, 단리주: TYM9
AB060592.1	ORF1, ORF2, ORF3, ORF4 에 대한 토르크 테노 바이러스 유전자, 클론: SAa-39
AB060593.1	ORF1, ORF2, ORF3, ORF4 에 대한 토르크 테노 바이러스 유전자, 완전한 cds, 클론: SAa-38
AB060595.1	ORF1, ORF2, ORF3, ORF4 에 대한 TT 바이러스 유전자, 완전한 cds, 클론:SAj-30
AB060596.1	ORF1, ORF2, ORF3, ORF4 에 대한 TT 바이러스 유전자, 완전한 cds, 클론:SAf-09
AB064596.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: CT25F
AB064597.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: CT30F
AB064599.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT03F
AB064600.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT05F
AB064601.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT14F

[0677]

AB064602.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT19F
AB064603.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT41F
AB064604.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: CT39F
AB064606.1	토르크 테노 바이러스 DNA, 완전한 게놈, 단리주: JT33F
AB290918.1	토르크 테노 미디 바이러스 1 DNA, 완전한 게놈, 단리주: MD1-073
AF079173.1	TT 바이러스 균주 TTVCHN1, 완전한 게놈
AF116842.1	TT 바이러스 균주 BDH1, 완전한 게놈
AF122914.3	TT 바이러스 단리주 JA20, 완전한 게놈
AF122917.1	TT 바이러스 단리주 JA4, 완전한 게놈
AF122919.1	TT 바이러스 단리주 JA10 미공지된 유전자
AF129887.1	TT 바이러스 TTVCHN2, 완전한 게놈
AF247137.1	TT 바이러스 단리주 TUPB, 완전한 게놈
AF254410.1	TT 바이러스 ORF2 단백질 및 ORF1 단백질 유전자, 완전한 cds
AF298585.1	TT 바이러스 폴리시(Polish) 단리주 P/1C1, 완전한 게놈
AF315076.1	TTV-유사 바이러스 DXL1 미공지된 유전자
AF315077.1	TTV-유사 바이러스 DXL2 미공지된 유전자
AF345521.1	TT 바이러스 단리주 TCHN-G1 Orf2 및 Orf1 유전자, 완전한 cds
AF345522.1	TT 바이러스 단리주 TCHN-E Orf2 및 Orf1 유전자, 완전한 cds
AF345525.1	TT 바이러스 단리주 TCHN-D2 Orf2 및 Orf1 유전자, 완전한 cds
AF345527.1	TT 바이러스 단리주 TCHN-C2 Orf2 및 Orf1 유전자, 완전한 cds
AF345528.1	TT 바이러스 단리주 TCHN-F Orf2 및 Orf1 유전자, 완전한 cds
AF345529.1	TT 바이러스 단리주 TCHN-G2 Orf2 및 Orf1 유전자, 완전한 cds
AF371370.1	TT 바이러스 ORF1, ORF3, 및 ORF2 유전자, 완전한 cds
AJ620212.1	토르크 테노 바이러스, 단리주 tth6, 완전한 게놈
AJ620213.1	토르크 테노 바이러스, 단리주 tth10, 완전한 게놈
AJ620214.1	토르크 테노 바이러스, 단리주 tth11g2, 완전한 게놈
AJ620215.1	토르크 테노 바이러스, 단리주 tth18, 완전한 게놈
AJ620216.1	토르크 테노 바이러스, 단리주 tth20, 완전한 게놈
AJ620217.1	토르크 테노 바이러스, 단리주 tth21, 완전한 게놈
AJ620218.1	토르크 테노 바이러스, 단리주 tth3, 완전한 게놈
AJ620219.1	토르크 테노 바이러스, 단리주 tth9, 완전한 게놈
AJ620220.1	토르크 테노 바이러스, 단리주 tth16, 완전한 게놈
AJ620221.1	토르크 테노 바이러스, 단리주 tth17, 완전한 게놈

[0678]

AJ620222.1	토르크 테노 바이러스, 단리주 tth25, 완전한 게놈
AJ620223.1	토르크 테노 바이러스, 단리주 tth26, 완전한 게놈
AJ620224.1	토르크 테노 바이러스, 단리주 tth27, 완전한 게놈
AJ620225.1	토르크 테노 바이러스, 단리주 tth31, 완전한 게놈
AJ620226.1	토르크 테노 바이러스, 단리주 tth4, 완전한 게놈
AJ620227.1	토르크 테노 바이러스, 단리주 tth5, 완전한 게놈
AJ620228.1	토르크 테노 바이러스, 단리주 tth14, 완전한 게놈
AJ620229.1	토르크 테노 바이러스, 단리주 tth29, 완전한 게놈
AJ620230.1	토르크 테노 바이러스, 단리주 tth7, 완전한 게놈
AJ620231.1	토르크 테노 바이러스, 단리주 tth8, 완전한 게놈
AJ620232.1	토르크 테노 바이러스, 단리주 tth13, 완전한 게놈
AJ620233.1	토르크 테노 바이러스, 단리주 tth19, 완전한 게놈
AJ620234.1	토르크 테노 바이러스, 단리주 tth22g4, 완전한 게놈
AJ620235.1	토르크 테노 바이러스, 단리주 tth23, 완전한 게놈
AM711976.1	TT 바이러스 sle1957 완전한 게놈
AM712003.1	TT 바이러스 sle1931 완전한 게놈
AM712004.1	TT 바이러스 sle1932 완전한 게놈
AM712030.1	TT 바이러스 sle2057 완전한 게놈
AM712031.1	TT 바이러스 sle2058 완전한 게놈
AM712032.1	TT 바이러스 sle2072 완전한 게놈
AM712033.1	TT 바이러스 sle2061 완전한 게놈
AM712034.1	TT 바이러스 sle2065 완전한 게놈
AY026465.1	TT 바이러스 단리주 L01 ORF2 및 ORF1 유전자, 완전한 cds
AY026466.1	TT 바이러스 단리주 L02 ORF2 및 ORF1 유전자, 완전한 cds
DQ003341.1	토르크 테노 바이러스 클론 P2-9-02 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 cds
DQ003342.1	토르크 테노 바이러스 클론 P2-9-07 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 cds
DQ003343.1	토르크 테노 바이러스 클론 P2-9-08 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 cds
DQ003344.1	토르크 테노 바이러스 클론 P2-9-16 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 cds

[0679]

DQ186994.1	토르크 테노 바이러스 클론 P601 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ186995.1	토르크 테노 바이러스 클론 P605 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ186996.1	토르크 테노 바이러스 클론 BM1A-02 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ186997.1	토르크 테노 바이러스 클론 BM1A-09 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ186998.1	토르크 테노 바이러스 클론 BM1A-13 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ186999.1	토르크 테노 바이러스 클론 BM1B-05 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187000.1	토르크 테노 바이러스 클론 BM1B-07 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187001.1	토르크 테노 바이러스 클론 BM1B-11 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187002.1	토르크 테노 바이러스 클론 BM1B-14 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187003.1	토르크 테노 바이러스 클론 BM1B-08 ORF2 (ORF2) 유전자, 완전한 cds; 및 비기능적 ORF1 (ORF1) 유전자, 완전한 서열
DQ187004.1	토르크 테노 바이러스 클론 BM1C-16 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187005.1	토르크 테노 바이러스 클론 BM1C-10 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 cds
DQ187007.1	토르크 테노 바이러스 클론 BM2C-25 ORF2 (ORF2) 유전자, 완전한 cds; 및 비기능적 ORF1 (ORF1) 유전자, 완전한 서열
DQ361268.1	토르크 테노 바이러스 단리주 ViPi04 ORF1 유전자, 완전한 cds
EF538879.1	토르크 테노 바이러스 단리주 CSC5 ORF2 및 ORF1 유전자, 완전한 cds
EU305675.1	토르크 테노 바이러스 단리주 LTT7 ORF1 유전자, 완전한 cds
EU305676.1	토르크 테노 바이러스 단리주 LTT10 ORF1 유전자, 완전한 cds
EU889253.1	토르크 테노 바이러스 단리주 ViPi08 비기능적 ORF1 유전자, 완전한 서열
FJ392105.1	토르크 테노 바이러스 단리주 TW53A25 ORF2 유전자, 부분 cds; 및 ORF1 유전자, 완전한 cds

[0680]

FJ392107.1	토르크 테노 바이러스 단리주 TW53A27 ORF2 유전자, 부분 cds; 및 ORF1 유전자, 완전한 cds
FJ392108.1	토르크 테노 바이러스 단리주 TW53A29 ORF2 유전자, 부분 cds; 및 ORF1 유전자, 완전한 cds
FJ392111.1	토르크 테노 바이러스 단리주 TW53A35 ORF2 유전자, 부분 cds; 및 ORF1 유전자, 완전한 cds
FJ392112.1	토르크 테노 바이러스 단리주 TW53A39 ORF2 유전자, 부분 cds; 및 ORF1 유전자, 완전한 cds
FJ392113.1	토르크 테노 바이러스 단리주 TW53A26 ORF2 유전자, 완전한 cds; 및 비기능적 ORF1 유전자, 완전한 서열
FJ392114.1	토르크 테노 바이러스 단리주 TW53A30 ORF2 및 ORF1 유전자, 완전한 cds
FJ392115.1	토르크 테노 바이러스 단리주 TW53A31 ORF2 및 ORF1 유전자, 완전한 cds
FJ392117.1	토르크 테노 바이러스 단리주 TW53A37 ORF1 유전자, 완전한 cds
FJ426280.1	토르크 테노 바이러스 균주 SIA109, 완전한 게놈
FR751500.1	토르크 테노 바이러스 완전한 게놈, 단리주 TTV-HD23a (rheu215)
GU797360.1	토르크 테노 바이러스 클론 8-17, 완전한 게놈
HC742700.1	특허 WO2010044889 로부터의 서열 7
HC742710.1	특허 WO2010044889 로부터의 서열 17
JX134044.1	TTV-유사 미니 바이러스 단리주 TTMV_LY1, 완전한 게놈
JX134045.1	TTV-유사 미니 바이러스 단리주 TTMV_LY2, 완전한 게놈
KU243129.1	TTV-유사 미니 바이러스 단리주 TTMV-204, 완전한 게놈
KY856742.1	TTV-유사 미니 바이러스 단리주 쯤지양(zhenjiang), 완전한 게놈
LC381845.1	토르크 테노 바이러스 인간/일본/KS025/2016 DNA, 완전한 게놈
MH648892.1	아넬로비리데 종 단리주 ctde048, 완전한 게놈
MH648893.1	아넬로비리데 종 단리주 ctde007, 완전한 게놈
MH648897.1	아넬로비리데 종 단리주 ctcb038, 완전한 게놈
MH648900.1	아넬로비리데 종 단리주 ctfc019, 완전한 게놈
MH648901.1	아넬로비리데 종 단리주 ctbb022, 완전한 게놈
MH648907.1	아넬로비리데 종 단리주 ctef040, 완전한 게놈
MH648911.1	아넬로비리데 종 단리주 cthi018, 완전한 게놈
MH648912.1	아넬로비리데 종 단리주 ctea38, 완전한 게놈
MH648913.1	아넬로비리데 종 단리주 ctbg006, 완전한 게놈
MH648916.1	아넬로비리데 종 단리주 ctbg020, 완전한 게놈

[0681]

MH648925.1	아넬로비리대 종 단리주 ctc1019, 완전한 계놈
MH648932.1	아넬로비리대 종 단리주 ctid031, 완전한 계놈
MH648946.1	아넬로비리대 종 단리주 ctdb017, 완전한 계놈
MH648957.1	아넬로비리대 종 단리주 ctch017, 완전한 계놈
MH648958.1	아넬로비리대 종 단리주 ctbh011, 완전한 계놈
MH648959.1	아넬로비리대 종 단리주 ctbc020, 완전한 계놈
MH648962.1	아넬로비리대 종 단리주 ctif015, 완전한 계놈
MH648966.1	아넬로비리대 종 단리주 ctei055, 완전한 계놈
MH648969.1	아넬로비리대 종 단리주 ctjg000, 완전한 계놈
MH648976.1	아넬로비리대 종 단리주 ctcj064, 완전한 계놈
MH648977.1	아넬로비리대 종 단리주 ctbj022, 완전한 계놈
MH648982.1	아넬로비리대 종 단리주 ctbf014, 완전한 계놈
MH648983.1	아넬로비리대 종 단리주 ctbd027, 완전한 계놈
MH648985.1	아넬로비리대 종 단리주 ctch016, 완전한 계놈
MH648986.1	아넬로비리대 종 단리주 ctbd020, 완전한 계놈
MH648989.1	아넬로비리대 종 단리주 ctga035, 완전한 계놈
MH648990.1	아넬로비리대 종 단리주 cthf001, 완전한 계놈
MH648995.1	아넬로비리대 종 단리주 ctbd067, 완전한 계놈
MH648997.1	아넬로비리대 종 단리주 ctce026, 완전한 계놈
MH648999.1	아넬로비리대 종 단리주 ctfb058, 완전한 계놈
MH649002.1	아넬로비리대 종 단리주 ctjj046, 완전한 계놈
MH649006.1	아넬로비리대 종 단리주 ctcf030, 완전한 계놈
MH649008.1	아넬로비리대 종 단리주 ctbg025, 완전한 계놈
MH649011.1	아넬로비리대 종 단리주 ctbh052, 완전한 계놈
MH649014.1	아넬로비리대 종 단리주 ctba003, 완전한 계놈
MH649017.1	아넬로비리대 종 단리주 ctbb016, 완전한 계놈
MH649022.1	아넬로비리대 종 단리주 ctch023, 완전한 계놈
MH649023.1	아넬로비리대 종 단리주 ctbd051, 완전한 계놈
MH649028.1	아넬로비리대 종 단리주 ctbf9, 완전한 계놈
MH649038.1	아넬로비리대 종 단리주 ctbi030, 완전한 계놈
MH649039.1	아넬로비리대 종 단리주 ctca057, 완전한 계놈
MH649040.1	아넬로비리대 종 단리주 ctch033, 완전한 계놈
MH649042.1	아넬로비리대 종 단리주 ctjd005, 완전한 계놈

[0682]

MH649045.1	아넬로비리대 종 단리주 ctde021, 완전한 계놈
MH649051.1	아넬로비리대 종 단리주 ctde044, 완전한 계놈
MH649056.1	아넬로비리대 종 단리주 ctcc062, 완전한 계놈
MH649061.1	아넬로비리대 종 단리주 ctid009, 완전한 계놈
MH649062.1	아넬로비리대 종 단리주 ctde018, 완전한 계놈
MH649063.1	아넬로비리대 종 단리주 ctbf012, 완전한 계놈
MH649068.1	아넬로비리대 종 단리주 ctcc066, 완전한 계놈
MH649070.1	아넬로비리대 종 단리주 ctde011, 완전한 계놈
MH649077.1	아넬로비리대 종 단리주 ctbh034, 완전한 계놈
MH649083.1	아넬로비리대 종 단리주 ctde028, 완전한 계놈
MH649084.1	아넬로비리대 종 단리주 ctii061, 완전한 계놈
MH649085.1	아넬로비리대 종 단리주 ctch021, 완전한 계놈
MH649092.1	아넬로비리대 종 단리주 ctbg012, 완전한 계놈
MH649101.1	아넬로비리대 종 단리주 ctif053, 완전한 계놈
MH649104.1	아넬로비리대 종 단리주 ctei657, 완전한 계놈
MH649106.1	아넬로비리대 종 단리주 ctca015, 완전한 계놈
MH649114.1	아넬로비리대 종 단리주 ctbf050, 완전한 계놈
MH649122.1	아넬로비리대 종 단리주 ctde002, 완전한 계놈
MH649125.1	아넬로비리대 종 단리주 ctbb15, 완전한 계놈
MH649127.1	아넬로비리대 종 단리주 ctba013, 완전한 계놈
MH649137.1	아넬로비리대 종 단리주 ctbb000, 완전한 계놈
MH649141.1	아넬로비리대 종 단리주 ctbc019, 완전한 계놈
MH649142.1	아넬로비리대 종 단리주 ctid026, 완전한 계놈
MH649144.1	아넬로비리대 종 단리주 ctif004, 완전한 계놈
MH649152.1	아넬로비리대 종 단리주 ctcj13, 완전한 계놈
MH649156.1	아넬로비리대 종 단리주 ctei006, 완전한 계놈
MH649157.1	아넬로비리대 종 단리주 ctbd025, 완전한 계놈
MH649158.1	아넬로비리대 종 단리주 ctbf005, 완전한 계놈
MH649161.1	아넬로비리대 종 단리주 ctcf045, 완전한 계놈
MH649165.1	아넬로비리대 종 단리주 ctcc29, 완전한 계놈
MH649169.1	아넬로비리대 종 단리주 ctib021, 완전한 계놈
MH649172.1	아넬로비리대 종 단리주 ctbh857, 완전한 계놈
MH649174.1	아넬로비리대 종 단리주 ctbj049, 완전한 계놈

[0683]

MH649178.1	아넬로비리대 종 단리주 ctfc006, 완전한 계놈
MH649179.1	아넬로비리대 종 단리주 ctbe000, 완전한 계놈
MH649183.1	아넬로비리대 종 단리주 ctbb031, 완전한 계놈
MH649186.1	아넬로비리대 종 단리주 ctcb33, 완전한 계놈
MH649189.1	아넬로비리대 종 단리주 ctcc12, 완전한 계놈
MH649196.1	아넬로비리대 종 단리주 ctc060, 완전한 계놈
MH649199.1	아넬로비리대 종 단리주 ctbb017, 완전한 계놈
MH649203.1	아넬로비리대 종 단리주 ctch018, 완전한 계놈
MH649204.1	아넬로비리대 종 단리주 ctbj003, 완전한 계놈
MH649206.1	아넬로비리대 종 단리주 ctbg010, 완전한 계놈
MH649208.1	아넬로비리대 종 단리주 ctid008, 완전한 계놈
MH649209.1	아넬로비리대 종 단리주 ctbg056, 완전한 계놈
MH649210.1	아넬로비리대 종 단리주 ctca001, 완전한 계놈
MH649212.1	아넬로비리대 종 단리주 ctcf004, 완전한 계놈
MH649217.1	아넬로비리대 종 단리주 ctbe029, 완전한 계놈
MH649223.1	아넬로비리대 종 단리주 ctc016, 완전한 계놈
MH649224.1	아넬로비리대 종 단리주 ctce11, 완전한 계놈
MH649228.1	아넬로비리대 종 단리주 ctcf013, 완전한 계놈
MH649229.1	아넬로비리대 종 단리주 ctcb036, 완전한 계놈
MH649241.1	아넬로비리대 종 단리주 ctca027, 완전한 계놈
MH649242.1	아넬로비리대 종 단리주 ctbf003, 완전한 계놈
MH649254.1	아넬로비리대 종 단리주 ctjb007, 완전한 계놈
MH649255.1	아넬로비리대 종 단리주 ctbb023, 완전한 계놈
MH649256.1	아넬로비리대 종 단리주 ctca002, 완전한 계놈
MH649258.1	아넬로비리대 종 단리주 ctcg010, 완전한 계놈
MH649263.1	아넬로비리대 종 단리주 ctgh3, 완전한 계놈
MK012439.1	아넬로비리대 종 단리주 ctch000, 완전한 계놈
MK012440.1	아넬로비리대 종 단리주 ctjd008, 완전한 계놈
MK012448.1	아넬로비리대 종 단리주 ctch012, 완전한 계놈
MK012457.1	아넬로비리대 종 단리주 ctca009, 완전한 계놈
MK012458.1	아넬로비리대 종 단리주 ctcd015, 완전한 계놈
MK012485.1	아넬로비리대 종 단리주 ctfd011, 완전한 계놈
MK012489.1	아넬로비리대 종 단리주 ctba003, 완전한 계놈

[0684]

MK012492.1	아넬로비리데 종 단리주 ctbb005, 완전한 게놈
MK012493.1	아넬로비리데 종 단리주 ctcj014, 완전한 게놈
MK012500.1	아넬로비리데 종 단리주 ctcb001, 완전한 게놈
MK012504.1	아넬로비리데 종 단리주 ctcj010, 완전한 게놈
MK012516.1	아넬로비리데 종 단리주 ctcf003, 완전한 게놈
NC_038336.1	토르크 테노 바이러스 5 단리주 TCHN-C1 Orf2 및 Orf1 유전자, 완전한 cds
NC_038338.1	토르크 테노 바이러스 11 단리주 TCHN-D1 Orf2 및 Orf1 유전자, 완전한 cds
NC_038339.1	토르크 테노 바이러스 13 단리주 TCHN-A Orf2 및 Orf1 유전자, 완전한 cds
NC_038340.1	토르크 테노 바이러스 20 ORF4, ORF3, ORF2, ORF1 유전자, 완전한 cds, 클론: SAa-10
NC_038341.1	토르크 테노 바이러스 21 단리주 TCHN-B ORF2 및 ORF1 유전자, 완전한 cds
NC_038342.1	토르크 테노 바이러스 23 ORF2, ORF1 유전자, 완전한 cds, 단리주: s-TTV CH65-2
NC_038343.1	토르크 테노 바이러스 24 ORF4, ORF3, ORF2, ORF1 유전자, 완전한 cds, 클론: SAa-01
NC_038344.1	토르크 테노 바이러스 29 ORF2, ORF1, ORF3 유전자, 완전한 cds, 단리주: TTVyon-KC009
NC_038345.1	토르크 테노 미니 바이러스 10 단리주 LIL-y1 ORF2, ORF1, ORF3, 및 ORF4 유전자, 완전한 cds
NC_038346.1	토르크 테노 미니 바이러스 11 단리주 LIL-y2 ORF2, ORF1, 및 ORF3 유전자, 완전한 cds
NC_038347.1	토르크 테노 미니 바이러스 12 단리주 LIL-y3 ORF2, ORF1, ORF3, 및 ORF4 유전자, 완전한 cds
NC_038350.1	토르크 테노 미디 바이러스 3 단리주 2PoSMA ORF2 및 ORF1 유전자, 완전한 cds
NC_038351.1	토르크 테노 미디 바이러스 4 단리주 6PoSMA ORF2, ORF1, 및 ORF3 유전자, 완전한 cds
NC_038352.1	토르크 테노 미디 바이러스 5 DNA, 완전한 게놈, 단리주: MDJHem2
NC_038353.1	토르크 테노 미디 바이러스 6 DNA, 완전한 게놈, 단리주: MDJHem3-1
NC_038354.1	토르크 테노 미디 바이러스 7 DNA, 완전한 게놈, 단리주: MDJHem3-2
NC_038355.1	토르크 테노 미디 바이러스 8 DNA, 완전한 게놈, 단리주: MDJN1
NC_038356.1	토르크 테노 미디 바이러스 9 DNA, 완전한 게놈, 단리주: MDJN2
NC_038357.1	토르크 테노 미디 바이러스 10 DNA, 완전한 게놈, 단리주: MDJN14
NC_038358.1	토르크 테노 미디 바이러스 11 DNA, 완전한 게놈, 단리주: MDJN47
NC_038359.1	토르크 테노 미디 바이러스 12 DNA, 완전한 게놈, 단리주: MDJN51

NC_038360.1	토르크 테노 미디 바이러스 13 DNA, 완전한 게놈, 단리주: MDJN69
NC_038361.1	토르크 테노 미디 바이러스 14 DNA, 완전한 게놈, 단리주: MDJN97
NC_038362.1	토르크 테노 미디 바이러스 15 DNA, 완전한 게놈, 단리주: Pt-TTMDV210

[0685]

[0686]

[0687]

일부 실시형태에서, 유전 요소는 하나 이상의 비-아넬로바이러스, 예를 들어, 아테노바이러스, 헤르페스 바이러스, 폭스 바이러스, 백시니아 바이러스, SV40, 파필로마 바이러스, RNA 바이러스, 예컨대 레트로바이러스, 예를 들어, 렌티 바이러스, 단일-가닥 RNA 바이러스, 예를 들어, 간염 바이러스, 또는 이중-가닥 RNA 바이러스, 예를 들어, 로타바이러스로부터의 하나 이상의 서열에 대하여 상동성 또는 동일성을 갖는 하나 이상의 서열을 포함한다. 일부 실시형태에서, 재조합 레트로바이러스가 결합이 있기 때문에, 감염성 입자를 생성하도록 지원이 제공될 수 있다. 이러한 지원은 예를 들어, LTR 내의 조절 서열의 제어 하에 레트로바이러스의 구조적 유전자의 전부를 인코딩하는 플라스미드를 함유하는 헬퍼 세포주를 사용함으로써 제공될 수 있다. 본원에 기재된 아넬로백터를 복제하는 데 적합한 세포주는 해당 분야에 알려진 세포주, 예를 들어, 본원에 기재된 바와 같이 변형될 수 있는 A549 세포를 포함한다. 상기 유전 요소는 선택 가능한 마커를 인코딩하는 유전자를 추가로 함유하여, 요망되는 유전 요소가 확인되게 할 수 있다.

[0688]

일부 실시형태에서, 유전 요소는 서열이 제1 뉴클레오티드 서열에 의해 인코딩되는 폴리펩티드와 적어도 약

70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일하게 유지되거나, 또는 다르게 본 발명을 실시하는 데 유용하다면, 인코딩된 폴리펩티드에 아미노산 차이를 초래하는 비-침묵 돌연변이, 예를 들어, 염기 치환, 결실 또는 부가를 포함한다. 이와 관련하여, 일반적으로 전체 단백질 기능을 불활성화시키지 않는 것으로 인식되는 특정 보존적 아미노산 치환이 이루어질 수 있다: 예컨대 양으로 하전된 아미노산(및 그의 역)에 관하여, 라이신, 아르기닌 및 히스티딘; 음으로 하전된 아미노산(및 그의 역)에 관하여, 아스파르트산 및 글루탐산; 및 특정 군의 중성으로 하전된 아미노산(및 모든 경우에, 또한 그의 역)에 관하여, (1) 알라닌 및 세린, (2) 아스파라긴, 글루타민 및 히스티딘, (3) 시스테인 및 세린, (4) 글리신 및 프롤린, (5) 이소류신, 류신 및 발린, (6) 메티오닌, 류신 및 이소류신, (7) 페닐알라닌, 메티오닌, 류신 및 티로신, (8) 세린 및 트레오닌, (9) 트립토판 및 티로신, (10) 및 예를 들어, 티로신, 트립토판 및 페닐알라닌. 아미노산은 물리적 특성 및 이차 및 삼차 단백질 구조에 대한 기여에 따라 분류될 수 있다. 보존적 치환은 하나의 아미노산의 유사한 특성을 갖는 또 다른 아미노산으로의 치환으로서 해당 분야에 인식된다.

[0689] 동일한 또는 특정 백분율의 동일한 뉴클레오티드 또는 아미노산 잔기를 갖는 2개 이상의 핵산 또는 폴리펩티드 서열의 동일성(예를 들어, 비교 윈도우 또는 표기된 영역에 걸쳐 최대 상응성을 위해 비교되고 정렬되는 경우, 특정 영역에 걸쳐 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상의 동일성)은 하기 기재된 디폴트 파라미터와 함께 BLAST 또는 BLAST 2.0 서열 비교 알고리즘을 사용하여 또는 수동의 정렬 및 육안의 검사(예를 들어, NCBI 웹 사이트 [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/) 등 참조)에 의해 측정될 수 있다. 동일성은 또한 시험 서열의 상보물을 지칭할 수 있거나 이에 적용될 수 있다. 동일성은 또한 결실 및/또는 부가를 갖는 서열 및 치환을 갖는 서열을 포함한다. 본원에 기재된 바와 같이, 알고리즘은 갭 등을 설명한다. 동일성은 적어도 약 10개의 아미노산 또는 뉴클레오티드 길이, 약 15개의 아미노산 또는 뉴클레오티드 길이, 약 20개의 아미노산 또는 뉴클레오티드 길이, 약 25개의 아미노산 또는 뉴클레오티드 길이, 약 30개의 아미노산 또는 뉴클레오티드 길이, 약 35개의 아미노산 또는 뉴클레오티드 길이, 약 40개의 아미노산 또는 뉴클레오티드 길이, 약 45개의 아미노산 또는 뉴클레오티드 길이, 약 50개의 아미노산 또는 뉴클레오티드 길이 이상인 영역에 걸쳐 존재할 수 있다. 유전 암호가 축퇴성이기 때문에, 상동성 뉴클레오티드 서열은 임의의 수의 침묵 염기 변경, 즉, 그럼에도 불구하고 동일한 아미노산을 인코딩하는 뉴클레오티드 치환을 포함할 수 있다.

[0690] **단백질성 외부**

[0691] 일부 실시형태에서, 아넬로백터, 예를 들어, 합성 아넬로백터는 유전 요소를 봉입하는 단백질성 외부를 포함한다. 단백질성 외부는 포유동물에서 원치 않는 면역 반응을 유도하지 못하는 실질적으로 비-병원성인 외부 단백질을 포함할 수 있다. 아넬로백터의 단백질성 외부는 전형적으로 단백질성 외부를 구성하는 정20면체 형성으로 자가-어셈블될 수 있는 실질적으로 비-병원성인 단백질을 포함한다.

[0692] 일부 실시형태에서, 단백질성 외부 단백질은 아넬로백터의 유전 요소의 서열에 의해 인코딩된다(예를 들어, 유전 요소와 시스로 존재한다). 다른 실시형태에서, 단백질성 외부 단백질은 아넬로백터의 유전 요소로부터 분리된 핵산에 의해 인코딩된다(예를 들어, 유전 요소와 트랜스로 존재한다).

[0693] 일부 실시형태에서, 단백질, 예를 들어, 실질적으로 비-병원성인 단백질 및/또는 단백질성 외부 단백질은 하나 이상의 글리코실화된 아미노산, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상을 포함한다.

[0694] 일부 실시형태에서, 단백질, 예를 들어, 실질적으로 비-병원성인 단백질 및/또는 단백질성 외부 단백질은 적어도 하나의 친수성 DNA-결합 영역, 아르기닌-풍부 영역, 트레오닌-풍부 영역, 글루타민-풍부 영역, N-말단 폴리 아르기닌 서열, 가변 영역, C-말단 폴리글루타민/글루탐산염 서열 및 하나 이상의 이황화 가교를 포함한다.

[0695] 일부 실시형태에서, 단백질은 캡시드 단백질이며, 예를 들어, 본원에 기재된 캡시드 단백질을 인코딩하는 뉴클레오티드 서열, 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 분자 및/또는 캡시드 단백질 서열 중 어느 하나에 의해 인코딩되는 단백질에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 갖는다. 일부 실시형태에서, 단백질 또는 캡시드 단백질의 기능적 단편은 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 핵산에 대하여 적어도 약 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 뉴클레오티드 서열에 의해 인코딩된다.

[0696] 일부 실시형태에서, 아넬로백터는 캡시드 단백질 또는 캡시드 단백질의 기능적 단편을 인코딩하는 뉴클레오티드 서열, 또는 본원에 기재된 바와 같은 아넬로바이러스 ORF1 분자에 대하여 적어도 약 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함한다.

[0697] 일부 실시형태에서, 더 낮은 서열 동일성을 갖는 아미노산의 범위는 본원에 기재된 특성 중 하나 이상 및 세포/

조직/종 특이성(예를 들어, 항성)의 차이를 제공할 수 있다.

- [0698] 일부 실시형태에서, 아넬로박터는 단백질성 외부 내에 지질이 결합된다. 일부 실시형태에서, 아넬로박터에는 지질 이중층, 예를 들어, 바이러스 외피가 결합된다. 일부 실시형태에서, 아넬로박터의 내부는 단백질성 외부에 의해 완전히 덮인다(예를 들어, 100% 커버리지). 일부 실시형태에서, 아넬로박터의 내부는 단백질성 외부에 의해 100% 미만, 예를 들어, 95%, 90%, 85%, 80%, 70%, 60%, 50% 이하의 커버리지로 덮인다. 일부 실시형태에서, 단백질성 외부는 유전 요소가 아넬로박터에 보유되는 한, 예를 들어, 물, 이온, 펩티드 또는 소분자에 대한 투과 가능성을 허용하는 겹 또는 불연속부를 포함한다.
- [0699] 일부 실시형태에서, 단백질성 외부는 숙주 세포 내로의 유전 요소의 유입을 매개하기 위하여 숙주 세포를 특이적으로 인식하고/하거나 이에 결합하는 하나 이상의 단백질 또는 폴리펩티드, 예를 들어, 상보성 단백질 또는 폴리펩티드를 포함한다.
- [0700] 일부 실시형태에서, 단백질성 외부는 다음 중 하나 이상을 포함한다: 예를 들어, 본원에 기재된 바와 같은 예를 들어, ORF1 분자의, 아르기닌-풍부 영역, 젤리-롤 영역, N22 도메인, 초가변 영역, 및/또는 C-말단 도메인. 일부 실시형태에서, 단백질성 외부는 하기 중 하나 이상을 포함한다: 하나 이상의 글리코실화된 단백질, 친수성 DNA-결합 영역, 아르기닌-풍부 영역, 트레오닌-풍부 영역, 글루타민-풍부 영역, N-말단 폴리아르기닌 서열, 가변 영역, C-말단 폴리글루타민/글루탐산염 서열 및 하나 이상의 이황화 가교. 예를 들어, 단백질성 외부는 예를 들어, 본원에 기재된 바와 같은 아넬로바이러스 ORF1 핵산에 의해 인코딩된 단백질을 포함한다.
- [0701] 일부 실시형태에서, 단백질성 외부는 다음의 특징 중 하나 이상을 포함한다: 정20면체 대칭, 하나 이상의 숙주 세포 분자와 상호작용하는 분자를 인식하고/하거나 이에 결합하여, 숙주 세포로의 유입을 매개함, 지질 분자가 결합됨, 탄수화물이 결합됨, pH 및 온도 안정성임, 세계 저항성임 및 숙주에서 실질적으로 비-면역원성이거나 실질적으로 비-병원성임.
- [0702] **III. 이용 방법**
- [0703] 본원에 기재된 아넬로박터 및 아넬로박터를 포함하는 조성물은 예를 들어, 질병, 장애 또는 질환의 치료를 필요로 하는 대상체(예를 들어, 포유동물 대상체, 예를 들어, 인간 대상체)에서의 질병, 장애 또는 질환의 치료 방법에 사용될 수 있다. 본원에 기재된 약제학적 조성물의 투여는 예를 들어, 비경구(정맥내, 종양내, 복강내, 근육내, 강내 및 피하 포함) 투여로 이루어질 수 있다. 아넬로박터는 단독으로 또는 약제학적 조성물로서 제형화되어 투여될 수 있다.
- [0704] 아넬로박터는 단위-용량 조성물, 예컨대 단위 용량 비경구 조성물의 형태로 투여될 수 있다. 이러한 조성물은 일반적으로 혼합에 의해 제조되며, 비경구 투여를 위하여 적합하게 조정될 수 있다. 이러한 조성물은 예를 들어, 주사 및 주입 용액 또는 현탁액 또는 좌제 또는 에어로졸의 형태로 존재할 수 있다.
- [0705] 일부 실시형태에서, 예를 들어, 본원에 기재된 바와 같은 아넬로박터 또는 이를 포함하는 조성물의 투여는 아넬로박터에 의해 포함되는 유전 요소가 예를 들어, 대상체 내의 표적 세포로 운반되게 할 수 있다.
- [0706] 예를 들어, 이펙터(예를 들어, 내인성 또는 외인성 이펙터)를 포함하는 본원에 기재된 아넬로박터 또는 그의 조성물을 사용하여 이펙터를 세포, 조직 또는 대상체에 운반할 수 있다. 일부 실시형태에서, 아넬로박터 또는 그의 조성물을 사용하여, 이펙터를 골수, 혈액, 심장, GI 또는 피부에 운반한다. 본원에 기재된 아넬로박터 조성물의 투여에 의한 이펙터의 운반은 세포, 조직 또는 대상체에서 비코딩 RNA 또는 폴리펩티드의 발현 수준을 조절할 수 있다(예를 들어, 증가 또는 감소시킬 수 있다). 이러한 방식으로의 발현 수준의 조절은 이펙터가 운반되는 세포에서 기능적 활성의 변경을 초래할 수 있다. 일부 실시형태에서, 조절된 기능적 활성은 성질이 효소적, 구조적 또는 조절적일 수 있다.
- [0707] 일부 실시형태에서, 아넬로박터 또는 그의 카피는 세포로의 운반 후 24시간(예를 들어, 1일, 2일, 3일, 4일, 5일, 6일, 1주, 2주, 3주, 4주, 30일 또는 1개월)에 세포에서 검출 가능하다. 실시형태들에서, 아넬로박터 또는 그의 조성물은 표적 세포 상에서 효과를 매개하며, 효과는 적어도 1, 2, 3, 4, 5, 6 또는 7일, 2, 3 또는 4주 또는 1, 2, 3, 6 또는 12개월 동안 지속된다. (예를 들어, 아넬로박터 또는 그의 조성물이 외인성 단백질을 인코딩하는 유전 요소를 포함하는) 일부 실시형태에서, 효과는 1, 2, 3, 4, 5, 6 또는 7일, 2, 3 또는 4주 또는 1, 2, 3, 6 또는 12개월 미만 동안 지속된다.
- [0708] 본원에 기재된 아넬로박터 또는 아넬로박터를 포함하는 조성물로 치료될 수 있는 질병, 장애 및 질환의 예는 제한 없이, 면역 장애, 인터페론병증(예를 들어, I형 인터페론병증), 감염성 질병, 염증성 장애, 자가면역 질환,

암(예를 들어, 고형 종양, 예를 들어, 폐암, 비-소세포 폐암, 예를 들어, miR-625, 예를 들어, 카스파제-3에 반응성인 유전자를 발현하는 종양) 및 위장 장애를 포함한다. 일부 실시형태에서, 아넬로백터는 아넬로백터가 접촉되는 세포 내에서 활성 또는 기능을 조절한다(예를 들어, 증가 또는 감소시킨다). 일부 실시형태에서, 아넬로백터는 아넬로백터가 접촉되는 세포에서 분자(예를 들어, 핵산 또는 단백질)의 수준 또는 활성을 조절한다(예를 들어, 증가 또는 감소시킨다). 일부 실시형태에서, 아넬로백터는 아넬로백터가 접촉되는 세포, 예를 들어, 암 세포의 생존력을 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% 이상 감소시킨다. 일부 실시형태에서, 아넬로백터는 아넬로백터가 접촉되는 세포, 예를 들어, 암 세포의 생존력을 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% 이상 감소시키는 이펙터, 예를 들어, miRNA, 예를 들어, miR-625를 포함한다. 일부 실시형태에서, 아넬로백터는 예를 들어, 카스파제-3 활성을 증가시킴으로써 아넬로백터가 접촉되는 세포, 예를 들어, 암 세포의 아포토시스를 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% 이상 증가시킨다. 일부 실시형태에서, 아넬로백터는 예를 들어, 카스파제-3 활성을 증가시킴으로써, 아넬로백터가 접촉되는 세포, 예를 들어, 암 세포의 아포토시스를 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% 이상 증가시키는 이펙터, 예를 들어, miRNA, 예를 들어, miR-625를 포함한다.

[0709] **IV. 투여/운반**

[0710] 조성물(예를 들어, 본원에 기재된 바와 같은 아넬로백터를 포함하는 약제학적 조성물)은 약제학적으로 허용되는 부형제를 포함하도록 제형화될 수 있다. 약제학적 조성물은 선택적으로 하나 이상의 추가의 활성 물질, 예를 들어, 치료적 및/또는 예방적 활성 물질을 포함할 수 있다. 본 발명의 약제학적 조성물은 멸균 및/또는 발열원-부재일 수 있다. 약제의 제형화 및/또는 제조에서의 일반적인 고려사항은 예를 들어, 문헌[Remington: The Science and Practice of Pharmacy 21st ed., Lippincott Williams & Wilkins, 2005](본원에 참조로 포함됨)에서 찾을 수 있다.

[0711] 본원에 제공되는 약제학적 조성물의 설명이 주로 인간으로의 투여에 적합한 약제학적 조성물에 관한 것이지만, 이러한 조성물이 일반적으로 임의의 다른 동물로의, 예를 들어, 비-인간 동물, 예를 들어, 비-인간 포유동물로의 투여에 적합하다는 것을 당업자라면 이해할 것이다. 조성물이 다양한 동물로의 투여에 적합하게 만들기 위해 인간으로의 투여에 적합한 약제학적 조성물을 변형하는 것이 충분히 이해되며, 숙련된 수의학 약리학자는 존재한다면, 단지 일상적인 실험을 사용하여 이러한 변형을 설계하고/하거나 수행할 수 있다. 약제학적 조성물의 투여가 고려되는 대상체는 인간 및/또는 기타 영장류; 포유동물, 예컨대 상업적으로 관련된 포유동물, 예컨대 소, 돼지, 말, 양, 고양이, 개, 마우스 및/또는 랫트; 및/또는 조류, 예컨대 상업적으로 관련된 조류, 예컨대 가금류, 닭, 오리, 거위 및/또는 칠면조를 포함하지만, 이들에 제한되지 않는다.

[0712] 본원에 기재된 약제학적 조성물의 제형은 약리학 분야에 알려져 있거나, 이후에 개발되는 임의의 방법에 의해 제조될 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분이 부형제 및/또는 하나 이상의 다른 보조 성분과 회합되게 한 다음, 필요하면 그리고/또는 바람직하면, 생성물을 나누고/나누거나, 성형하고/하거나 패키징하는 단계를 포함한다.

[0713] 일 양태에서, 본 발명은 대상체로 아넬로백터를 운반하는 방법을 특징으로 한다. 당해 방법은 본원에 기재된 바와 같은 아넬로백터를 포함하는 약제학적 조성물을 대상체로 투여하는 단계를 포함한다. 일부 실시형태에서, 투여된 아넬로백터는 대상체에서 복제한다(예를 들어, 대상체의 바이러스체의 일부가 된다).

[0714] 약제학적 조성물은 야생형 또는 고유 바이러스 요소 및/또는 변형된 바이러스 요소를 포함할 수 있다. 아넬로백터는 하나 이상의 아넬로바이러스 서열(예를 들어, 핵산 서열 또는 그의 아미노산 서열을 인코딩하는 핵산 서열) 또는 이에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 및 99% 뉴클레오티드 서열 동일성을 갖는 서열을 포함할 수 있다. 아넬로백터는 하나 이상의 아넬로바이러스 서열(예를 들어, 아넬로바이러스 ORF1 핵산 서열)에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 및 99% 서열 동일성을 갖는 핵산 서열을 포함하는 핵산 분자를 포함할 수 있다. 아넬로백터는 아넬로바이러스 아미노산 서열(예를 들어, 아넬로바이러스 ORF1 분자의 아미노산 서열)에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 및 99% 서열 동일성을 갖는 아미노산 서열을 인코딩하는 핵산 분자를 포함할 수 있다. 아넬로백터는 아넬로바이러스 아미노산 서열(예를 들어, 아넬로바이러스 ORF1 분자의 아미노산 서열)에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 및 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드를 포함할 수 있다.

[0715] 일부 실시형태에서, 아넬로백터는 내인성 유전자 및 단백질 발현을 기준, 예를 들어, 건강한 대조군에 비하여

예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% 이상 증가(자극)시키기에 충분하다. 특정 실시형태에서, 아넬로백터는 내인성 유전자 및 단백질 발현을 기준, 예를 들어, 건강한 대조군에 비하여 예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% 이상 감소(저해)시키기에 충분하다.

[0716] 일부 실시형태에서, 아넬로백터는 숙주 또는 숙주 세포에서 하나 이상의 바이러스 특성, 예를 들어, 항성, 감염성, 면역억제/활성화를 기준, 예를 들어, 건강한 대조군에 비하여 예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% 이상 저해/향상시킨다.

[0717] 일부 실시형태에서, 바이러스 유전 정보에서 표현되지 않는 하나 이상의 바이러스 주를 추가로 포함하는 억제학적 조성물이 대상체에게 투여된다.

[0718] 일부 실시형태에서, 본원에 기재된 아넬로백터를 포함하는 억제학적 조성물은 바이러스 감염을 조절하기에 충분한 용량 및 시간으로 투여된다. 바이러스 감염의 일부 비제한적인 예는 아데노-연관 바이러스, 아이치(Aichi) 바이러스, 호주 박쥐 리사 바이러스(lyssavirus), BK 폴리오마바이러스(polyomavirus), 반나 바이러스(Banna virus), 바르마 포레스트(Barmah forest) 바이러스, 부니아무에라(Bunyamwera) 바이러스, 부니아바이러스 라 크로세(Bunyavirus La Crosse), 부니아바이러스 스노우슈 하레(Bunyavirus snowshoe hare), 세르코피테신 헤르페스바이러스(Cercopithecine herpesvirus), 찬디푸라 바이러스(Chandipura virus), 치쿤구니아 바이러스(Chikungunya virus), 코스바이러스(Cosavirus) A, 우두 바이러스, 콕사키바이러스(Coxsackievirus), 크리민-콩고 출혈성 열 바이러스(Crimean-Congo hemorrhagic fever virus), 땡기(Dengue) 바이러스, 도리(Dhori) 바이러스, 더그비(Dugbe) 바이러스, 두벤하계(Duvenhage) 바이러스, 동부 말 뇌염 바이러스, 에볼라바이러스, 에코 바이러스(Echovirus), 뇌심근염(Encephalomyocarditis) 바이러스, 엡스테인-바 바이러스, 유럽 박쥐 리사바이러스, GB 바이러스 C/G형 간염 바이러스, 한탄(Hantaan) 바이러스, 헨드라(Hendra) 바이러스, A형 간염 바이러스, B형 간염 바이러스, C형 간염 바이러스, E형 간염 바이러스, 간염 델타 바이러스, 마두 바이러스, 인간 아데노 바이러스, 인간 아스트로바이러스, 인간 코로나바이러스, 인간 사이토메갈로바이러스, 인간 엔테로바이러스 68, 인간 엔테로바이러스 70, 인간 헤르페스바이러스 1, 인간 헤르페스바이러스 2, 인간 헤르페스바이러스 6, 인간 헤르페스바이러스 7, 인간 헤르페스바이러스 8, 인간 면역결핍 바이러스, 인간 파필로마바이러스 1, 인간 파필로마바이러스 2, 인간 파필로마바이러스 16, 인간 파필로마바이러스 18, 인간 파라인플루엔자, 인간 파보바이러스 B19, 인간 호흡기 융합 바이러스, 인간 리노바이러스, 인간 SARS 코로나바이러스, 인간 스푸마레트로바이러스(spumaretrovirus), 인간 T-림포트로픽 바이러스, 인간 토로바이러스(torovirus), 인플루엔자 A 바이러스, 인플루엔자 B 바이러스, 인플루엔자 C 바이러스, 이스파한(Isfahan) 바이러스, JC 폴리오마바이러스, 일본 뇌염 바이러스, 주닌 아레나바이러스(Junin arenavirus), KI 폴리오마바이러스, 쿤진(Kunjin) 바이러스, 라고스(Lagos) 박쥐 바이러스, 레이크 빅토리아 마르부르그바이러스(Lake Victoria marburgvirus), 란가트(Langat) 바이러스, 라사(Lassa) 바이러스, 로즈데일(Lordsdale) 바이러스, 라우핑 일(Louping ill) 바이러스, 림프구 맥락수막염 바이러스, 마추포(Machupo) 바이러스, 마이아로(Mayaro) 바이러스, MERS 코로나바이러스, 홍역 바이러스, 멘고(Mengo) 뇌심근염 바이러스, 메르켈 세포 폴리오마바이러스(Merkel cell polyomavirus), 모콜라(Mokola) 바이러스, 전염성 연속종 바이러스, 원두 바이러스, 볼거리 바이러스, 무레이 밸리(Murray valley) 뇌염 바이러스, 뉴욕 바이러스, 니파(Nipah) 바이러스, 노르워크(Norwalk) 바이러스, 오농-농(O'nyong-nyong) 바이러스, 올프(Orf) 바이러스, 오로푸스(Oropouche) 바이러스, 피친데(Pichinde) 바이러스, 폴리오바이러스, 폰타 토로 플레보바이러스(Punta toro phlebovirus), 푸말라(Puumala) 바이러스, 광견병 바이러스, 리프트게곡열 바이러스, 로사바이러스(Rosavirus) A, 로스 강(Ross river) 바이러스, 로타바이러스 A, 로타바이러스 B, 로타바이러스 C, 풍진 바이러스, 사기야마(Sagiyama) 바이러스, 살리바이러스(Salivirus) A, 모래파리 열 시실리안(Sandfly fever sicilian) 바이러스, 삿포로(Sapporo) 바이러스, 셴리키 포레스트(Semliki forest) 바이러스, 서울(Seoul) 바이러스, 유인원 포미(Simian foamy) 바이러스, 유인원 바이러스 5, 신드비스(Sindbis) 바이러스, 사우샘프턴(Southampton) 바이러스, 세인트 루이스 뇌염(St. louis encephalitis) 바이러스, 틱-본 포와산(Tick-borne powassan) 바이러스, 토르크 테노 바이러스, 토스카나(Toscana) 바이러스, 우우쿠니에미(Uukuniemi) 바이러스, 백시니아 바이러스, 바리셀라-조스터(Varicella-zoster) 바이러스, 바리올라(Variola) 바이러스, 베네수엘란(Venezuelan) 말 뇌염 바이러스, 수포성 구내염 바이러스, 서부 말 뇌염 바이러스, WU 폴리오마바이러스, 웨스트 나일(West Nile) 바이러스, 야바(Yaba) 원숭이 종양 바이러스, 야바-유사 질병 바이러스, 황열 바이러스 및 지카 바이러스를 포함한다. 특정 실시형태에서, 아넬로백터는 대상체에 이미 존재하는 바이러스를 예를 들어, 기준에 비하여 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% 이상 증가하고/하거나 대체하기에 충분하다. 특정 실시형태에서, 아넬로백터는 만성 또는 급성 바이러스 감염과 경쟁하기에 충분하다. 특정 실시형태에서, 아넬로백터는 바이러스 감염으로부터 보호하기 위하여 예방적으로 투여될 수 있다(예를 들어, 바이러스전(provirotic)). 일부 실시형태에서, 아넬로백터는 (예를 들어, 표현형, 바이러스

수준, 유전자 발현, 다른 바이러스와의 경쟁, 병태 등을 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% 이상 조절하기에 충분한 양으로 존재한다. 일부 실시형태에서, 치료, 치료하는 및 그의 동원어 (cognate)는, 예를 들어, 질병, 병리학적 상태 또는 장애를 개선시키거나, 호전시키거나, 안정화시키거나, 예방하거나, 치유하려는 의도로 (예를 들어, 아넬로백터, 예를 들어, 본원에 기재된 바와 같이 제조된 아넬로백터를 투여함으로써) 대상체의 의학적 관리를 포함한다. 일부 실시형태에서, 치료는 적극 치료(질병, 병리학적 상태 또는 장애를 개선시키기 위한 치료), 원인 치료(연관된 질병, 병리학적 상태 또는 장애의 원인에 대한 치료), 완화 치료(증상의 경감을 위해 설계된 치료), 예방적 치료(연관된 질병, 병리학적 상태 또는 장애의 발달을 예방하거나, 최소화시키거나, 부분적으로 또는 완전히 억제하기 위한 치료) 및/또는 지지 치료(다른 요법을 보완하기 위해 사용된 치료)를 포함한다.

[0719] 본원에 언급된 모든 참고문헌 및 간행물은 본원에 참조로 포함된다. 다음의 실시예는 본 발명의 일부 실시형태를 추가로 예시하기 위해 제공되지만, 본 발명의 범주를 제한하고자 하지 않으며; 당업자에게 알려져 있는 다른 절차, 방법 또는 기법이 대안적으로 사용될 수 있다는 것이 그들의 예시적인 성질에 의해 이해될 것이다.

[0720] **실시예**

[0721] **목차**

[0722] 실시예 1: 배칼로바이러스를 사용하여 생성된 ORF1을 사용한 아넬로백터의 시험관 내 어셈블리

[0723] 실시예 2: mRNA를 캡시드화하는 고리2 ORF1-기반 아넬로백터의 시험관 내 어셈블리

[0724] 실시예 3: 변형된 ORF1 단백질을 사용한 mRNA-캡시드화 아넬로백터의 시험관 내 어셈블리

[0725] 실시예 4: 변형된 mRNA를 사용한 mRNA-캡시드화 아넬로백터의 시험관 내 어셈블리

[0726] 실시예 5: 아넬로바이러스 ORF1 캡시드 단백질의 구조 분석

[0727] **실시예 1: 배칼로바이러스를 사용하여 생성된 ORF1을 사용한 아넬로백터의 시험관 내 어셈블리**

[0728] 이 실시예에서, 아넬로바이러스 단백질(예를 들어, ORF1)의 발현에 적합한 배칼로바이러스 작제물을 시험관 내 어셈블리에 의해 생성하였다.

[0729] 제1 실시예에서, N-말단 HIS<sub>6</sub>-태그(HIS-ORF1)에 융합된 고리2 ORF1을 인코딩하는 DNA를 곤충 발현을 위해 코돈 최적화하였고, 제조업체 지침(ThermoFisher Scientific)에 따라 배칼로바이러스 발현 벡터 pFASTbac 시스템 내로 클로닝하였다. 10 리터의 곤충 세포 배양액(Sf9)을 고리2 HIS-ORF1 배칼로바이러스로 감염시키고, 감염 3-일 후 원심분리에 의해 세포를 수확하였다. 세포를 용해시키고, 용해물을 킬레이팅 수지 컬럼을 사용한, 현장에서 표준 방법을 사용하여 정제하였다. HIS-ORF1을 함유하는 용출 분획을 투석하고, DNase로 처리하여, 숙주 세포 DNA를 소화시켰다. 생성된 물질을 킬레이팅 수지 컬럼을 사용하여 다시 정제하고, ORF1을 함유하는 분획을 핵산 캡시드화 및 바이러스 벡터 정제를 위해 보유하였다. ORF1-함유 분획도 음성 염색 전자 현미경에 의해 분석하였다.

[0730] 제2 실시예에서, N-말단 HIS<sub>6</sub>-태그(HIS-ORF1)에 융합된 고리10 ORF1을 인코딩하는 DNA를 곤충 발현을 위해 코돈 최적화하였고, 제조업체 지침(ThermoFisher Scientific)에 따라 배칼로바이러스 발현 벡터 pFASTbac 시스템 내로 클로닝하였다. 곤충 세포(Sf9)를 고리 HIS-ORF1 배칼로바이러스로 감염시키고, 감염 3-일 후 원심분리에 의해 세포를 수확하였다. 세포를 용해시키고, 단백질을 킬레이팅 수지 친화도 컬럼(HisTrap, GE Healthcare)을 사용한, 현장에서 표준 방법을 사용하여 정제하였다. 생성된 물질을 다시 헤파린 친화도 컬럼(Heparin HiTrap, GE Healthcare)을 사용하여 정제하고, ORF1을 함유하는 분획을 음성 염색 전자 현미경에 의해 분석하였다.

[0731] 제3 실시예에서, N-말단 HIS<sub>6</sub>-플래그-태그(HIS-Flag-Vp1) 및 헬퍼 단백질(Vp2)에 융합된 닭 빈혈 바이러스(CAV) 캡시드 단백질(Vp1)을 인코딩하는 DNA를 포유동물 발현을 위해 코돈 최적화하였고, CMV 프로모터를 사용하는 포유동물 발현 벡터 내로 클로닝하였다. 포유동물 세포(293expi)를 CAV Vp1 및 Vp2 발현 벡터로 형질주입시켰다. 감염 3-일 후 원심분리에 의해 세포를 수확하였다. 세포를 용해시키고, 킬레이트화 및 헤파린 정제 프로세스를 사용하여 용해물을 정제하였다. 닭 빈혈 바이러스(CAV) Vp1을 함유하는 용출 분획을 음성 염색 전자현미경에 의해 분석하였다.

[0732] 도 1에 나타난 바와 같이, 고리 2 ORF1 및 고리 10 ORF1 둘 모두는 약 35 nm 바이러스-유사 입자를 형성하는 경향을 보였다.

- [0733] **핵산 캡시드화 및 바이러스 벡터 정제:** 고리 ORF1(야생형 단백질, 키메라 단백질 또는 그의 단편)은 VLP 또는 바이러스 캡시드를 분리하여 핵산 카고와 재어셈블리할 수 있도록 하기에 충분한 조건으로 처리될 것이다. 핵산 카고는 예를 들어, 치료제로서 운반하고자 하는 관심 유전자를 인코딩하는 RNA로 정의될 수 있다. 정의된 농도의 핵산 카고는 정의된 농도의 고리 ORF1과 조합되고, 핵산 캡시드화를 허용하기에 충분한 조건으로 처리될 것이며, 바이러스 벡터로서 정의된 생성된 입자는 후속적으로 표준 바이러스 정제 절차를 사용하여 정제될 것이다.
- [0734] **실시예 2: mRNA를 캡시드화하는 고리2 ORF1-기반 아넬로벡터의 시험관 내 어셈블리**
- [0735] 고리2 ORF1은 바이러스 입자 또는 VLP를 분산된 단백질 또는 캡소머로 분리하기 위해, 500 mM NaCl을 갖는 Tris pH 8.0, 0.1% SDS와 함께 500 mM NaCl을 갖는 Tris pH 8.0, 150 mM NaCl을 갖는 CAPS 완충액 pH 10.5, 500 mM NaCl을 갖는 CAPS 완충액 pH 10.5 또는 0.1% SDS와 함께 500 mM NaCl을 갖는 CAPS 완충액 pH 10.5를 포함하는 이동상을 이용하여 크기 배제 크로마토그래피(SEC)에 의해 정제된다.
- [0736] 제1 실시예에서, ORF1은 mRNA, 형광 표지된 mRNA 또는 벡터 형성을 유도하기에 적절한 것으로 실시예 1에 나타난 ssDNA의 세그먼트에 화학적으로 컨주게이션된 mRNA 트랜스진과 혼합된다. 바이러스 벡터는 (예를 들어, 유지된 형광 흡수에 의해 측정된 바와 같은) 아넬로벡터 캡시드화 RNA를 단리하기 위해 Tris pH 8.0 완충액을 사용하여 투석 및 SEC 정제를 통해 형성된다. 아넬로벡터 어셈블리는 DLS 또는 전자 현미경과 같은 생물물리학적 평가에 의해 추가로 평가된다.
- [0737] 제2 실시예에서, 정제된 ORF1은 0.1% SDS와 함께 1 M NaCl로 처리되어 올리고머 또는 VLP를 분산된 단백질 또는 캡소머로 분리한다. 그런 다음, ORF1은 관심 유전자(예를 들어, 리포터 유전자, 예를 들어, GFP, mCherry; 또는 관심 이펙터, 예를 들어, EPO)를 번역하는 mRNA와 같은 mRNA와 혼합되고, VLP 형성을 허용하기 위해 150 mM NaCl을 갖는 Tris pH 8.0에 대해 투석된다. 후속 복합체는 AV 벡터 캡시드화 mRNA를 단리하기 위해 Tris pH 8.0 완충액을 사용하는 SEC에 의해 정제된다. 아넬로벡터 벡터 어셈블리는 예를 들어, 세포를 형질도입하고 리포터 유전자(예를 들어, mCherry 또는 GFP)의 발현을 관찰함으로써 또는 관심 이펙터의 발현을 통해(예를 들어, EPO와 같은 유전자의 발현을 검출하기 위해 ELISA를 사용하여) 시험관 내 또는 생체 내 관독에 의해 추가로 평가될 수 있다.
- [0738] *GFP mRNA를 캡시드화하는 고리2 ORF1-기반 아넬로벡터의 시험관 내 어셈블리*
- [0739] 추가 실시예에서, 고리2 ORF1 단백질을 곤충 세포에서 전장 단백질로서 발현시키고, 어셈블된 VLP를 해파린 친화도 컬럼에 이어서, Tris 완충액 이동상을 사용하는 크기 배제 크로마토그래피(SEC)로 정제하였다. 단리된 고리2 ORF1 단백질로부터 형성된 VLP를 음성 염색 전자 현미경(EM)으로 관찰하였고, 이는  $10^{10}$ 개의 입자/ml(pts/ml; 도 4a)의 추정된 입자 역가를 가졌다. VLP를 분해하기 위해 2몰(2 M) 요소를 사용하여 VLP를 처리하였다. EM에 의한 재이미징은 관찰된 VLP가 없음을 보여주었다(도 4b). 이어서, 요소-처리된 VLP를 투석하여, mRNA의 부재(도 4c) 또는 약 10x 과량의 GFP를 인코딩하는 mRNA의 존재(도 4d 및 4e) 하에 요소를 제거하였다. 요소로 처리되고 mRNA의 부재 하에 투석된 VLP 샘플의 경우, EM에 의해 입자가 거의 관찰되지 않았다( $10^8$ 개 미만의 입자/ml; 도 4c). 대조적으로, 과량의 mRNA의 존재 하에서의 투석은 EM에 의해 실질적으로 더 높은 역가의 입자(약  $10^9$  내지  $10^{10}$ 개의 입자/ml; 도 4d 및 4e)의 관찰을 초래하였다. 이들 데이터는 VLP의 분해 및 재어셈블리가 mRNA의 존재 하에서 보다 효율적임을 입증하고, 아넬로바이러스 ORF1 단백질이 아넬로벡터를 형성하기 위해 시험관 내에서 mRNA를 캡시드화하는데 사용될 수 있음을 나타낸다.
- [0740] **실시예 3: 변형된 ORF1 단백질을 사용한 mRNA-캡시드화 아넬로벡터의 시험관 내 어셈블리**
- [0741] 이 실시예에서, mRNA에 결합하는 접촉 잔기를 보유하도록 ORF1 단백질을 변형시킴으로써 mRNA 유전 요소의 패키징이 개선된다. 이 실시예에서, ssDNA 및/또는 N-말단 아르기닌-풍부 모티프(ARM)와 접촉하는 ssDNA 접촉 잔기 및/또는 젤리롤 베타 가닥을 mRNA 결합 바이러스 단백질 또는 기타 mRNA-결합 단백질의 구성성분으로 대체할 수 있어, mRNA의 효율적인 결합 및 패키징을 허용하였다. 이 mRNA-결합 키메라 ORF1은 0.1% SDS를 갖는 1 M NaCl로 처리되어, 올리고머 또는 VLP를 분산된 단백질 또는 캡소머로 분리한다. 그런 다음, 키메라 ORF1은 관심 유전자(예를 들어, 리포터 유전자, 예를 들어, GFP, mCherry; 또는 관심 이펙터, 예를 들어, EPO)를 인코딩하는 mRNA와 같은 mRNA와 혼합되고, VLP 형성을 허용하기 위해 150 mM NaCl을 갖는 Tris pH 8.0에 대해 투석된다. 후속 복합체는 아넬로벡터 캡시드화 mRNA를 단리하기 위해 Tris pH 8.0 완충액을 사용하는 SEC에 의해 정제된다. 아넬로벡터 어셈블리는 예를 들어, 세포를 형질도입하고 리포터 유전자(예를 들어, mCherry 또는 GFP)의 발현을

관찰함으로써 또는 관심 이펙터의 발현을 통해(예를 들어, EPO와 같은 유전자의 발현을 검출하기 위해 ELISA를 사용하여) 시험관 내 또는 생체 내 판독에 의해 추가로 평가될 수 있다.

- [0742] **ORF1 분자에 대한 예시적인 변형:** 본원에 기재된 방법에 사용될 수 있는 고리 ORF1 분자는 예를 들어, 여러 야생형 아넬로바이러스 ORF1 단백질; CAV 캡시드 단백질(VP1) 변이체; 어셈블리 효율, 수율 또는 안정성을 개선시키기 위한 돌연변이를 보유하는 아넬로바이러스 ORF1 단백질; 및 키메라 ORF1 균주 또는 그의 기능적 단편을 포함한다. 일부 예에서, 친화도 태그가 예를 들어, N-말단(SEQ ID NO: 561 내지 562)에서 ORF1 분자에 부착된다. 일부 예에서, ORF1 분자는 태그가 부착되지 않은 단백질이다. 고리 ORF1 분자는 단독으로 또는 비제한적으로 아넬로바이러스 ORF2 및/또는 ORF3 단백질을 포함하는 임의의 수의 헬퍼 단백질과 조합하여 발현될 수 있다.
- [0743] 어셈블리 효율을 개선시키기 위한 돌연변이를 보유하는 고리 ORF1 단백질은 예를 들어, pH 민감성 핵산 결합을 허용하여 입자 어셈블리를 촉발할 수 있는, ARG 아암의 pI를 변경하기 위해 N-말단 아르기닌-풍부 모티프(ARM) 내로 도입된 돌연변이를 보유하는 ORF1 단백질을 포함할 수 있으나, 이에 제한되지 않는다 (SEQ ID NO: 563 내지 565). 안정성을 개선시키는 ORF1 돌연변이는 예를 들어, 프로토머 표면의 소수성 상태를 변경시키고/변경시키거나 캡시드 형성이 보다 열역학적으로 선호되도록 만들기 위해 예를 들어, 표준 젤리를 베타-배럴(F 및 G 베타 가닥)의 베타 가닥 F 및 G와 접촉하는 인터프로토머에 대한 돌연변이를 포함할 수 있다.
- [0744] 키메라 ORF1 단백질은 또 다른 캡시드 단백질, 예컨대, BFDV 캡시드 단백질, E형 간염 캡시드 단백질로부터의 비슷한 부분(예를 들어, ARG 아암 및/또는 또는 F 및 G 베타 가닥, 또는 그의 비슷한 구성성분)으로 대체된 이들의 서열의 일부 또는 일부들을 갖는 ORF1 단백질을 포함할 수 있으나, 이에 제한되지 않는다. 키메라 ORF1 단백질은 다른 아넬로바이러스 ORF1 단백질의 비슷한 부분(예컨대, 젤리 롤 단편 또는 고리 9 ORF1의 비슷한 부분으로 대체된 고리 2 ORF1의 C-말단 부분; 예를 들어, SEQ ID NO: 568 내지 575 참고)으로 대체된 이들의 서열의 일부 또는 일부들을 갖는 ORF1 단백질을 포함할 수도 있다.
- [0745] 일반적으로 ORF1 분자는 비제한적으로 킬레이팅 정제, 헤파린 정제, 구배 침강 정제 및/또는 SEC 정제를 포함하는 정제 기술을 사용하여 정제될 수 있다.
- [0746] **실시예 4: 변형된 mRNA를 사용한 mRNA-캡시드화 아넬로바이러스의 시험관 내 어셈블리**
- [0747] 이 실시예에서, mRNA 기반 유전 요소의 캡시드화는 mRNA 분자를 ssDNA에 결합시키거나 백본의 섹션이 야생형 ORF1의 ssDNA 접촉 잔기에 대한 결합을 허용하는 방식으로 mRNA 이식유전자를 변형함으로써 최적화된다. mRNA는 일반적으로 리포터 유전자(예를 들어, GFP 또는 mCherry) 및/또는 이펙터 유전자(예를 들어, EPO)와 같은 관심 유전자를 인코딩한다.
- [0748] 일 실시예에서, 당 사슬 백본 덕분에 ORF1에 결합할 수 있지만 mRNA와 비공유적으로 쌍을 이룰 수도 있는 변형된 ssDNA는 mRNA/DNA 복합체를 생성하기 위해 관심 있는 mRNA와 혼합된다. 그런 다음 이 mRNA/DNA 복합체는 고리 ORF1을 사용하여 캡시드화되어 예를 들어, 하기에 설명된 대로 아넬로바이러스를 형성할 수 있다.
- [0749] 다른 실시예에서 mRNA 분자는, 운반될 유전자(예를 들어, 리포터 유전자 또는 이펙터 유전자)를 인코딩하는 mRNA의 일부분을 간직하면서도 ORF1과의 결합 및 캡시드화가 가능하도록 하는 DNA 골격을 보유한 mRNA 분자의 절편 또는 절편들을 사용하여 합성된다. 그런 다음 이 mRNA/DNA 혼성 분자는 예를 들어, 아래에 설명된 바와 같이 아넬로바이러스를 형성하기 위해 고리 ORF1을 사용하여 캡시드화될 수 있다.
- [0750] **시험관 내 어셈블리에 의한 캡시드화:** 위에서 설명한 mRNA/DNA 유전 요소는 시험관 내 어셈블리에 의해 캡시드화된다. 간략하게, 아넬로바이러스 ORF1은 올리고머 또는 VLP를 분산된 단백질 또는 캡소머로 분리시키기 위해 0.1% SDS를 갖는 1 M NaCl로 처리된다. 그런 다음 ORF1을 합성 mRNA 복합체 또는 혼성 분자와 혼합하고 150 mM NaCl로 Tris pH 8.0에 대해 투석하여 VLP 형성을 허용한다. 후속 입자는 아넬로바이러스 캡시드화 mRNA를 단리하기 위해 Tris pH 8.0 버퍼를 사용하여 SEC에 의해 정제된다. 아넬로바이러스 어셈블리는 예를 들어, 본원에 기재된 바와 같이 세포를 형질도입하고 리포터 유전자 또는 이펙터 유전자의 발현을 관찰함으로써 시험관내 또는 생체내 판독에 의해 추가로 평가될 수 있다.
- [0751] **실시예 5: 아넬로바이러스 ORF1 캡시드 단백질의 구조 분석**
- [0752] 아넬로바이러스는 조류 병원체인 부리 및 깃털 질병 바이러스(BFDV) 또는 닭 빈혈 바이러스(CAV)와 같은 잘 특성화된 다른 바이러스에 대해 예측되는 구조적 특징을 공유한다.
- [0753] 아넬로바이러스 ORF1 캡시드 단백질은 BFDV의 것과 유사한 N-말단 ARM 서열을 함유한다. 2차 구조 예측은 균주에 좌우되는 ORF1의 처음 약 250개의 잔기가 8개의 예측된 베타 가닥을 포함함을 보여주었다(도 3). 젤리 롤 도

메인 명명 규칙에 따라 B에서 I로 명명된 8개의 ORF1의 예측된 베타 가닥이 BFDV 및 Hep E의 캡시드 단백질에 대한 2차 구조 예측에 의해 정렬 가이드될 때, ORF1에 보존된 라이신 및 아르기닌 잔기는 BFDV 및 Hep E 캡시드 단백질의 알려진 ssDNA 접촉 잔기와 정렬된다 (도 3, 별표로 표시됨).

[0754] 서열

[0755] 아래 열거된 서열은 다음과 같이 주석 처리된다. 볼드체 및 밑줄이 그어진 텍스트는 킬레이팅 정제에 사용되는 His6 태그(HHHHHH) 및 예를 들어, 저-발현 단백질의 웨스턴 블롯 검출에 사용되는 강한 에피토프인 플래그 태그(DYKDDDDK)를 포함하는 서열을 나타낸다. 볼드체 및 이탤릭체 서열은 고리9 ORF1 서열 또는 그의 일부를 나타낸다. 볼드체가 아니고 밑줄이 그어지지 않은 서열은 고리2 서열 또는 그의 일부이다. 볼드체가 아닌 밑줄이 그어진 서열은 부리 및 깃털 질병 바이러스(BFDV)로부터 유래된 것이다. 회색 강조 표시는 예를 들어, 아르기닌-풍부 영역과 고리 9 ORF1의 제1 베타 가닥에서 라이신에서 히스티딘으로의 돌연변이의 위치를 나타낸다.

[0756] SEQ ID NO: 561:

[0757] 고리 2 N-말단 HIS-FLAG-3C프로테아제-ORF1:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFGQPSG**MPYYRRRRRYNYRRPRWYGRGWIRRPFRRRFRKRVRPTYT  
TIPLKQWQPPYKRTCYIKGQDCLIIYSNLRGLMNSTMYEKSIVPVHWPGGGSFVSMLTLDALYDIHKLCRNWWTSTN  
QDLPLVRYKGCKITFYQSTFTDYVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIIPSRQTRRKKKPYTKIFVKPPPQFE  
NKWYFATDLYKIPLLQIHCTACNLQNPVFKPKDKLSNNVTLWLSLNTISIQNRNMSVDQGGQSWPFKILGTQSFYFYFTGA  
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKKNFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT  
TENMKWNLNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQYLKLLSNATGTGWDPGPELILEGFPLWLIYWGAFD  
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
PNIQQQLSDNISMFKYFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPESILYSFDYRHHGNYTTTALSRI  
SQDWALKDTSKITEPDRQQLLKQALECLQISEETQEKEKEVQQLISNLRQQQQLYRERIISLLKDDQ

[0758]

[0759] SEQ ID NO: 562:

[0760] 고리 9 N-말단 HIS-FLAG-3C프로테아제-ORF1:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFGQPSG**MPYYWRQKYRRRYPFSWRTRRIIQRKRWRVYRKPRKTY  
**WRRKLRVRRKRYKRLKIVLKQFQPKIIRRCTIFGTICLFGQSPERANNNYIQTIYSYVDPKEPGGGWTLITESLSSL**  
**WEDWEHLKNVWTSNAGLPLVRYGGVTLFYFQSAITDYIAQVFNCPMTDTKYTHADSAPNRMLLKHHVIRVPS**  
**RETRKRRKPYKRVVGGPPSQMKNKWFQRDICEIPLIMIAATAVDFRYPFCASDCASNLTCLNPLLFQNDQDFDH**  
**PSDTQGYFPKPGVLYSTQRSNKSSSDCIYLGNTKDNQEGKSASSMLTKTKITDWGNPFWHYYIDGSKKIFSYFK**  
**PPSQLDSSDFEHMTELAEPMFIVRYNPERDTGQGNLIYVTENFRGQHWDPSSDNLKLDGFPLYDMCWGFIDWI**

[0761]

**EKVHETENLLTNYCFIRSSAFNEKTKVFIPVDHSLTGFSPYETPVKSSDAQHWHPQIRFQTKSINDICLTGPGCARSP**  
**YGNMQAKMSYKHFVKGWGGCPKTYEKPYDPCSQPNWITPHNLNETIQIQNPNTCPQTELEQWDRRDIVTKKAI**  
**ERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQQLNKLKHKHQHLKQLKQYLKPNQIE**

[0762]

[0763] SEQ ID NO: 563:

[0764] 고리 9의 ARG 아암을 갖는 고리 2 ORF1(고리 291)

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFGQPSG**MPYYWRQKYRRRYPFSWRTRRIIQRKRWRVYRKPRKTY  
**WRRKLRVRRRRRPTTYTIPLKQWQPPYKRTCYIKGQDCLIIYSNLRGLMNSTMYEKSIVPVHWPGGGSFVSMLTLD**  
**ALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIIPSRQT**  
**RRKKKPYTKIFVKPPPQFENKWFATDLYKIPLLQIHCTACNLQNPVFKPKDKLSNNVTLWLSLNTISIQNRNMSVDQGGQ**  
**WPFKILGTQSFYFYFTGANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKKNFTNYWGNPFNKHIQE**  
**HLDMILYSLKSPEAIKNEWTTENMKWNLNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQYLKLLSNATGTGWDPG**  
**IPELILEGFPLWLIYWGAFDQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYP**  
**QVQYQLEAQNKLLQTGPFTPNIQQQLSDNISMFKYFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLSQSPGEAPE**  
**SILYSFDYRHHGNYTTTALSRIISLLKDDQ**

[0765]

[0766] SEQ ID NO: 564:

[0767] 고리 9의 ARG 아암 및 베타 가닥 1 + 2722 에피토프를 갖는 고리 2 ORF1 (고리 292)

**MGSSHHHHHGS**DKDDDDKSGSLEVL**FQGP**SGMPPYWRQKYRRRYRPF**SWRTRRII**QRRKRWRYR**KPRKTY**  
WRRKLRVRKRFYKRLKKIVL**KQFQPKI**IRRCTIFGT**ICLFQGS**NLRLGMNSTMYEKSIVPVHWPGGGFSVSM**LTDLA**  
LYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTD**YVRIHTELPANS**NKLTYPNTHPLMMMSKYKH**IIPSRQTR**  
RKKKPYTKIFVPPPQFENKWYFATDLYKIPLLQ**IHCTACNLQ**NPVFKPKDKLSNNVTLW**SLNTISIQNRN**MSVDQGGQ**SW**  
PFKILGTQSFYFYTGANLPGD**TTQIPVADLL**PLTNPRINRPGQ**SLNEAKITD**HITFTEYKNKFTNYWGNPFNK**HIQEHL**  
DMILYSLKSPEAIKNEWTTENMKWNQL**NNAGTMAL**TPFNEIFTQIQYNPDRDTG**EDTQLYLLS**NATGTGWDP**PGIP**  
ELILEGFPLWLIYWGFAD**FQKNLKKVTN**IDTNYMLVAKTKFTQKPGTFYLVILNDTFV**EGNSPYEKQ**PLPEDNIKWY**PQV**  
QYQLEAQNKLLQ**TGPFTPN**IQQGLSDNISMFYKFKWGGSP**PKAINVEN**PAHQIQYPIRNEHETT**SLQSPGEA**ESIL  
YSFDYR**HGN**YTTALSRISQDWALKDTVSKIT**EPDRQ**LLKQALECLQISEETQEKKEKEV**QQLIS**NRQQQ**LYRERI**SL  
LK**DQ**

[0768] SEQ ID NO: 565:  
[0769] ARG 아암 및 제1 베타 가닥에 LYS/HIS 돌연변이를 갖는 고리 9

**MGSSHHHHHGS**DKDDDDKSGSLEVL**FQGP**SGMPPYWRQKYRRRYRPF**SWRTRRII**QRRKRWRYR**KPRKTY**  
WRRHLRVRHRFYRHLHHI**VLKQFQPKI**IRRCTIFGT**ICLFQGS**PERANNNYI**QTIYSY**VPDKE**PGGG**WTLIT**ESLSS**  
WEDWEHLKNV**WTQSNAGL**PLVRYGGV**TLYFYQ**SAYTD**YIAQV**FN**CYPMTD**TKYTHADSAP**NRMLL**KKHVIR**VPS**  
RETRKKR**KPKYKRV**RGPPSQ**MQNKWY**QRDICEI**PLIMIAA**TA**VD**FRY**PF**CASDCAS**NNLT**LTCL**NPL**L**FQ**NQ**DFD**H  
**PSDTQGY**FPKPGV**YLYST**QRSNK**PSSDCI**YLGNTKDNQ**EGKS**ASS**MLTK**TQKITDWGN**PFWHY**YIDG**SKKIF**SYFK  
PP**QLDSS**DFEHMTE**LAEP**MFQVRYN**PERDTG**QGNLIY**TENFR**GQHWD**PPSSD**NL**KLDG**FPLYDMC**WGF**IDW**I**  
EK**VHETEN**LLTNY**CFIR**SSAFNEK**TVFIP**VDH**SFLT**GFSPY**ETPV**KSSDQA**HWH**PQIR**FQ**TKS**INDI**CLT**GP**GCAR**SP**  
Y**GN**Y**MQA**KMSYK**FHV**KWGG**CPKTYE**KPYD**PCS**QPN**WTIP**HN**NETI**QIQNP**NTCP**QTELQ**EW**D**WRR**DIV**TK**KA**I**  
ER**IR**QHTE**PHET**LQ**ISTG**SKHN**PPV**HR**QTSP**WTD**SETD**SE**E**E**KD**Q**TQ**EIQI**Q**LN**KLR**KK**HQ**HL**KQ**L**KQ**YL**KP**Q**NI**E

[0772] SEQ ID NO: 566:

[0774] BFDV의 ARG 아암을 갖는 고리 9:

**MGSSHHHHHGS**DKDDDDKSGSLEVL**FQGP**SGM**WGT**SN**CA**CA**KFQ**IRRRYAR**PY**RRR**IR**RYRRR**RFR**  
RR**FT**TNR**RR**KRFYKRLKKIVL**KQFQPKI**IRRCTIFGT**ICLFQGS**PERANNNYI**QTIYSY**VPDKE**PGGG**WTLIT**ESLSS**  
WEDWEHLKNV**WTQSNAGL**PLVRYGGV**TLYFYQ**SAYTD**YIAQV**FN**CYPMTD**TKYTHADSAP**NRMLL**KKHVIR**VPS**  
RETRKKR**KPKYKRV**RGPPSQ**MQNKWY**QRDICEI**PLIMIAA**TA**VD**FRY**PF**CASDCAS**NNLT**LTCL**NPL**L**FQ**NQ**DFD**H  
**PSDTQGY**FPKPGV**YLYST**QRSNK**PSSDCI**YLGNTKDNQ**EGKS**ASS**MLTK**TQKITDWGN**PFWHY**YIDG**SKKIF**SYFK  
PP**QLDSS**DFEHMTE**LAEP**MFQVRYN**PERDTG**QGNLIY**TENFR**GQHWD**PPSSD**NL**KLDG**FPLYDMC**WGF**IDW**I**  
EK**VHETEN**LLTNY**CFIR**SSAFNEK**TVFIP**VDH**SFLT**GFSPY**ETPV**KSSDQA**HWH**PQIR**FQ**TKS**INDI**CLT**GP**GCAR**SP**  
Y**GN**Y**MQA**KMSYK**FHV**KWGG**CPKTYE**KPYD**PCS**QPN**WTIP**HN**NETI**QIQNP**NTCP**QTELQ**EW**D**WRR**DIV**TK**KA**I**  
ER**IR**QHTE**PHET**LQ**ISTG**SKHN**PPV**HR**QTSP**WTD**SETD**SE**E**E**KD**Q**TQ**EIQI**Q**LN**KLR**KK**HQ**HL**KQ**L**KQ**YL**KP**Q**NI**E

[0776] SEQ ID NO: 567:

[0777] BFDV 캡시드 단백질의 베타 가닥 F 및 G를 갖는 고리 9:

**MGSSHHHHHGS**DKDDDDKSGSLEVL**FQGP**SGMPPYWRQKYRRRYRPF**SWRTRRII**QRRKRWRYR**KPRKTY**  
WRRKLRVRKRFYKRLKKIVL**KQFQPKI**IRRCTIFGT**ICLFQGS**PERANNNYI**QTIYSY**VPDKE**PGGG**WTLIT**ESLSS**  
WEDWEHLKNV**WTQSNAGL**PLVRYGGV**TLYFYQ**SAYTD**YIAQV**FN**CYPMTD**TKYTHADSAP**NRMLL**KKHAK**KW**FS  
RETRKKR**KPGFKR**LL**GP**PSQ**MQNKWY**QRDICEI**PLIMIAA**TA**VD**FRY**PF**CASDCAS**NNLT**LTCL**NPL**L**FQ**NQ**DFD**H  
**SDTQGY**FPKPGV**YLYST**QRSNK**PSSDCI**YLGNTKDNQ**EGKS**ASS**MLTK**TQKITDWGN**PFWHY**YIDG**SKKIF**SYFK  
PS**QLDSS**DFEHMTE**LAEP**MFQVRYN**PERDTG**QGNLIY**TENFR**GQHWD**PPSSD**NL**KLDG**FPLYDMC**WGF**IDW**I**  
KV**HETEN**LLTNY**CFIR**SSAFNEK**TVFIP**VDH**SFLT**GFSPY**ETPV**KSSDQA**HWH**PQIR**FQ**TKS**INDI**CLT**GP**GCAR**SPY**  
G**NY**Y**MQA**KMSYK**FHV**KWGG**CPKTYE**KPYD**PCS**QPN**WTIP**HN**NETI**QIQNP**NTCP**QTELQ**EW**D**WRR**DIV**TK**KA**I**  
R**IR**QHTE**PHET**LQ**ISTG**SKHN**PPV**HR**QTSP**WTD**SETD**SE**E**E**KD**Q**TQ**EIQI**Q**LN**KLR**KK**HQ**HL**KQ**L**KQ**YL**KP**Q**NI**E

[0779] SEQ ID NO: 568:

[0780] 고리 9의 베타 C를 갖는 고리 2:

**MGSSHHHHHGS**DKDDDDKSGSLEVL**FQGP**SGMPPYRRRYRNYRRPRWYGRGWIRRPFRRRFRKRRVRPT**YT**  
TIPLKQWQPPYK**RCTIFGTICLFQGS**NLRLGMNSTMYEKSIVPVHWPGGGFSVSM**LTDLA**LYDIHKLCRNWWTST**N**  
QDLPLVRYKGCKITFYQSTFTD**YVRIHTELPANS**NKLTYPNTHPLMMMSKYKH**IIPSRQTR**RKKKPYTKIFV**KPP**QFE  
NKWYFATDLYKIPLLQ**IHCTACNLQ**NPVFKPKDKLSNNVTLW**SLNTISIQNRN**MSVDQGGQ**SW**PFKILGTQSFYFYTG**A**  
NLPGD**TTQIPVADLL**PLTNPRINRPGQ**SLNEAKITD**HITFTEYKNKFTNYWGNPFNK**HIQEHL**DMILYSLKSPEAIKNEW**T**  
TENMKWNQL**NNAGTMAL**TPFNEIFTQIQYNPDRDTG**EDTQLYLLS**NATGTGWDP**PGIP**ELILEGFPLWLIYWGFAD

[0781] TENMKWNQLNNAGTMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGFAD

FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
PNIQQQLSDNISMIFYKFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI  
SQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQQLYRERIISLLKDQ

[0782]

SEQ ID NO: 569:

[0783]

고리 9의 링커 1을 갖는 고리 2:

[0784]

**MGSSHHHHHGS**DYKDDDDKSGS**LEVL**FQ**GP**SGMPYYYRRRRYNYRRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
TIPLKQWQPPYKRTCYIKGQDCLIIYS**PERANN**YIQTIYS**VPDKEP**GGGG**WTLITESL**SS**LWEDWEHLK**NV**WT**QSN  
**AGL**PLVRYKGCKITFYQSTFTDYIVRIHELTPANSNKLTYPNTHPLMMMMSKYKHIIIPSRQTRRKKKPYTKIFVKPPPQFE  
NKWYFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNVTLWSLNTISIQNRNMSVDQGGQSWPFKILGTQSFYFYFTGA  
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT  
TENMKWNQLNNAAGTMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGFAD  
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
PNIQQQLSDNISMIFYKFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI  
SQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQQLYRERIISLLKDQ

[0785]

SEQ ID NO: 570:

[0786]

고리 9의 베타 가닥 D를 갖는 고리 2:

[0787]

**MGSSHHHHHGS**DYKDDDDKSGS**LEVL**FQ**GP**SGMPYYYRRRRYNYRRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR**LGM**NSTMYEKSIVPVHWPGGGSFSVSM**LT**LDALYDIH**KLC**RN**W**WTSTN  
QD**LP**L**VRYGGV**TL**YFY**QSTFTDYIVRIHELTPANSNKLTYPNTHPLMMMMSKYKHIIIPSRQTRRKKKPYTKIFVKPPPQF  
ENKWFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNVTLWSLNTISIQNRNMSVDQGGQSWPFKILGTQSFYFYFTG  
ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE  
WTENMKWNQLNNAAGTMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGF  
ADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG  
PFTPN**IQ**QLSDNISMIFYKFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI  
SRISQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQQLYRERIISLLKDQ

[0788]

SEQ ID NO: 571:

[0789]

고리 9의 링커 2를 갖는 고리 2:

[0790]

**MGSSHHHHHGS**DYKDDDDKSGS**LEVL**FQ**GP**SGMPYYYRRRRYNYRRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR**LGM**NSTMYEKSIVPVHWPGGGSFSVSM**LT**LDALYDIH**KLC**RN**W**WTSTN  
QD**LP**L**VRY**KGCKITFYQSTFTDYIVRIH**NCYP**MTDT**KYTHADS**AP**NR**M**LLK**KHIIIPSRQTRRKKKPYTKIFVKPPPQF  
NKWYFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNVTLWSLNTISIQNRNMSVDQGGQSWPFKILGTQSFYFYFTGA  
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT  
TENMKWNQLNNAAGTMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGFAD  
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT  
PNIQQQLSDNISMIFYKFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI  
SQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQQLYRERIISLLKDQ

[0791]

SEQ ID NO: 572:

[0792]

고리 9의 베타 가닥 G DNA 결합을 갖는 고리 2:

[0793]

**MGSSHHHHHGS**DYKDDDDKSGS**LEVL**FQ**GP**SGMPYYYRRRRYNYRRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR**LGM**NSTMYEKSIVPVHWPGGGSFSVSM**LT**LDALYDIH**KLC**RN**W**WTSTN  
QD**LP**L**VRY**KGCKITFYQSTFTDYIVRIHELTPANSNKLTYPNTHPLMMMMSKYKHIIIPSRQTRRKKK**PKRVRV**KPPPQF  
ENKWFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNVTLWSLNTISIQNRNMSVDQGGQSWPFKILGTQSFYFYFTG  
ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE  
WTENMKWNQLNNAAGTMALTPFNEIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPIPELILEGFPLWLIYWGF  
ADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG  
PFTPN**IQ**QLSDNISMIFYKFKWGGSPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI  
SRISQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQQLYRERIISLLKDQ

[0795]

SEQ ID NO: 573:

[0796]

[0797] 고리 9의 베타 가닥 F 인터프로토머 접촉부를 갖는 고리 2:

**MGSSHHHHHGS DYKDDDDKSGSLEVL FQGPSG**MPYYRRRRRYNRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR LGMNSTMYEKSIVPVHWPGGGSFVSMLTLDALYDIHKLCRNWWTSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMMSK**HVIRVPS**RQTRRKKKPYTKIFVKPPPQF  
 ENKWFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNVTLWLSLNTISIQNRNMSVDQGSWPFKILGTQSFYFYFTG  
 ANLPGDTTQIPVADLLPTNPRINRPGQSLNEAKITDHITFEYKKNFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE  
 WTTENMKWNQLNNAAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPGPELILEGFPLWLIYWG  
 ADFQKNLKVNTIDNYMLVAKTKFTQKPGTFYLVLNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLQGTG  
 PFTPNIQGLSDNISMFYKFFYKGGSPKAINVENPAHQIQYPIPRNEHETTSLOSPGEAPESILYSFDYRHGNYTTTAL  
 SRISQDWALKDVTSKITEPDRQQLLKQALECLQISEETQEKKKEVQQLISNLRQQQLYRERIISLLKQD

[0798]

SEQ ID NO: 574:

[0799]

[0800] 고리 9의 C-말단 단편을 갖는 가닥 H 및 I를 갖는 고리 2 ORF1:

**MGSSHHHHHGS DYKDDDDKSGSLEVL FQGPSG**MPYYRRRRRYNRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR LGMNSTMYEKSIVPVHWPGGGSFVSMLTLDALYDIHKLCRNWWTSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE  
 NKWYFATDLYKIPLIMIAATAVDFRYPFCASDCASNLTCLNPLLFQNFQDFDHPSTQGYFPKPGVYLYSTQRSNK  
**PSSSDCIYLGNTKDNQEGKSASSMLTKTKITDWGNPFWHYYIDGSKKIFSYFKPPS QLDSSDFEHMTELAEP MFIQ**  
**VRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFIRSSAFNEK**  
**KTVFIPVDHSLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNMQAKMSYKFHVKWGGCP**  
**KTYEKPYPDCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDRRDIIVTKKAIERIRQHTEPHETLQISTGSKHNPPV**  
**HRQTSPWTDSETDSEEEKDQTQEIQIQLNKLKHKHQQHLKQQLKQYLKPNIE**

[0801]

SEQ ID NO: 575:

[0802]

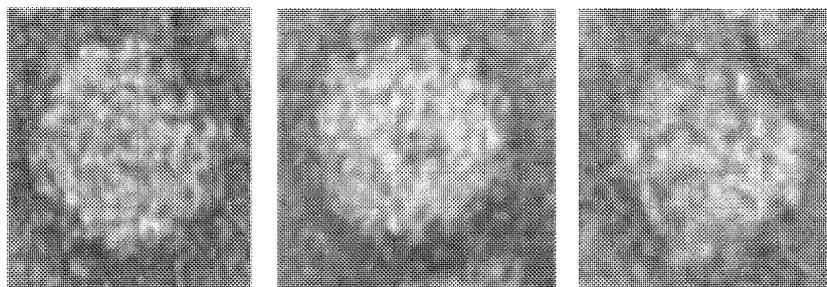
[0803] 고리 9의 가닥 I 및 C-말단 단편을 갖는 고리 2 ORF1:

**MGSSHHHHHGS DYKDDDDKSGSLEVL FQGPSG**MPYYRRRRRYNRRPRWYGRGWIRRPFRFRFRKRVRPTYT  
 TIPLKQWQPPYKRTCYIKGQDCLIIYSNLR LGMNSTMYEKSIVPVHWPGGGSFVSMLTLDALYDIHKLCRNWWTSTN  
 QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE  
 NKWYFATDLYKIPLLQIHCTACNLQNPVFKPKDLSNNLTCLNPLLFQNFQDFDHPSTQGYFPKPGVYLYSTQRSNK  
**SSSDCIYLGNTKDNQEGKSASSMLTKTKITDWGNPFWHYYIDGSKKIFSYFKPPS QLDSSDFEHMTELAEP MFIQV**  
**RYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFIRSSAFNEK**  
**TVFIPVDHSLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNMQAKMSYKFHVKWGGCPK**  
**TYEKPYPDCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDRRDIIVTKKAIERIRQHTEPHETLQISTGSKHNPPV**  
**HRQTSPWTDSETDSEEEKDQTQEIQIQLNKLKHKHQQHLKQQLKQYLKPNIE**

[0804]

도면

도면1

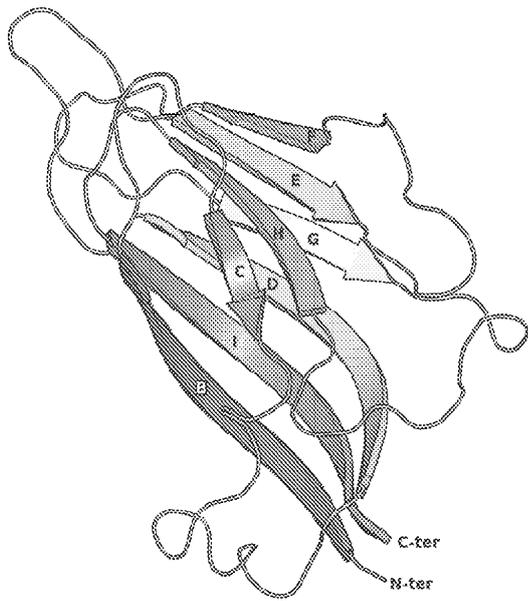


(A) 고리 2 ORF1

(B) 고리 10 ORF1

(C) CAV VP1

도면2

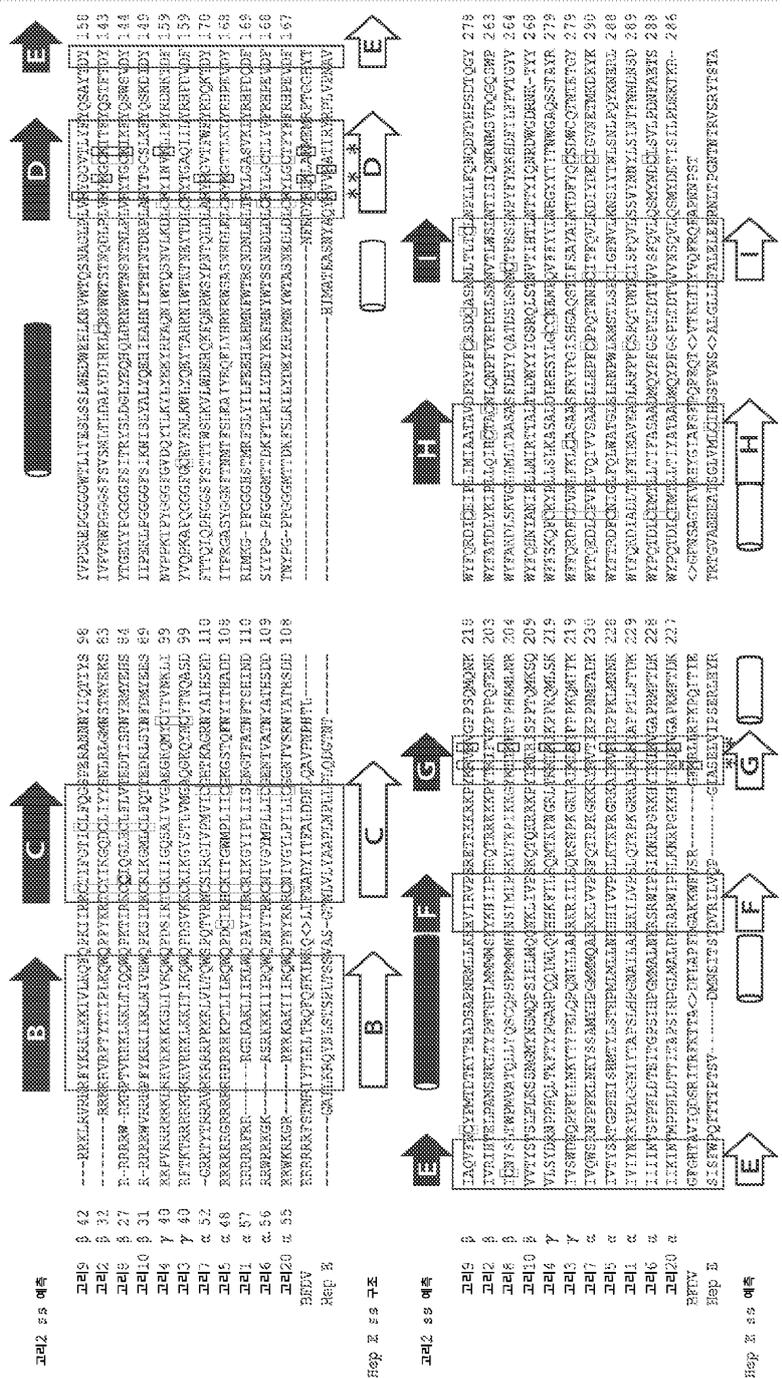


**1. C - H - E - F**

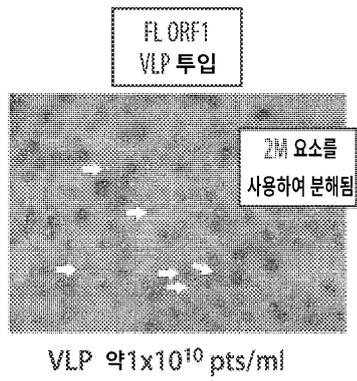
**2. B - I - D - G**

1차  
DNA 접촉물

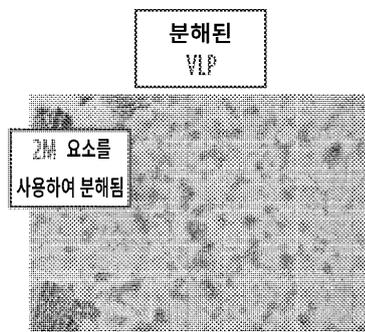
도면3



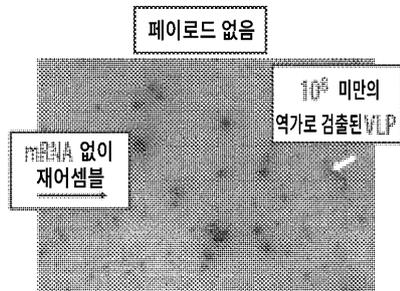
도면4a



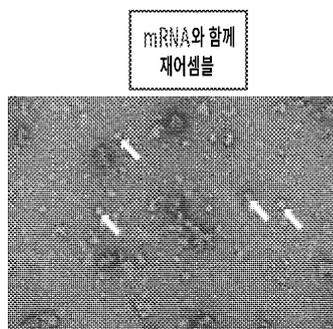
도면4b



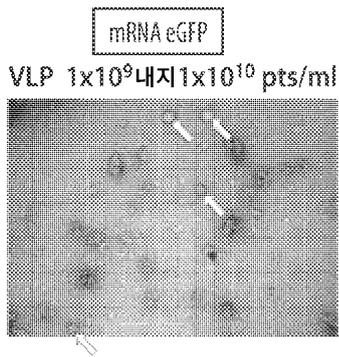
도면4c



도면4d



도면4e



서열 목록

SEQUENCE LISTING

<110> FLAGSHIP PIONEERING INNOVATIONS V, INC.

<120> IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA

<130> V2057-7017WO

<140><141><150> 63/147,064

<151> 2021-02-08

<150> 63/130,360

<151> 2020-12-23

<160> 968

<170> PatentIn version 3.5

<210> 1

<211> 699

<212> PRT

<213> Artificial Sequence

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

<400> 1

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1 5 10 15

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

20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro

35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe

50                      55                      60  
 Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys  
 65                      70                      75                      80  
  
 Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp  
    85                      90                      95  
 Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met  
    100                      105                      110  
 Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe  
    115                      120                      125  
 Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu  
    130                      135                      140  
  
 Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg  
 145                      150                      155                      160  
 Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr  
    165                      170                      175  
 Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr  
    180                      185                      190  
 Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His  
    195                      200                      205  
  
 Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys  
    210                      215                      220  
 Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala  
 225                      230                      235                      240  
 Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys  
    245                      250                      255  
 Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val  
    260                      265                      270  
  
 Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
    275                      280                      285  
 Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
    290                      295                      300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305                      310                      315                      320  
 Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
                                  325                      330                      335  
  
 Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr  
                                  340                      345                      350  
 Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
                                  355                      360                      365  
 Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
                                  370                      375                      380  
 Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
 385                      390                      395                      400  
  
 Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
                                  405                      410                      415  
 Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
                                  420                      425                      430  
 Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
                                  435                      440                      445  
 Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
                                  450                      455                      460  
  
 Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr  
 465                      470                      475                      480  
 Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln  
                                  485                      490                      495  
 Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu  
                                  500                      505                      510  
 Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
                                  515                      520                      525  
  
 Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
                                  530                      535                      540  
 Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn

545                      550                      555                      560  
Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
                                 565                      570                      575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
                                 580                      585                      590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
                                 595                      600                      605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
                                 610                      615                      620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
625                      630                      635                      640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
                                 645                      650                      655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
                                 660                      665                      670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
                                 675                      680                      685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
                                 690                      695

<210> 2

<211> 688

<212> PRT

<213> Artificial Sequence

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

<400> 2

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1                      5                      10                      15

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

20                      25                      30

Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg

35                      40                      45

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

290                      295                      300  
 Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr  
 305                      310                      315                      320  
 Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu  
  
                                  325                      330                      335  
 Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met  
                                  340                      345                      350  
 Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His  
                                  355                      360                      365  
 Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro  
 370                      375                      380  
 Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu  
  
 385                      390                      395                      400  
 Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln  
                                  405                      410                      415  
 Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp  
                                  420                      425                      430  
 Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp  
                                  435                      440                      445  
 Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu  
  
 450                      455                      460  
 Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn  
 465                      470                      475                      480  
 Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly  
                                  485                      490                      495  
 Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp  
                                  500                      505                      510  
 His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu  
  
                                  515                      520                      525  
 Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala  
 530                      535                      540

Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr  
 545                      550                      555                      560  
 Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro  
                                  565                      570                      575  
 His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro  
                                  580                      585                      590  
 Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys  
                                  595                      600                      605  
 Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu  
                                  610                      615                      620  
 Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr  
 625                      630                      635                      640  
 Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln  
                                  645                      650                      655  
 Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln  
                                  660                      665                      670  
 His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu  
                                  675                      680                      685  
 <210> 3  
 <211> 714  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
                                  polypeptide  
 <400> 3  
 Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp  
  
 1                      5                      10                      15  
 Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser  
                                  20                      25                      30  
 Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg  
                                  35                      40                      45  
 Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp

50                      55                      60  
 Arg Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val  
  
 65                      70                      75                      80  
 Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln  
                                  85                      90                      95  
 Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys  
                                  100                      105                      110  
 Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr  
                                  115                      120                      125  
 Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser  
  
 130                      135                      140  
 Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys  
 145                      150                      155                      160  
 Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr  
                                  165                      170                      175  
 Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile  
                                  180                      185                      190  
 Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr  
  
 195                      200                      205  
 Pro Asn Thr His Pro Leu Met Met Met Ser Lys Tyr Lys His Ile  
                                  210                      215                      220  
 Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile  
 225                      230                      235                      240  
 Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr  
                                  245                      250                      255  
 Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn  
  
 260                      265                      270  
 Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr  
                                  275                      280                      285  
 Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val  
                                  290                      295                      300

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe  
 305                      310                      315                      320  
 Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln  
    325                      330                      335  
 Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg  
    340                      345                      350  
 Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe  
    355                      360                      365  
 Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn  
    370                      375                      380  
 Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser  
  
 385                      390                      395                      400  
 Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn  
    405                      410                      415  
 Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro  
    420                      425                      430  
 Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp  
    435                      440                      445  
 Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro  
  
    450                      455                      460  
 Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile  
 465                      470                      475                      480  
 Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn  
    485                      490                      495  
 Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys  
    500                      505                      510  
 Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly  
  
    515                      520                      525  
 Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp  
    530                      535                      540  
 Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln

545                    550                    555                    560  
 Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile  
                          565                    570                    575  
 Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys  
  
                          580                    585                    590  
 Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro  
                          595                    600                    605  
 Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro  
                          610                    615                    620  
 Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr  
 625                    630                    635                    640  
 Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val  
  
                          645                    650                    655  
 Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu  
                          660                    665                    670  
 Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu  
                          675                    680                    685  
 Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg  
                          690                    695                    700  
 Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 705                    710

<210> 4

<211> 714

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 4

Met Gly Ser Ser His His His His His Gly Ser Asp Tyr Lys Asp  
 1                    5                    10                    15  
 Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser  
                          20                    25                    30

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

275                                    280                                    285  
 Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val  
 290                                    295                                    300  
  
 Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe  
 305                                    310                                    315                                    320  
 Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln  
 325                                    330                                    335  
 Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg  
 340                                    345                                    350  
 Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe  
 355                                    360                                    365  
  
 Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn  
 370                                    375                                    380  
 Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser  
 385                                    390                                    395                                    400  
 Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn  
 405                                    410                                    415  
 Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro  
 420                                    425                                    430  
  
 Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp  
 435                                    440                                    445  
 Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro  
 450                                    455                                    460  
 Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile  
 465                                    470                                    475                                    480  
 Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn  
 485                                    490                                    495  
  
 Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys  
 500                                    505                                    510  
 Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly  
 515                                    520                                    525

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp  
 530 535 540  
 Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln  
 545 550 555 560  
  
 Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile  
 565 570 575  
 Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys  
 580 585 590  
 Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro  
 595 600 605  
 Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro  
 610 615 620  
  
 Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr  
 625 630 635 640  
 Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val  
 645 650 655  
 Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu  
 660 665 670  
 Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu  
 675 680 685  
  
 Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg  
 690 695 700  
 Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 705 710  
 <210> 5  
 <211> 688  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polypeptide  
 <400> 5  
 Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1                    5                    10                    15  
 Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser  
  
                   20                    25                    30  
 Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg  
                   35                    40                    45  
 Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg His Arg Trp  
                   50                    55                    60  
 Arg Tyr Arg Lys Pro Arg His Thr Tyr Trp Arg Arg His Leu Arg Val  
 65                    70                    75                    80  
 Arg His Arg Phe Tyr His Arg His Leu His His Ile Val Leu Lys Gln  
  
                   85                    90                    95  
 Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys  
                   100                    105                    110  
 Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr  
                   115                    120                    125  
 Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr  
                   130                    135                    140  
 Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu  
  
 145                    150                    155                    160  
 Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr  
                   165                    170                    175  
 Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile  
                   180                    185                    190  
 Ala Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His  
                   195                    200                    205  
 Ala Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg  
  
                   210                    215                    220  
 Val Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val  
 225                    230                    235                    240  
 Arg Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg  
                   245                    250                    255

Asp Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp  
 260 265 270  
 Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr  
 275 280 285  
 Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His  
 290 295 300  
 Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr  
 305 310 315 320  
 Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu  
 325 330 335  
 Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met  
 340 345 350  
 Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His  
 355 360 365  
 Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro  
 370 375 380  
 Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu  
 385 390 395 400  
 Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln  
 405 410 415  
 Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp  
 420 425 430  
 Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp  
 435 440 445  
 Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu  
 450 455 460  
 Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn  
 465 470 475 480  
 Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly  
 485 490 495  
 Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp

500 505 510  
 His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu

515 520 525  
 Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala

530 535 540  
 Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr

545 550 555 560  
 Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro

565 570 575  
 His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro

580 585 590  
 Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys

595 600 605  
 Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu

610 615 620  
 Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr

625 630 635 640  
 Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln

645 650 655  
 Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln

660 665 670  
 His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu

675 680 685

<210> 6

<211> 687

<212> PRT

<213> Artificial Sequence

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

<400> 6

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp  
 1 5 10 15

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

260                              265                              270  
 Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr Leu  
  
 275                              280                              285  
 Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His Pro  
 290                              295                              300  
 Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr Ser  
 305                              310                              315                              320  
 Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu Gly  
 325                              330                              335  
 Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met Thr  
  
 340                              345                              350  
 Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His Tyr  
 355                              360                              365  
 Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro Ser  
 370                              375                              380  
 Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu Pro  
 385                              390                              395                              400  
 Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln Gly  
  
 405                              410                              415  
 Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp Pro  
 420                              425                              430  
 Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp Met  
 435                              440                              445  
 Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu Asn  
 450                              455                              460  
 Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn Glu  
  
 465                              470                              475                              480  
 Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly Phe  
 485                              490                              495  
 Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp His  
 500                              505                              510

Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu Thr  
 515 520 525

Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala Lys  
 530 535 540

Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr Tyr  
 545 550 555 560

Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro His  
 565 570 575

Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro Gln  
 580 585 590

Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys Lys  
 595 600 605

Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu Gln  
 610 615 620

Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr Ser  
 625 630 635 640

Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln Thr  
 645 650 655

Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln His  
 660 665 670

Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu  
 675 680 685

<210> 7  
 <211> 688  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polypeptide  
 <400> 7

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp  
 1 5 10 15

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



His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu  
 515 520 525  
 Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala  
 530 535 540  
 Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr  
 545 550 555 560  
 Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro  
 565 570 575  
 His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro  
 580 585 590  
 Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys  
 595 600 605  
 Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu  
 610 615 620  
 Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr  
 625 630 635 640  
 Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln  
 645 650 655  
 Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln  
 660 665 670  
 His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu  
 675 680 685  
 <210> 8  
 <211> 699  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polypeptide  
 <400> 8  
 Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp  
 1 5 10 15

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

260 265 270  
 Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
  
 275 280 285  
 Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
 290 295 300  
 Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305 310 315 320  
 Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
 325 330 335  
 Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr  
  
 340 345 350  
 Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
 355 360 365  
 Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
 370 375 380  
 Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
 385 390 395 400  
 Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
  
 405 410 415  
 Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
 420 425 430  
 Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
 435 440 445  
 Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
 450 455 460  
 Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr  
  
 465 470 475 480  
 Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln  
 485 490 495  
 Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu  
 500 505 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
 515 520 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
 530 535 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn  
 545 550 555 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
 565 570 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
 580 585 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
 595 600 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
 610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
 625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
 645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
 660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
 675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 690 695

<210> 9  
 <211> 699  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polypeptide  
 <400> 9  
 Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1                    5                    10                    15  
  
 Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser  
                   20                    25                    30  
 Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro  
                   35                    40                    45  
 Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe  
                   50                    55                    60  
 Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys  
 65                    70                    75                    80  
  
 Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp  
                   85                    90                    95  
 Cys Leu Ile Tyr Tyr Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln  
                   100                    105                    110  
 Thr Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp  
                   115                    120                    125  
 Thr Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His  
                   130                    135                    140  
  
 Leu Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg  
 145                    150                    155                    160  
 Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr  
                   165                    170                    175  
 Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr  
                   180                    185                    190  
 Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His  
                   195                    200                    205  
  
 Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys  
                   210                    215                    220  
 Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala  
 225                    230                    235                    240  
 Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys  
                   245                    250                    255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val  
 260 265 270  
  
 Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
 275 280 285  
 Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
 290 295 300  
 Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305 310 315 320  
 Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
 325 330 335  
  
 Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr  
 340 345 350  
 Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
 355 360 365  
 Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
 370 375 380  
 Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
 385 390 395 400  
  
 Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
 405 410 415  
 Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
 420 425 430  
 Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
 435 440 445  
 Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
 450 455 460  
  
 Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr  
 465 470 475 480  
 Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln  
 485 490 495  
 Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu

500 505 510  
 Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
 515 520 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
 530 535 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn  
 545 550 555 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
 565 570 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
 580 585 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
 595 600 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
 610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
 625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
 645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
 660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
 675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 690 695

<210> 10

<211> 699

<212> PRT

<213> Artificial Sequence

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

<400> 10

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

245                                  250                                  255  
 Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val  
  
 260                                  265                                  270  
 Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
 275                                  280                                  285  
 Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
 290                                  295                                  300  
 Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305                                  310                                  315                                  320  
 Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
  
 325                                  330                                  335  
 Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr  
 340                                  345                                  350  
 Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
 355                                  360                                  365  
 Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
 370                                  375                                  380  
 Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
  
 385                                  390                                  395                                  400  
 Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
 405                                  410                                  415  
 Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
 420                                  425                                  430  
 Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
 435                                  440                                  445  
 Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
  
 450                                  455                                  460  
 Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr  
 465                                  470                                  475                                  480  
 Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln  
 485                                  490                                  495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu  
 500 505 510  
 Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
 515 520 525  
 Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
 530 535 540  
 Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn  
 545 550 555 560  
 Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
 565 570 575  
 Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
 580 585 590  
 Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
 595 600 605  
 Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
 610 615 620  
 Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
 625 630 635 640  
 Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
 645 650 655  
 Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
 660 665 670  
 Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
 675 680 685  
 Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 690 695  
 <210> 11  
 <211> 699  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 11

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

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys  
 65                    70                    75                    80  
 Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp  
                   85                    90                    95  
 Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met  
                   100                    105                    110  
 Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe  
                   115                    120                    125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu  
                   130                    135                    140  
 Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg  
 145                    150                    155                    160  
 Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr  
                   165                    170                    175  
 Ile Val Arg Ile His Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr  
                   180                    185                    190

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

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys  
 245 250 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val  
 260 265 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
 275 280 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
 290 295 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305 310 315 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
 325 330 335

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

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
 355 360 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
 370 375 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
 385 390 395 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
 405 410 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
 420 425 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
 435 440 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
 450 455 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr  
 465 470 475 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln

485 490 495  
 Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu  
 500 505 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
 515 520 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
 530 535 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn  
 545 550 555 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
 565 570 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
 580 585 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
 595 600 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
 610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
 625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
 645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
 660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
 675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 690 695

<210> 12

<211> 699

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 12

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1                    5                    10                    15

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

                  20                    25                    30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro

                  35                    40                    45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe

                  50                    55                    60

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

65                    70                    75                    80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp

                  85                    90                    95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met

                  100                    105                    110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe

                  115                    120                    125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu

                  130                    135                    140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg

145                    150                    155                    160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr

                  165                    170                    175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr

                  180                    185                    190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His

                  195                    200                    205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Lys Arg

                  210                    215                    220

Val Arg Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala



Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln  
 485 490 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu  
 500 505 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys  
 515 520 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu  
 530 535 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn  
 545 550 555 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro  
 565 570 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile  
 580 585 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala  
 595 600 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr  
 610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr  
 625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala  
 645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys  
 660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr  
 675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln  
 690 695

<210> 13

<211> 699

<212> PRT

<213> Artificial Sequence

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

<400> 13

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

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala  
 225                                230                                235                                240  
 Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys  
    245                                250                                255  
 Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val  
    260                                265                                270  
 Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser  
    275                                280                                285  
 Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser  
  
 290                                295                                300  
 Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr  
 305                                310                                315                                320  
 Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn  
    325                                330                                335  
 Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr  
    340                                345                                350  
 Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe  
  
 355                                360                                365  
 Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys  
 370                                375                                380  
 Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp  
 385                                390                                395                                400  
 Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu  
    405                                410                                415  
 Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu  
  
 420                                425                                430  
 Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp  
 435                                440                                445  
 Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu  
 450                                455                                460  
 Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr

465                    470                    475                    480  
 Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln

                         485                    490                    495  
 Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu

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

                         515                    520                    525  
 Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu

                         530                    535                    540  
 Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn

545                    550                    555                    560  
 Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro

                         565                    570                    575  
 Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile

                         580                    585                    590  
 Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala

                         595                    600                    605  
 Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr

                         610                    615                    620  
 Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr

625                    630                    635                    640  
 Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala

                         645                    650                    655  
 Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys

                         660                    665                    670  
 Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr

                         675                    680                    685  
 Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln

                         690                    695  
 <210> 14

<211> 673

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 14

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp  
1                   5                   10                   15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser  
                  20                   25                   30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro  
                  35                   40                   45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe  
                  50                   55                   60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys  
65                   70                   75                   80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp  
                  85                   90                   95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met  
                  100                   105                   110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe  
                  115                   120                   125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu  
                  130                   135                   140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg  
145                   150                   155                   160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr  
                  165                   170                   175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr  
                  180                   185                   190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His  
                  195                   200                   205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys

210	215	220																				
Ile	Phe	Val	Lys	Pro	Pro	Pro	Gln	Phe	Glu	Asn	Lys	Trp	Tyr	Phe	Ala							
225				230					235					240								
Thr	Asp	Leu	Tyr	Lys	Ile	Pro	Leu	Ile	Met	Ile	Ala	Ala	Thr	Ala	Val							
			245					250						255								
Asp	Phe	Arg	Tyr	Pro	Phe	Cys	Ala	Ser	Asp	Cys	Ala	Ser	Asn	Asn	Leu							
		260						265						270								
Thr	Leu	Thr	Cys	Leu	Asn	Pro	Leu	Leu	Phe	Gln	Asn	Gln	Asp	Phe	Asp							
	275						280							285								
His	Pro	Ser	Asp	Thr	Gln	Gly	Tyr	Phe	Pro	Lys	Pro	Gly	Val	Tyr	Leu							
	290				295						300											
Tyr	Ser	Thr	Gln	Arg	Ser	Asn	Lys	Pro	Ser	Ser	Ser	Asp	Cys	Ile	Tyr							
305				310						315				320								
Leu	Gly	Asn	Thr	Lys	Asp	Asn	Gln	Glu	Gly	Lys	Ser	Ala	Ser	Ser	Leu							
			325						330					335								
Met	Thr	Leu	Lys	Thr	Gln	Lys	Ile	Thr	Asp	Trp	Gly	Asn	Pro	Phe	Trp							
		340						345						350								
His	Tyr	Tyr	Ile	Asp	Gly	Ser	Lys	Lys	Ile	Phe	Ser	Tyr	Phe	Lys	Pro							
	355						360						365									
Pro	Ser	Gln	Leu	Asp	Ser	Ser	Asp	Phe	Glu	His	Met	Thr	Glu	Leu	Ala							
	370					375					380											
Glu	Pro	Met	Phe	Ile	Gln	Val	Arg	Tyr	Asn	Pro	Glu	Arg	Asp	Thr	Gly							
385				390							395			400								
Gln	Gly	Asn	Leu	Ile	Tyr	Val	Thr	Glu	Asn	Phe	Arg	Gly	Gln	His	Trp							
			405								410			415								
Asp	Pro	Pro	Ser	Ser	Asp	Asn	Leu	Lys	Leu	Asp	Gly	Phe	Pro	Leu	Tyr							
		420							425					430								
Asp	Met	Cys	Trp	Gly	Phe	Ile	Asp	Trp	Ile	Glu	Lys	Val	His	Glu	Thr							
		435							440					445								
Glu	Asn	Leu	Leu	Thr	Asn	Tyr	Cys	Phe	Cys	Ile	Arg	Ser	Ser	Ala	Phe							
	450						455							460								

Asn Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr  
 465 470 475 480

Gly Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His  
 485 490 495

Trp His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys  
 500 505 510

Leu Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln  
 515 520 525

Ala Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys  
 530 535 540

Thr Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile  
 545 550 555 560

Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys  
 565 570 575

Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr  
 580 585 590

Lys Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr  
 595 600 605

Leu Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln  
 610 615 620

Thr Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp  
 625 630 635 640

Gln Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln  
 645 650 655

Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile  
 660 665 670

Glu

<210> 15

<211> 673

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 15

Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp

1                    5                    10                    15

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

20                    25                    30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro

35                    40                    45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe

50                    55                    60

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

65                    70                    75                    80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp

85                    90                    95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met

100                    105                    110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe

115                    120                    125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu

130                    135                    140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg

145                    150                    155                    160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr

165                    170                    175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr

180                    185                    190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His

195                    200                    205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys

210                    215                    220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala  
 225 230 235 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys  
 245 250 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Leu  
 260 265 270

Thr Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp  
 275 280 285

His Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu  
 290 295 300

Tyr Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr  
 305 310 315 320

Leu Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu  
 325 330 335

Met Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp  
 340 345 350

His Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro  
 355 360 365

Pro Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala  
 370 375 380

Glu Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly  
 385 390 395 400

Gln Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp  
 405 410 415

Asp Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr  
 420 425 430

Asp Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr  
 435 440 445

Glu Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe  
 450 455 460

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

465                                 470                                 475                                 480  
 Gly Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His  
   485                                 490                                 495

Trp His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys  
   500                                 505                                 510

Leu Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln  
   515                                 520                                 525

Ala Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys  
   530                                 535                                 540

Thr Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile  
 545                                 550                                 555                                 560

Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys  
   565                                 570                                 575

Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr  
   580                                 585                                 590

Lys Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr  
   595                                 600                                 605

Leu Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln  
   610                                 615                                 620

Thr Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp  
 625                                 630                                 635                                 640

Gln Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln  
   645                                 650                                 655

Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile  
   660                                 665                                 670

Glu

<210> 16

<211> 3753

<212> DNA

<213> Alphatorquevirus sp.

<400> 16

tgctacgtca ctaaccacag tgtctctac aggccaatcg cagtctatgt cgtgcacttc 60  
  
 ctgggcatgg tctacataat tataataatg cttgcacttc cgaatggctg agtttttct 120  
 gcccgctccg ggagaggagc cacggcaggg gatccgaacg tctgagggc ggggtgccgga 180  
 ggtgagttta cacaccgaag tcaaggggca attcgggctc aggactggcc ggcttttggg 240  
 caaggctctt aaaaatgcac ttttctcgaa taagcagaaa gaaaaggaaa gtgctactgc 300  
 tttgcgtgcc agcagctaag aaaaaaccaa ctgctatgag cttctggaaa cctccggtac 360  
 acaatgtcac ggggatccaa cgcatgtggt atgagtcctt tcaccgtggc cacgcttctt 420  
 tttgtggttg tgggaatcct atacttcaca ttactgcact tgctgaaaca tatggccatc 480  
  
 caacaggccc gagaccttct gggccaccgg gagtagacc caacccccac atccgtagag 540  
 ccaggcctgc cccggccgct ccggagccct cacaggttga ttcgagacca gccttgacat 600  
 ggcatgggga tgggtggaagc gacggaggcg ctggtggttc cggaagcggg ggaccctggg 660  
 cagacttcgc agacgatggc ctcgatcagc tcgtcgccgc cctagacgac gaagagtaag 720  
 gaggcgcaga cgggtggagga gggggagacg aaaaacaagg acttacagac gcaggagacg 780  
 ctttagacgc aggggacgaa aagcaaaact tataataaaa ctgtggcaac ctgcagtaat 840  
 taaaagatgc agaataaagg gatacatacc actgattata agtgggaacg gtacctttgc 900  
  
 cacaaacttt accagtcaaa taaatgacag aataatgaaa ggccccttcg ggggaggaca 960  
 cagcactatg aggttcagcc tctacathtt gtttgaggag cacctcagac acatgaactt 1020  
 ctggaccaga agcaacgata acctagagct aaccagatac ttgggggctt cagtaaaaat 1080  
 atacaggcac ccagaccaag actttatagt aatatacaac agaagaacct ctctaggagg 1140  
 caacatctac acagcacctt ctctacacc aggcaatgcc attttagcaa aacacaaaat 1200  
 attagtacca agtttacaga caagaccaa gggtagaaaa gcaattagac taagaatagc 1260  
 acccccaca ctctttacag acaagtggta ctttcaaaag gacatagccg acctcacctt 1320  
  
 tttcaacatc atggcagttg aggctgactt gcggtttccg ttctgctcac cacaaactga 1380  
 caacacttgc atcagcttcc aggtccttag ttcggtttac aacaactacc tcagtattaa 1440  
 tacctttaat aatgacaact cagactcaaa gttaaaagaa ttttaataa aagcatttcc 1500  
 aacaacaggc acaaaaggaa caagttaaa tgcactaat acatttagaa cagaaggatg 1560  
 cataagtca ccaactaa aaaaaccaa cccacaaata aacaaacct tagagtcaca 1620  
 atactttgca cctttagatg ccctctgggg agacccata tactataatg atctaataga 1680  
 aaacaaaagt ttgaacgata tcattgagaa aataactaata aaaaacatga ttacatacca 1740

tgcaaaacta agagaatttc caaattcata ccaaggaaac aaggcctttt gccacctaac 1800  
 aggcataatac agcccacat acctaacca aggcagaata tctccagaaa tatttggact 1860  
 gtacacagaa ataatttaca acccttacac agacaaagga actggaaaca aagtatggat 1920  
 ggacccacta actaaagaga acaacatata taaagaagga cagagcaaat gcctactgac 1980  
 tgacatgccc ctatggactt tactttttgg atatacagac tgggttaaaa aggacactaa 2040  
 taactgggac ttaccactaa actacagact agtactaata tgccttata cttttccaaa 2100  
 attgtacaat gaaaaagtaa aagactatgg gtacatcccg tactcctaca aattcggagc 2160  
  
 gggtcagatg ccagacggca gcaactacat accctttcag tttagagcaa agtggtagcc 2220  
 cacagtacta caccagcaac aggtaatgga ggacataagc aggagcgggc cttttgcacc 2280  
 taaggtagaa aaaccaagca ctacagctgtt aatgaagtac tgttttaact ttaactgggg 2340  
 cggtaacct atcattgaac agattgttaa agacccagc ttccagccca cctatgaaat 2400  
 acccggtagc ggtaacatcc ctagaagaat acaagtcatc gaccgcgggg tcttgggacc 2460  
 gcactactcg ttccggtcat gggacatgag cagacacaca tttagcagag caagtattaa 2520  
 gagagtgtca gaacaacaag aaacttctga cttgtattc tcaggcccaa aaaagcctcg 2580  
  
 ggtcgacatc caaaacaag aaaccaaga agaaagtca cattcactcc aaagagaatc 2640  
 gagaccgtgg gagaccgagg aagaaagcga gacagaagcc ctctcgcaag agagccaaga 2700  
 ggtccccctc caacagcagt tgcagcagca gtaccaagag cagctcaagc tcagacaggg 2760  
 aatcaaagtc ctcttcgagc agctcataag gaccaacaa ggggtccatg taaacctatg 2820  
 cctacggtag gtcccaggca gtggctgttt ccagagagaa agccagcccc agctcctagc 2880  
 agtggagact gggccatgga gtttctcgca gcaaaaatat ttgataggcc agttagaagc 2940  
 aaccttaaag atacccctta ctaccatata gttaaaacc aatacaatgt ctactttgac 3000  
  
 cttaaatttg aataaacagc agcttcaaac ttgcaaggcc gtgggagttt cactggtcgg 3060  
 tgtctacctc taaaggtcac taagcactcc gagcgtaacg gaggagtgcg accctcccc 3120  
 ctggaacaac ttcttcggag tccggcgcta cgccttcggc tgcgccggac acctcagacc 3180  
 cccccccac cegaaacgct tgcgcgtttc ggaccttcgg cgtcgggggg gtcgggagct 3240  
 ttattaaacg gactccgaag tgctcttga cactgagggg gtgaacagca acgaaagtga 3300  
 gtggggccag acttcgcat aaggccttta tcttcttccc atttctcagt gtcgggggtc 3360  
 gccataggct tcgggctcgt ttttaggcct tccggactac aaaaatcgcc attttggatga 3420  
  
 cgtcacggcc gccatcttaa gtagttgagg cggacggtgg cgtgagttca aaggtcacca 3480  
 tcagccacac ctactcaaaa tggtaggaaa tttcttcggg gtcaaaggtt acagccgcca 3540  
 tgttaaaaca cgtgacgtat gacgtcacgg ccgccatctt gtgacacaag atggccgact 3600

tccttctct ttttcaaaaa aaagcgggaag tgccgcccgc gcggcggggg gcggcgcgct 3660  
 gcgcgcgccg ccagtaggg ggagccatgc gcccccccc gcgcatgcgc ggggcccccc 3720  
 cccgcggggg gctccgcccc ccggcccccc ccg 3753

<210> 17

<211> 127

<212> PRT

<213> Alphatorquevirus sp.

<400> 17

Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg  
 1                    5                    10                    15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys  
                   20                    25                    30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His  
                   35                    40                    45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro  
                   50                    55                    60

His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln  
 65                    70                    75                    80

Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp  
                   85                    90                    95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala  
                   100                    105                    110

Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu  
                   115                    120                    125

<210> 18

<211> 268

<212> PRT

<213> Alphatorquevirus sp.

<400> 18

Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg  
 1                    5                    10                    15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys

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

<210> 19

<211> 276

<212> PRT

<213> Alphatorquevirus sp.

<400> 19

Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg

1                    5                    10                    15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys

                  20                    25                    30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His

                  35                    40                    45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro

50                    55                    60

His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln

65                    70                    75                    80

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

                  85                    90                    95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala

                  100                    105                    110

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

115                    120                    125

Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro Arg Arg Lys

130                    135                    140

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

145                    150                    155                    160

Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly Pro Leu Pro

                  165                    170                    175

Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala Gln Thr Gly

                  180                    185                    190

Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr Arg Gly Pro

195                    200                    205

Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu Phe Pro Glu

210                      215                      220  
 Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala Met Glu Phe

225                      230                      235                      240  
 Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn Leu Lys Asp

                         245                      250                      255  
 Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val Tyr Phe Asp

                         260                      265                      270  
 Leu Lys Phe Glu

                         275  
 <210> 20  
 <211> 167  
 <212> PRT  
 <213> Alphatorquevirus sp.  
 <400> 20

Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg  
 1                      5                      10                      15

Met Trp Pro Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro  
                          20                      25                      30

Arg Arg Lys Leu Thr Phe Thr Pro Lys Arg Ile Glu Thr Val Gly Asp  
                          35                      40                      45

Arg Gly Arg Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly  
                          50                      55                      60

Pro Leu Pro Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala  
 65                      70                      75                      80

Gln Thr Gly Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr  
                          85                      90                      95

Arg Gly Pro Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu  
                          100                      105                      110

Phe Pro Glu Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala  
                          115                      120                      125

Met Glu Phe Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn

```

130              135              140

Leu Lys Asp Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val
145              150              155              160
Tyr Phe Asp Leu Lys Phe Glu
              165
<210> 21
<211> 743
<212> PRT
<213> Alphatorquevirus sp.
<400> 21
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1          5          10          15
Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
              20              25              30

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

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

Ile Tyr Arg His Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg

```



Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp  
                   420                                  425                                  430  
 Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn  
                   435                                  440                                  445  
 Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro  
                   450                                  455                                  460  
 Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr  
 465                                  470                                  475                                  480  
  
 Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro  
                   485                                  490                                  495  
 Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr  
                   500                                  505                                  510  
 Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser  
                   515                                  520                                  525  
 Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu  
                   530                                  535                                  540  
  
 His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala  
 545                                  550                                  555                                  560  
 Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe  
                   565                                  570                                  575  
 Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp  
                   580                                  585                                  590  
 Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro  
                   595                                  600                                  605  
  
 Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser  
                   610                                  615                                  620  
 Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile  
 625                                  630                                  635                                  640  
 Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly  
                   645                                  650                                  655  
 Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu

660 665 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu

675 680 685

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

690 695 700

Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln

705 710 715 720

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

725 730 735

His Val Asn Pro Cys Leu Arg

740

<210> 22

<211> 194

<212> PRT

<213> Alphatorquevirus sp.

<400> 22

Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys

1 5 10 15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg

20 25 30

Arg Arg Pro Arg Arg Arg Arg Ile Val Lys Asp Pro Ser Phe Gln Pro

35 40 45

Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro Arg Arg Ile Gln Val

50 55 60

Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser Phe Arg Ser Trp Asp

65 70 75 80

Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile Lys Arg Val Ser Glu

85 90 95

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

100 105 110

Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu Ser Ser His Ser Leu

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

Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val His Val Asn Pro Cys  
 180 185 190  
 Leu Arg

<210> 23

<211> 113

<212> PRT

<213> Alphatorquevirus sp.

<400> 23

Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys  
 1 5 10 15  
 Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg  
 20 25 30  
 Arg Arg Pro Arg Arg Arg Arg Ala Gln Lys Ser Leu Gly Ser Thr Ser

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

100 105 110  
 Ser

<210> 24  
<400> 24  
000  
<210> 25  
<400> 25  
000  
<210> 26  
<400> 26  
000  
<210> 27  
<400> 27  
000  
<210> 28  
<400> 28  
000  
<210> 29  
<400> 29  
000  
<210> 30  
<400> 30  
000  
<210> 31  
<400> 31  
000  
<210> 32  
<400> 32  
000  
<210> 33  
<400> 33  
000  
<210> 34  
<400> 34  
000  
<210> 35  
<400> 35  
000

<210> 36  
<400> 36  
000  
<210> 37  
<400> 37  
000  
<210> 38  
<400> 38  
000  
<210> 39  
<400>  
> 39  
000  
<210> 40  
<400> 40  
000  
<210> 41  
<400> 41  
000  
<210> 42  
<400> 42  
000  
<210> 43  
<400> 43  
000  
<210> 44  
<400> 44  
000  
<210> 45  
<400> 45  
000  
<210> 46  
<400> 46  
000  
<210> 47

<400> 47  
 000  
 <210> 48  
 <400> 48  
 000  
 <210> 49  
 <400> 49  
 000  
 <210> 50  
 <400> 50  
 000  
 <210> 51  
 <400> 51  
 000  
 <210> 52  
 <400> 52  
 000  
 <210> 53  
 <400> 53  
 000  
 <210> 54  
 <211> 2979  
 <212> DNA  
 <213> Betatorquevirus sp.  
 <400> 54  
 taataaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa cegtcactta 60  
  
 gttccccttt ttgcaacaac ttctgctttt ttccaactgc cggaaaacca cataatttgc 120  
 atggctaacc acaaactgat atgctaatta acttccacaa aacaacttcc ctttttaaaa 180  
 ccacacctac aaattaatta ttaaacacag tcacatctg ggaggtaacta ccacactata 240  
 ataccaagt cacttccgaa tggctgagtt tatgccgcta gacggagaac gcatcagtta 300  
 ctgactgceg actgaacttg ggcggtgcc gaaggtgagt gaaaccacg aagtcaaggg 360  
 gcaattcggg ctagttcagt ctagcggaac gggcaagaaa cttaaaatta ttttatTTTT 420  
 cagatgagcg actgctttaa accaacatgc tacaacaaca aaacaaagca aactcactgg 480

attaataacc tgcatttaac ccacgacctg atctgcttct gcccaacacc aactagacac 540  
 ttattactag ctttagcaga acaacaagaa acaattgaag tgtctaaaca agaaaaagaa 600  
 aaaataacaa gatgccttat tactacagaa gaagacggta caactacaga cgtcctagat 660  
 ggtatggacg aggttggatt agacgccctt ttcgcagaag atttcgaaga aaaagaaggg 720  
 taagacctac ttatactact attcctctaa agcaatggca accgcatat aaaagaacat 780  
 gctatataaa aggacaagac tgtttaatat actatagcaa ctttaagactg ggaatgaata 840  
 gtacaatgta tgaaaaaagt attgtacctg tacattggcc gggaggggggt tcttttctg 900  
  
 taagcatggt aacttttagat gccttgtatg atatacataa actttgtaga aactggtgga 960  
 catccacaaa ccaagactta ccactagtaa gatataaagg atgcaaaaata acattttatc 1020  
 aaagcacatt tacagactac atagtaagaa tacatacaga actaccagct aacagtaaca 1080  
 aactaacata cccaacaca catccactaa tgatgatgat gtctaagtac aaacacatta 1140  
 tacctagtag acaacaaga agaaaaaga aaccatacac aaaaatattt gtaaaaccac 1200  
 ctccgcaatt tgaaaacaaa tggactttg ctacagacct ctacaaaatt ccattactac 1260  
 aaatacactg cacagcatgc aacttacaaa acccatttgt aaaaccagac aaattatcaa 1320  
  
 acaatgttac attatggta ctaaacacca taagcataca aaatagaac atgtcagtgg 1380  
 atcaaggaca atcatggcca tttaaatac taggaacaca aagcttttat ttttactttt 1440  
 acaccggagc aaacctacca ggtgacacaa cacaataacc agtagcagac ctattaccac 1500  
 taacaaacc aagaataaac agaccaggac aatcactaaa tgaggcaaaa attacagacc 1560  
 atattacttt cacagaatac aaaaacaaat ttacaaatta ttgggtaac ccatttaata 1620  
 aacacattca agaacaccta gatatgatac tatactcact aaaaagtcca gaagcaataa 1680  
 aaaacgaatg gacaacagaa aacatgaaat ggaaccaatt aaacaatgca ggaacaatgg 1740  
  
 cattaacacc atttaacgag ccaatattca cacaataca atataacca gatagagaca 1800  
 caggagaaga cactcaatta tactactct ctaacgctac aggaacagga tgggaccac 1860  
 caggaattcc agaattaata ctagaaggat ttccactatg gttaatatat tggggatttg 1920  
 cagactttca aaaaaaccta aaaaaagtaa caaacataga cacaattac atgttagtag 1980  
 caaaaacaaa atttacaaa aaacctggca cattctactt agtaatacta aatgacacct 2040  
 ttgtagaagg caatagccca tatgaaaaac aacctttacc tgaagacaac attaaatggt 2100  
 acccacaagt acaatacaca ttagaagcac aaaacaaact actacaaact gggccattta 2160  
  
 caccaaacat acaaggacaa ctatcagaca atatatcaat gttttataaa ttttacttta 2220  
 aatggggagg aagcccacca aaagcaatta atgttgaaaa tctgcccac cagattcaat 2280  
 atcccatacc ccgtaacgag catgaacaa cttcgttaca gagtccaggg gaagccccag 2340

aatccatctt atactccttc gactatagac acgggaacta cacaacaaca gctttgtcac 2400  
 gaattagcca agactgggca cttaaagaca ctgtttctaa aattacagag ccagatcgac 2460  
 agcaactgct caaacaagcc ctcgaatgcc tgcaaatctc ggaagaaacg caggagaaaa 2520  
 aagaaaaaga agtacagcag ctcatcagca acctcagaca gcagcagcag ctgtacagag 2580

agcgaataat atcattatta aaggaccaat aacttttaac tgtgtaaaaa aggtgaatt 2640  
 gtttgatgat aaaccaaaaa accgtagatt tacacctgag gaatttgaaa ctgagttaca 2700  
 aatagcaaaa tggtaaaga gaccccaag atcctttgta aatgatcctc ccttttacc 2760  
 atggttacca cctgaacctg ttgtaaactt taagcttaat tttactgaat aaaggccagc 2820  
 attaattcac ttaaggagtc tgtttattta agttaaactt taataaacgg tcaccgctc 2880  
 cctaatacgc aggcgcagaa agggggctcc gccccctta acccccaggg ggctccgcc 2940  
 cctgaaacc ccaagggggc tacgccccct tacaccccc 2979

<210> 55

<211> 99

<212> PRT

<213> Betatorquevirus sp.

<400> 55

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

<210> 56

<211> 203

<212> PRT

<213> Betatorquevirus sp.

<400> 56

Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln  
 1                    5                    10                    15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe  
                   20                    25                    30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln  
                   35                    40                    45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys  
                   50                    55                    60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly  
 65                    70                    75                    80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu  
                   85                    90                    95

Lys Glu Gly Phe Asn Ile Pro Tyr Pro Val Thr Ser Met Lys Gln Leu  
                   100                    105                    110

Arg Tyr Arg Val Gln Gly Lys Pro Gln Asn Pro Ser Tyr Thr Pro Ser  
                   115                    120                    125

Thr Ile Asp Thr Gly Thr Thr Gln Gln Gln Leu Cys His Glu Leu Ala  
                   130                    135                    140

Lys Thr Gly His Leu Lys Thr Leu Phe Leu Lys Leu Gln Ser Gln Ile  
 145                    150                    155                    160

Asp Ser Asn Cys Ser Asn Lys Pro Ser Asn Ala Cys Lys Ser Arg Lys  
                   165                    170                    175

Lys Arg Arg Arg Lys Lys Lys Lys Lys Tyr Ser Ser Ser Ser Ala Thr  
                   180                    185                    190

Ser Asp Ser Ser Ser Ser Cys Thr Glu Ser Glu  
                   195                    200

<210> 57

<211> 219

<212> PRT

<213> Betatorquevirus sp.

<400> 57

Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln

1                    5                    10                    15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe

                  20                    25                    30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln

                  35                    40                    45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys

                  50                    55                    60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly

65                    70                    75                    80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu

                  85                    90                    95

Lys Glu Gly Ala Arg Ser Thr Ala Thr Ala Gln Thr Ser Pro Arg Met

                  100                    105                    110

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

                  115                    120                    125

Ala Ala His Gln Gln Pro Gln Thr Ala Ala Ala Ala Val Gln Arg Ala

                  130                    135                    140

Asn Asn Ile Ile Ile Lys Gly Pro Ile Thr Phe Asn Cys Val Lys Lys

145                    150                    155                    160

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

                  165                    170                    175

Glu Phe Glu Thr Glu Leu Gln Ile Ala Lys Trp Leu Lys Arg Pro Pro

                  180                    185                    190

Arg Ser Phe Val Asn Asp Pro Pro Phe Tyr Pro Trp Leu Pro Pro Glu

                  195                    200                    205

Pro Val Val Asn Phe Lys Leu Asn Phe Thr Glu

                  210                    215

<210> 58

<211> 666

<212> PRT

<213> Betatorquevirus sp.

<400> 58

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

210                    215                    220  
 Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr  
 225                    230                    235                    240  
 Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val  
                          245                    250                    255  
 Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe  
                          260                    265                    270  
 Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln  
  
                          275                    280                    285  
 Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg  
                          290                    295                    300  
 Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe  
 305                    310                    315                    320  
 Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn  
                          325                    330                    335  
 Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser  
  
                          340                    345                    350  
 Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn  
                          355                    360                    365  
 Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro  
                          370                    375                    380  
 Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp  
 385                    390                    395                    400  
 Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro  
  
                          405                    410                    415  
 Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile  
                          420                    425                    430  
 Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn  
                          435                    440                    445  
 Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys  
                          450                    455                    460

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly

465 470 475 480

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp

485 490 495

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln

500 505 510

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

515 520 525

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys

530 535 540

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro

545 550 555 560

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

565 570 575

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr

580 585 590

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

595 600 605

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu

610 615 620

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu

625 630 635 640

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Leu Tyr Arg

645 650 655

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln

660 665

<210> 59

<211> 148

<212> PRT

<213> Betatorquevirus sp.

<400> 59

Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg  
 1                    5                    10                    15  
 Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg  
                   20                    25                    30  
 Arg Lys Arg Arg Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr  
                   35                    40                    45  
 Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser  
  
                   50                    55                    60  
 Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile  
 65                    70                    75                    80  
 Ser Gln Asp Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro  
                   85                    90                    95  
 Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser  
                   100                    105                    110  
 Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu Val Gln Gln Leu Ile Ser  
  
                   115                    120                    125  
 Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu  
                   130                    135                    140  
 Leu Lys Asp Gln  
 145  
 <210> 60  
 <211> 82  
 <212> PRT  
 <213> Betatorquevirus sp.  
 <400> 60  
 Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg  
 1                    5                    10                    15  
 Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg  
                   20                    25                    30  
  
 Arg Lys Arg Arg Ser Gln Ile Asp Ser Asn Cys Ser Asn Lys Pro Ser  
                   35                    40                    45  
 Asn Ala Cys Lys Ser Arg Lys Lys Arg Arg Arg Lys Lys Lys Lys Lys



000

<210> 71

<400> 71

000

<210> 72

<400> 72

000

<210> 73

<400> 73

000

<210> 74

<400> 74

000

<210> 75

<400> 75

000

<210> 76

<400> 76

000

<210> 77

<400> 77

000

<210> 78

<400> 78

000

<210> 79

<400> 79

000

<210> 80

<400> 80

000

<210> 81

<400> 81

000

<210> 82

<400> 82

000

<210> 83

<400> 83

000

<210> 84

<400> 84

000

<210> 85

<400> 85

000

<210> 86

<400> 86

000

<210> 87

<400> 87

000

<210> 88

<400> 88

000

<210> 89

<400> 89

000

<210> 90

<400> 90

000

<210> 91

<400> 91

000

<210> 92

<400> 92

000

<210> 93

<400> 93

000

<210> 94

<400> 94

000

<210> 95

<400> 95

000

<210> 96

<400> 96

000

<210> 97

<400> 97

000

<210> 98

<400> 98

000

<210> 99

<400> 99

000

<210> 100

<400> 100

000

<

210> 101

<400> 101

000

<210> 102

<400> 102

000

<210> 103

<400> 103

000

<210> 104

<400> 104

000

<210> 105  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 105  
 cgggtgccgk aggtgagttt acacaccgma gtcaaggggc aattcgggct crggactggc      60  
 cgggcyhtgg g      71  
 <210> 106  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 106  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc      60  
 cgggctwtgg g      71  
 <210> 107  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 107  
 cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc      60  
 cgggctatgg g      71  
 <210> 108  
  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide

<400> 108  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggccctgg g 71  
 <210> 109  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 109  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggctttgg g 71  
 <210> 110  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 110  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggctatgg g 71  
 <210> 111  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 111  
 cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60  
 cgggctttgg g 71  
 <210> 112  
 <211> 71  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 112  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggcyhtgg g 71  
 <210> 113  
 <211> 71  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 113  
 cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggctatgg g 71  
 <210> 114  
 <211> 70  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 114  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggcccggg 70

<210> 115  
 <211> 71  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 115  
 cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60  
 cgggctttgg g 71

<210> 116  
 <211> 69  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 116  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggaggccg 60  
  
 ggccatggg 69  
 <210> 117  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 117  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggccccgg g 71  
 <210> 118  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
  
 <400> 118  
 cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggctatgg g 71  
 <210> 119  
 <211> 71  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide

<400> 119  
 cgggtgccga aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60  
 cgggctatgg g 71

<210> 120  
 <211> 117  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polynucleotide  
 <220><221> misc\_feature  
 <222> (10)..(10)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (12)..(12)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (30)..(32)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (34)..(34)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (43)..(46)  
 <223> May or may not be present  
 <220><221>  
 > misc\_feature  
 <222> (52)..(54)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (70)..(71)  
 <223> May or may not be present  
 <220><221> misc\_feature  
 <222> (89)..(90)  
 <223> May or may not be present

<220><221> misc\_feature  
 <222> (103)..(103)  
 <223> May or may not be present  
 <400> 120  
 cggcggsggs gcsscgcgct dgcgcgcgsg cccrsyruggg grdssmmwgc skscscccc 60  
 cscgcgcatg cgcrcgggkc cccccccyv sggggggctc egcccccccg gcccccc 117  
 <210> 121  
 <211> 169  
 <212> DNA  
 <213  
 > Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     polynucleotide  
 <220><221> modified\_base  
 <222> (20)..(20)  
 <223> a, c, t, g, unknown or other  
 <220><221> modified\_base  
 <222> (22)..(22)  
 <223> a, c, t, g, unknown or other  
 <220><221> modified\_base  
 <222> (40)..(42)  
 <223> a, c, t, g, unknown or other  
 <220><221> modified\_base  
 <222> (53)..(56)  
 <223> a, c, t, g, unknown or other  
 <220><221> modified\_base  
 <222> (62)..(62)  
 <223> a, c, t, g, unknown or other  
  
 <220><221> modified\_base  
 <222> (64)..(64)  
 <223> a, c, t, g, unknown or other  
 <220><221> modified\_base  
 <222> (97)..(98)  
 <223> a, c, t, g, unknown or other

<400> 121  
 gccgccgagg cgaggggsggn gnsagcagct daggcagcsn nncrccragg ggnnnncwgc 60  
 sncnccccc cccagcagatg aggggkccc cccccnncg gggggctccg cccccagcc 120  
 cccccagtg ctaaaccac agcagatgag agaccagcc cccagcagcc 169

<210> 122

<211> 79

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified\_base

<222> (20)..(20)

<223> a, c, t, g, unknown or other

<220><221> modified\_base

<222> (22)..(22)

<223> a, c, t, g, unknown or other

<220><221> modified\_base

<222> (40)..(42)

<223> a, c, t, g, unknown or other

<220><221> modified\_base

<222> (53)..(56)

<223> a, c, t, g, unknown or other

<220><221> modified\_base

<222> (62)..(62)

<223> a, c, t, g, unknown or other

<220><221> modified\_base

<222> (64)..(64)

<223> a, c, t, g, unknown or other

<400> 122

gccgccgagg cgaggggsggn gnsagcagct daggcagcsn nncrccragg ggnnnncwgc 60  
 sncnccccc cccagcagatg 79

<210> 123

<211> 31

<212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <220><221> modified\_base  
 <222> (18)..(19)  
 <223> a, c, t, g, unknown or other  
 <400> 123  
 gcgcgggkcc ccccccnnc ggggggctcc g 31

<210> 124  
 <211> 59  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 124  
 cccccggcc cccccctg ctaaaccac cgcgatgcg cgaccagcc cccgccc 59

<210> 125  
 <211> 156  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     polynucleotide  
 <400> 125  
 gcggcggggg ggcggccgcg ttcgcgcgc gccaccagg ggggtctgcg ccccccccc 60  
 cgcgatgcg cggggccccc ccccgggggg gctccgcccc cccggcccc ccccgtgcta 120  
 aaccaccgc gcatgcgcga ccacgcccc gccgcc 156

<210> 126  
 <211> 7  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide

<400> 126  
 gcggcgg 7  
 <210> 127  
 <211> 7  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 127  
 gggggcg 7  
 <210> 128  
 <211> 6  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 128  
 gccgcg 6  
 <210> 129  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 129  
 ttcgcgcgcc gccaccagg ggtg 25  
 <210> 130  
 <211> 5  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
  
 oligonucleotide  
 <400> 130

ctgcg	5
<210> 131	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 131	
cgccccccc cgcgcat	17
<210> 132	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 132	
gcgcggggcc ccccc	17
<210> 133	
<211> 72	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 133	
gggggggctc cgccccccg gccccccc gtgctaaacc caccgcat gcgcgaccac	60
gccccgccc cc	72
<210> 134	
<211> 115	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic polynucleotide	
<400> 134	

cggcggcggc ggcgcgcgcg ctgcgcgcgc gcgccggggg ggcgccagcg ccccccccc 60  
  
 cgcgcatgca cgggtcccc cccccaggg gggtccgcc ccccggcccc ccccc 115  
 <210> 135  
 <211> 14  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 135  
 cggcggcggc ggcg 14  
 <210> 136  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 136  
 cgcgcgctgc ggcgcg 17  
  
 <210> 137  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 137  
 cgccgggggg ggcgccagcg 19  
 <210> 138  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     oligonucleotide  
 <400> 138

ccccccccc cgcgcat 17

<210> 139

<211> 31

<212> DNA

<213> Artificial Sequence

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

<400> 139

gcacgggtcc cccccccac ggggggctcc g 31

<210> 140

<211> 17

<212> DNA

<213> Artificial Sequence

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

<400> 140

ccccccggcc ccccccc 17

<210> 141

<211> 121

<212> DNA

<213> Artificial Sequence

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

<400> 141

ccgtcggcgg gggggccgcg cgctgcgcgc gcggcccccg ggggaggcac agcctcccc 60

ccccgcgcgc atgcgcgagg gtccccccc ctccgggggg ctccgcccc cggccccccc 120

c 121

<210> 142

<211> 37

<212> DNA

<213> Artificial Sequence

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

<400> 142  
 ccgtcggcgg gggggccgcg cgctgcgcgc gcggccc 37  
 <210> 143

<211> 84  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 143  
 ccgggggagg cacagcctcc ccccccgcg cgcgatgcgcg cgggtcccc ccctccggg 60  
 gggctccgcc ccccgcccc cccc 84  
 <210> 144  
 <211> 104  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> 144  
 cggcggcggc gcgcgcgcta cgcgcgcgcg cggggggct gccgcccc ccccgcat 60  
 gcgcggggcc ccccccgcg gggggctccg cccccggcc cccc 104  
 <210> 145  
 <211> 11  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 145  
 cggcggcggc g 11  
 <210> 146  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide  
 <400> 146  
 cgcgcgctac gcgcgcg 17

<210> 147  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 147  
 cgccgggggg 10  
 <210> 148  
 <211> 7  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 148  
 ctgccgc 7  
 <210> 149  
 <211> 15  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 149  
 ccccccccg cgcgcat 15  
 <210> 150  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 150  
gcgcggggcc cccccc 17  
<210> 151  
<211> 13  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
oligonucleotide  
<400> 151  
gcggggggct ccg 13  
<210> 152  
<211> 14  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
oligonucleotide  
<400> 152  
cccccggcc cccc 14  
<210> 153  
<211> 122  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
polynucleotide  
<400> 153  
gcccgcgcgg cggcgggggg cggcgcgctg cgcgccgc ccagtagggg gagccatgcg 60  
ccccccccg cgcgatgcgcg gggcccccc cgcggggggg ctccgcccc cggcccccc 120  
cg 122  
<210> 154  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 154

gccgccgcgg cggcggggg 19

<210> 155

<211> 41

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 155

gcggcgcgct gcgcgcgccg cccagtaggg ggagccatgc g 41

<210> 156

<211> 15

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 156

ccccccccg cgcgcat 15

<210> 157

<211> 17

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 157

gcgcggggcc ccccccc 17

<210> 158

<211> 13

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 158  
gccccgggct ccg 13  
<210> 159  
<211> 17  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
oligonucleotide  
<400> 159  
ccccccggcc cccccg 17  
<210> 160  
<211> 36  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
oligonucleotide  
<400> 160  
cgcgctgcg cgcgcccca gtaggggag ccatgc 36  
<210> 161  
<211> 78  
<212> DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
oligonucleotide  
<400> 161  
ccgccatctt aagtagttga ggcggacggt ggcgtgagtt caaaggtcac catcagccac 60  
acctactcaa aatggtgg 78  
<210> 162  
<211> 172  
<212>  
DNA  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 162  
 cttaagtagt tgaggcggac ggtggcgtga gttcaaaggt caccatcagc cacacctact 60  
 caaaatggtg gacaatttct tccgggtcaa aggttacagc cgccatgtta aaacacgtga 120  
 cgtatgacgt cacggccgcc atttgtgac acaagatggc cgacttcctt cc 172

<210> 163  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 163

cgcgctgcgc gcgcccacca gtagggggag ccatgc 36

<210> 164  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 164

gcgctdcgcg cgcgcccggg ggggctgcgc ccccc 36

<210> 165  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 165

gcgcttcgcg cgccgcccac tagggggcgt tgcgcg 36

<210> 166  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide  
 <400> 166  
 gcgctgcgcg cgccgccag tagggggcgc aatgcg 36  
 <210> 167  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide  
 <400> 167  
 gcgctgcgcg cgcgccccc gggggaggca ttgctt 36  
 <210> 168  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 168  
 gcgctgcgcg cgcgcccg gggggcgcca gcgccc 36  
 <210> 169  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 169  
 gcgcttcgcg cgcgcccg ggggtccgc ccccc 36  
 <210> 170  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence

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

oligonucleotide

<400> 170  
 gcgcttcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 171  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 171  
 gcgctacgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 172  
 <211> 36  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> 172  
 gcgctacgcg cgcgcgccgg ggggctctgc cccccc 36

<210> 173  
 <400> 173  
 000

<210> 174  
 <400> 174  
 000

<210> 175  
 <400> 175  
 000

<210> 176  
 <400> 176  
 000

<210> 177  
 <400> 177  
 000

<210> 178

<400> 178

000

<210> 179

<400> 179

000

<210> 180

<400> 180

000

<210> 181

<400> 181

000

<210> 182

<400> 182

000

<210> 183

<400> 183

000

<210> 184

<400> 184

000

<210> 185

<211> 743

<212> PRT

<213> Alphatorquevirus sp.

<400> 185

Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys

1                    5                                    10    15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg

                  20                                    25    30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg

                  35                                    40    45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg

                  50                                    55    60

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



Cys Ile Ser His Pro Gln Leu Lys Lys Pro Asn Pro Gln Ile Asn Lys  
 325 330 335  
  
 Pro Leu Glu Ser Gln Tyr Phe Ala Pro Leu Asp Ala Leu Trp Gly Asp  
 340 345 350  
 Pro Ile Tyr Tyr Asn Asp Leu Asn Glu Asn Lys Ser Leu Asn Asp Ile  
 355 360 365  
 Ile Glu Lys Ile Leu Ile Lys Asn Met Ile Thr Tyr His Ala Lys Leu  
 370 375 380  
 Arg Glu Phe Pro Asn Ser Tyr Gln Gly Asn Lys Ala Phe Cys His Leu  
 385 390 395 400  
  
 Thr Gly Ile Tyr Ser Pro Pro Tyr Leu Asn Gln Gly Arg Ile Ser Pro  
 405 410 415  
 Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp  
 420 425 430  
 Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn  
 435 440 445  
 Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro  
 450 455 460  
  
 Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr  
 465 470 475 480  
 Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro  
 485 490 495  
 Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr  
 500 505 510  
 Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser  
 515 520 525  
  
 Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu  
 530 535 540  
 His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala  
 545 550 555 560  
 Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe

565 570 575  
 Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp  
 580 585 590

Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro  
 595 600 605

Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser  
 610 615 620

Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile  
 625 630 635 640

Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly  
 645 650 655

Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu  
 660 665 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu  
 675 680 685

Glu Ser Glu Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe  
 690 695 700

Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln  
 705 710 715 720

Gly Ile Lys Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val  
 725 730 735

His Val Asn Pro Cys Leu Arg  
 740

<210> 186

<211> 68

<212> PRT

<213> Alphatorquevirus sp.

<400> 186

Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys  
 1 5 10 15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg

20 25 30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg  
 35 40 45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg  
 50 55 60

Arg Gly Arg Lys  
 65

<210> 187

<211> 212

<212> PRT

<213> Alphatorquevirus sp.

<400> 187

Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys  
 1 5 10 15

Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe  
 20 25 30

Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro  
 35 40 45

Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe  
 50 55 60

Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn  
 65 70 75 80

Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His  
 85 90 95

Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg Thr Pro Leu Gly  
 100 105 110

Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly Asn Ala Ile Leu  
 115 120 125

Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr Arg Pro Lys Gly  
 130 135 140

Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp

145                    150                    155                    160  
 Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr Leu Phe Asn Ile  
                                  165                    170                    175  
 Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr  
                                  180                    185                    190  
 Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser Val Tyr Asn Asn  
                                  195                    200                    205

Tyr Leu Ser Ile

210

<210> 188

<211> 133

<212> PRT

<213> Alphatorquevirus sp.

<400> 188

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

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

Gly Asn Lys Ala Phe Cys His Leu Thr Gly Ile Tyr Ser Pro Pro Tyr  
                                  115                    120                    125  
 Leu Asn Gln Gly Arg  
 130

<210> 189

<211> 166

<212> PRT

<213> Alphatorquevirus sp.

<400> 189

Ile Ser Pro Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro

1                    5                    10                    15

Tyr Thr Asp Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr

                  20                    25                    30

Lys Glu Asn Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr

                  35                    40                    45

Asp Met Pro Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys

                  50                    55                    60

Lys Asp Thr Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu

65                    70                    75                    80

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

                  85                    90                    95

Tyr Gly Tyr Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro

                  100                    105                    110

Asp Gly Ser Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro

                  115                    120                    125

Thr Val Leu His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly

                  130                    135                    140

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

145                    150                    155                    160

Tyr Cys Phe Asn Phe Asn

                  165

<210> 190

<211> 164

<212> PRT

<213> Alphatorquevirus sp.

<400> 190

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

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

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

<210> 191

<400> 191

000

<210> 192

<400> 192

000

<210> 193

<400> 193

000

<210> 194

<400> 194  
000  
<210> 195  
<400> 195  
000  
<210> 196  
<400> 196  
000  
<210> 197  
<400> 197  
000  
<210> 198  
<400> 198  
000  
<210> 199  
<400> 199  
000  
<210> 200  
<400> 200  
000  
<210> 201  
<400> 201  
000  
<210> 202  
<400> 202  
000  
<210> 203  
<400> 203  
000  
<210> 204  
<400> 204  
000  
<210> 205  
<400> 205  
000  
<210> 206

<400> 206

000

<210> 207

<400> 207

000

<210> 208

<400> 208

000

<210> 209

<400> 209

000

<210> 210

<400> 210

000

<210> 211

<400> 211

000

<210> 212

<400> 212

000

<210> 213

<400> 213

000

<210> 214

<400> 214

000

<210> 215

<211> 666

<212> PRT

<213> Betatorquevirus sp.

<400> 215

Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg

1                    5                    10                    15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg

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

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

515 520 525  
 Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys

530 535 540  
 Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro

545 550 555 560  
 Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro

565 570 575  
 Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr

580 585 590  
 Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val

595 600 605  
 Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu

610 615 620  
 Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu

625 630 635 640  
 Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg

645 650 655  
 Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln

660 665  
 <210> 216

<211> 38  
 <212> PRT

<213> Betatorquevirus sp.  
 <400> 216

Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg

1 5 10 15  
 Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg

20 25 30  
 Arg Lys Arg Arg Val Arg

35  
 <210> 217

<211> 208

<212> PRT

<213> Betatorquevirus sp.

<400> 217

Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln Trp Gln Pro Pro Tyr Lys

1                    5                    10                    15

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

20                    25                    30

Leu Arg Leu Gly Met Asn Ser Thr Met Tyr Glu Lys Ser Ile Val Pro

35                    40                    45

Val His Trp Pro Gly Gly Gly Ser Phe Ser Val Ser Met Leu Thr Leu

50                    55                    60

Asp Ala Leu Tyr Asp Ile His Lys Leu Cys Arg Asn Trp Trp Thr Ser

65                    70                    75                    80

Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr Lys Gly Cys Lys Ile Thr

85                    90                    95

Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile Val Arg Ile His Thr Glu

100                    105                    110

Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr Pro Asn Thr His Pro Leu

115                    120                    125

Met Met Met Met Ser Lys Tyr Lys His Ile Ile Pro Ser Arg Gln Thr

130                    135                    140

Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile Phe Val Lys Pro Pro Pro

145                    150                    155                    160

Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr Asp Leu Tyr Lys Ile Pro

165                    170                    175

Leu Leu Gln Ile His Cys Thr Ala Cys Asn Leu Gln Asn Pro Phe Val

180                    185                    190

Lys Pro Asp Lys Leu Ser Asn Asn Val Thr Leu Trp Ser Leu Asn Thr

195                    200                    205

<210> 218

<211> 128

<212> PRT

<213> Betatorquevirus sp.

<400> 218

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

Glu Ala Lys Ile Thr Asp His Ile Thr Phe Thr Glu Tyr Lys Asn Lys  
 65                    70                    75                    80  
 Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn Lys His Ile Gln Glu His  
                   85                    90                    95  
 Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser Pro Glu Ala Ile Lys Asn  
                   100                    105                    110  
 Glu Trp Thr Thr Glu Asn Met Lys Trp Asn Gln Leu Asn Asn Ala Gly  
                   115                    120                    125

<210> 219

<211> 163

<212> PRT

<213> Betatorquevirus sp.

<400> 219

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

Phe Gln Lys Asn Leu Lys Lys Val Thr Asn Ile Asp Thr Asn Tyr Met  
 65                      70                      75                      80  
 Leu Val Ala Lys Thr Lys Phe Thr Gln Lys Pro Gly Thr Phe Tyr Leu  
                     85                      90                      95  
 Val Ile Leu Asn Asp Thr Phe Val Glu Gly Asn Ser Pro Tyr Glu Lys  
                     100                      105                      110  
 Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp Tyr Pro Gln Val Gln Tyr  
  
                     115                      120                      125  
 Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln Thr Gly Pro Phe Thr Pro  
                     130                      135                      140  
 Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile Ser Met Phe Tyr Lys Phe  
 145                      150                      155                      160  
 Tyr Phe Lys

<210> 220

<211> 129

<212> PRT

<213> Betatorquevirus sp.

<400> 220

Trp Gly Gly Ser Pro Pro Lys Ala Ile Asn Val Glu Asn Pro Ala His  
 1                      5                      10                      15

Gln Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr Thr Ser Leu  
                     20                      25                      30

Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr  
                     35                      40                      45

Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp  
                     50                      55                      60

Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro Asp Arg Gln  
 65                      70                      75                      80

Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr  
                     85                      90                      95

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

100 105 110  
 Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp

115 120 125

Gln

<210> 221

<400> 221

000

<210> 222

<400> 222

000

<210> 223

<400> 223

000

<210> 224

<400> 224

000

<210> 225

<400> 225

000

<210> 226

<400> 226

000

<210> 227

<211> 220

<212> PRT

<213> Artificial Sequence

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

polypeptide

<220><221> MOD\_RES

<222> (29)..(31)

<223> Any amino acid

<220><221> SITE

<222> (29)..(31)

<223> This region may encompass 0-3 residues

<220><221> MOD\_RES

<222> (100)..(100)

<223> Any amino acid

<220><221> MOD\_RES

<222> (125)..(129)

<223> Any amino acid

<220><221> SITE

<222> (125)..(129)

<223> This region may encompass 1-5 residues

<220><221> MOD\_RES

<222> (181)..(181)

<223> Any amino acid

<220><221> MOD\_RES

<222> (211)..(211)

<223> Any amino acid

<400> 227

Leu Val Leu Thr Gln Trp Gln Pro Asn Thr Val Arg Arg Cys Tyr Ile

1                    5                    10                    15

Arg Gly Tyr Leu Pro Leu Ile Ile Cys Gly Glu Asn Xaa Xaa Xaa Thr

                  20                    25                    30

Thr Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Ile Gln Lys Gly

                  35                    40                    45

Pro Phe Gly Gly Gly Met Ser Thr Thr Thr Phe Ser Leu Arg Val Leu

50                    55                    60

Tyr Asp Glu Tyr Gln Arg Phe Met Asn Arg Trp Thr Tyr Ser Asn Glu

65                    70                    75                    80

Asp Leu Asp Leu Ala Arg Tyr Leu Gly Cys Lys Phe Thr Phe Tyr Arg

                  85                    90                    95

His Pro Asp Xaa Asp Phe Ile Val Gln Tyr Asn Thr Asn Pro Pro Phe

                  100                    105                    110

Lys Asp Thr Lys Leu Thr Ala Pro Ser Ile His Pro Xaa Xaa Xaa Xaa

115                      120                      125  
 Xaa Gly Met Leu Met Leu Ser Lys Arg Lys Ile Leu Ile Pro Ser Leu  
 130                      135                      140  
 Lys Thr Arg Pro Lys Gly Lys His Tyr Val Lys Val Arg Ile Gly Pro  
 145                      150                      155                      160  
 Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Ser Asp Leu Cys Asp  
  
 165                      170                      175  
 Val Pro Leu Val Xaa Leu Tyr Ala Thr Ala Ala Asp Leu Gln His Pro  
 180                      185                      190  
 Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys Val Thr Phe Gln Val Leu  
 195                      200                      205  
 Gly Ser Xaa Tyr Asn Lys His Leu Ser Ile Ser Pro  
 210                      215                      220

<210> 228

<211> 172

<212> PRT

<213> Artificial Sequence

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

polypeptide

<220><221> MOD\_RES

<222> (38)..(38)

<223> Any amino acid

<220><221> MOD\_RES

<222> (44)..(46)

<223> Any amino acid

<220><221> SITE

<222> (44)..(46)

<223> This region may encompass 0-3 residues

<220><221> MOD\_RES

<222> (77)..(77)

<223> Any amino acid

<220><221> MOD\_RES

<222> (79)..(79)

<223> Any amino acid

<220><221> MOD\_RES

<222> (98)..(101)

<223> Any amino acid

<220><221> SITE

<222> (98)..(101)

<223> This region may encompass 0-4 residues

<400> 228

Ser Asn Phe Glu Phe Pro Gly Ala Tyr Thr Asp Ile Thr Tyr Asn Pro  
 1                    5                    10                    15

Leu Thr Asp Lys Gly Val Gly Asn Met Val Trp Ile Gln Tyr Leu Thr  
                   20                    25                    30

Lys Pro Asp Thr Ile Xaa Asp Lys Thr Gln Ser Xaa Xaa Xaa Lys Cys  
                   35                    40                    45

Leu Ile Glu Asp Leu Pro Leu Trp Ala Ala Leu Tyr Gly Tyr Val Asp  
                   50                    55                    60

Phe Cys Glu Lys Glu Thr Gly Asp Ser Ala Ile Ile Xaa Asn Xaa Gly  
 65                    70                    75                    80

Arg Val Leu Ile Arg Cys Pro Tyr Thr Lys Pro Pro Leu Tyr Asp Lys  
                   85                    90                    95

Thr Xaa Xaa Xaa Xaa Asn Lys Gly Phe Val Pro Tyr Ser Thr Asn Phe  
                   100                    105                    110

Gly Asn Gly Lys Met Pro Gly Gly Ser Gly Tyr Val Pro Ile Tyr Trp  
                   115                    120                    125

Arg Ala Arg Trp Tyr Pro Thr Leu Phe His Gln Lys Glu Val Leu Glu  
                   130                    135                    140

Asp Ile Val Gln Ser Gly Pro Phe Ala Tyr Lys Asp Glu Lys Pro Ser  
 145                    150                    155                    160

Thr Gln Leu Val Met Lys Tyr Cys Phe Asn Phe Asn  
                   165                    170

<210> 229

<211> 258

<212> PRT  
<213> Artificial Sequence  
<220><223> Description of Artificial Sequence: Synthetic  
polypeptide  
<220  
><221> MOD\_RES  
<222> (20)..(22)  
<223> Any amino acid  
<220><221> SITE  
<222> (20)..(22)  
<223> This region may encompass 0-3 residues  
<220><221> MOD\_RES  
<222> (25)..(25)  
<223> Any amino acid  
<220><221> MOD\_RES  
<222> (78)..(78)  
<223> Any amino acid  
<220><221> MOD\_RES  
<222> (89)..(89)  
<223> Any amino acid  
<220><221> MOD\_RES  
<222> (91)..(91)  
<223> Any amino acid  
<220><221> MOD\_RES  
<222> (95)..(98)  
<223> Any amino acid  
<220><221> SITE  
<222> (95)..(98)  
<223> This region may encompass 1-4 residues  
  
<220><221> MOD\_RES  
<222> (107)..(120)  
<223> Any amino acid  
<220><221> SITE  
<222> (107)..(120)

<223> This region may encompass 2-14 residues

<220><221> MOD\_RES

<222> (129)..(129)

<223> Any amino acid

<220><221> MOD\_RES

<222> (139)..(168)

<223> Any amino acid

<220><221> SITE

<222> (139)..(168)

<223> This region may encompass 0-30 residues

<220><221> MOD\_RES

<222> (201)..(204)

<223> Any amino acid

<220><221> SITE

<222> (201)..(204)

<223> This region may encompass 0-4 residues

<220><221> MOD\_RES

<222> (219)..(258)

<223> Any amino acid

<220><221> SITE

<222> (219)..(258)

<223> This region may encompass 0-40 residues

<400> 229

Trp Gly Gly Asn Pro Ile Ser Gln Gln Val Val Arg Asn Pro Cys Lys

1                    5                    10                    15

Asp Ser Gly Xaa Xaa Xaa Ser Gly Xaa Gly Arg Gln Pro Arg Ser Val

                  20                    25                    30

Gln Val Val Asp Pro Lys Tyr Met Gly Pro Glu Tyr Thr Phe His Ser

                  35                    40                    45

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

                  50                    55                    60

Ser Glu Gln Pro Thr Asp Asp Glu Ile Phe Thr Gly Gly Xaa Pro Lys

65                    70                    75                    80

Arg Pro Arg Arg Asp Pro Pro Thr Xaa Gln Xaa Pro Glu Glu Xaa Xaa  
 85 90 95

Xaa Xaa Gln Lys Glu Ser Ser Ser Phe Arg Xaa Xaa Xaa Xaa Xaa Xaa  
 100 105 110

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Trp Glu Ser Ser Ser Gln Glu  
 115 120 125

Xaa Glu Ser Glu Ser Gln Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa  
 130 135 140

Xaa  
 145 150 155 160

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Gln Thr Val Gln Gln Gln Leu  
 165 170 175

Arg Gln Gln Leu Arg Glu Gln Arg Arg Leu Arg Val Gln Leu Gln Leu  
 180 185 190

Leu Phe Gln Gln Leu Leu Lys Thr Xaa Xaa Xaa Xaa Gln Ala Gly Leu  
 195 200 205

His Ile Asn Pro Leu Leu Leu Ser Gln Ala Xaa Xaa Xaa Xaa Xaa Xaa  
 210 215 220

Xaa  
 225 230 235 240

Xaa  
 245 250 255

Xaa Xaa

<210> 230

<211> 214

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD\_RES

<222> (136)..(136)

<223> Any amino acid

<220><221> MOD\_RES

<222> (138)..(141)

<223> Any amino acid

<220><221> SITE

<222> (138)..(141)

<223> This region may encompass 1-4 residues

<220><221> MOD\_RES

<222> (179)..(179)

<223> Any amino acid

<400> 230

Leu Lys Gln Trp Gln Pro Ser Thr Ile Arg Lys Cys Lys Ile Lys Gly

1                    5                    10                    15

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

                  20                    25                    30

Thr Gln Tyr Lys Glu Ser Ile Val Pro His His Glu Pro Gly Gly Gly

                  35                    40                    45

Gly Trp Ser Ile Gln Gln Phe Thr Leu Gly Ala Leu Tyr Glu Glu His

                  50                    55                    60

Leu Lys Leu Arg Asn Trp Trp Thr Lys Ser Asn Asp Gly Leu Pro Leu

65                    70                    75                    80

Val Arg Tyr Leu Gly Cys Thr Ile Lys Leu Tyr Arg Ser Glu Asp Thr

                  85                    90                    95

Asp Tyr Ile Val Thr Tyr Gln Arg Cys Tyr Pro Met Thr Ala Thr Lys

                  100                    105                    110

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

                  115                    120                    125

Lys Ile Ile Val Pro Ser Lys Xaa Thr Xaa Xaa Xaa Xaa Asn Lys Lys

                  130                    135                    140

Lys Lys Pro Tyr Lys Lys Ile Phe Ile Lys Pro Pro Ser Gln Met Gln

145                    150                    155                    160

Asn Lys Trp Tyr Phe Gln Gln Asp Ile Ala Asn Thr Pro Leu Leu Gln

165 170 175  
 Leu Thr Xaa Thr Ala Cys Ser Leu Asp Arg Met Tyr Leu Ser Ser Asp

180 185 190  
 Ser Ile Ser Asn Asn Ile Thr Phe Thr Ser Leu Asn Thr Asn Phe Phe

195 200 205  
 Gln Asn Pro Asn Phe Gln

210

<210> 231

<211> 187

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD\_RES

<222> (1)..(10)

<223> Any amino acid

<220><221> SITE

<222> (1)..(10)

<223> This region may encompass 4-10 residues

<220><221> MOD\_RES

<222> (38)..(45)

<223> Any amino acid

<220><221> SITE

<222> (38)..(45)

<223> This region may encompass 1-8 residues

<220><221> MOD\_RES

<222> (94)..(94)

<223> Any amino acid

<220><221> MOD\_RES

<222> (100)..(102)

<223> Any amino acid

<220><221> SITE

<222> (100)..(102)

<223> This region may encompass 1-3 residues

<220><221> MOD\_RES

<222> (112)..(112)

<223> Any amino acid

<220><221> MOD\_RES

<222> (114)..(115)

<223> Any amino acid

<220><221> SITE

<222

> (114)..(115)

<223> This region may encompass 0-2 residues

<220><221> MOD\_RES

<222> (124)..(139)

<223> Any amino acid

<220><221> SITE

<222> (124)..(139)

<223> This region may encompass 3-16 residues

<220><221> MOD\_RES

<222> (154)..(154)

<223> Any amino acid

<400> 231

Xaa Thr Pro Leu Tyr Phe Glu

1                    5                    10                    15

Cys Arg Tyr Asn Pro Phe Lys Asp Lys Gly Thr Gly Asn Lys Val Tyr

                  20                    25                    30

Leu Val Ser Asn Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Gly Trp

                  35                    40                    45

Asp Pro Pro Thr Asp Pro Asp Leu Ile Ile Glu Gly Phe Pro Leu Trp

                  50                    55                    60

Leu Leu Leu Trp Gly Trp Leu Asp Trp Gln Lys Lys Leu Gly Lys Ile

65                    70                    75                    80

Gln Asn Ile Asp Thr Asp Tyr Ile Leu Val Ile Gln Ser Xaa Tyr Tyr

                  85                    90                    95

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

Ala Leu Thr Gly Pro Gly Thr Pro Lys Leu Pro Asn Gln Lys Ser Ile  
 165 170 175  
 Gln Ala His Met Lys Tyr Lys Phe Tyr Phe Lys  
 180 185

<210> 232

<211> 163

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD\_RES

<222> (34)..(34)

<223> Any amino acid

<220><221> MOD\_RES

<222> (65)..(65)

<223> Any amino acid

<220><221> MOD\_RES

<222> (77)..(78)

<223> Any amino acid

<220><221> MOD\_RES

<222> (86)..(87)

<223> Any amino acid

<220><221> MOD\_RES

<222> (96)..(96)

<223> Any amino acid

<220><221> MOD\_RES  
 <222> (102)..(106)  
 <223> Any amino acid  
 <220><221> SITE  
 <222> (102)..(106)  
 <223> This region may encompass 0-5 residues  
 <220><221> MOD\_RES  
 <222> (125)..(125)  
 <223> Any amino acid  
 <220><221> MOD\_RES  
 <222> (135)..(135)  
 <223> Any amino acid  
 <220><221> MOD\_RES  
 <222> (138)..(163)  
 <223> Any amino acid  
  
 <220><221> SITE  
 <222> (138)..(163)  
 <223> This region may encompass 0-26 residues  
 <400> 232  
 Trp Gly Gly Cys Pro Ala Pro Met Glu Thr Ile Thr Asp Pro Cys Lys  
 1                    5                    10                    15  
 Gln Pro Lys Tyr Pro Ile Pro Asn Asn Leu Leu Gln Thr Thr Ser Leu  
                   20                    25                    30  
 Gln Xaa Pro Thr Thr Pro Ile Glu Thr Tyr Leu Tyr Lys Phe Asp Glu  
                   35                    40                    45  
 Arg Arg Gly Leu Leu Thr Lys Lys Ala Ala Lys Arg Ile Lys Lys Asp  
  
                   50                    55                    60  
 Xaa Thr Thr Glu Thr Thr Leu Phe Thr Asp Thr Gly Xaa Xaa Thr Ser  
 65                    70                    75                    80  
 Thr Thr Leu Pro Thr Xaa Xaa Gln Thr Glu Thr Thr Gln Glu Glu Xaa  
                   85                    90                    95  
 Thr Ser Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Glu Thr Leu Leu Gln Gln  
                   100                    105                    110

Leu Gln Gln Leu Arg Arg Lys Gln Lys Gln Leu Arg Xaa Arg Ile Leu

115 120 125  
 Gln Leu Leu Gln Leu Leu Xaa Leu Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
 130 135 140  
 Xaa  
 145 150 155 160  
 Xaa Xaa Xaa

<210> 233

<211> 203

<212> PRT

<213> Artificial Sequence

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

<220><221

> MOD\_RES

<222> (79)..(79)

<223> Any amino acid

<220><221> MOD\_RES

<222> (104)..(104)

<223> Any amino acid

<220><221> MOD\_RES

<222> (116)..(116)

<223> Any amino acid

<220><221> MOD\_RES

<222> (120)..(121)

<223> Any amino acid

<220><221> MOD\_RES

<222> (125)..(125)

<223> Any amino acid

<220><221> MOD\_RES

<222> (170)..(170)

<223> Any amino acid

<400> 233

Thr Ile Pro Leu Lys Gln Trp Gln Pro Glu Ser Ile Arg Lys Cys Lys  
 1 5 10 15

Ile Lys Gly Tyr Gly Thr Leu Val Leu Gly Ala Glu Gly Arg Gln Phe  
 20 25 30

Tyr Cys Tyr Thr Asn Glu Lys Asp Glu Tyr Thr Pro Pro Lys Ala Pro  
 35 40 45

Gly Gly Gly Gly Phe Gly Val Glu Leu Phe Ser Leu Glu Tyr Leu Tyr  
 50 55 60

Glu Gln Trp Lys Ala Arg Asn Asn Ile Trp Thr Lys Ser Asn Xaa Tyr  
 65 70 75 80

Lys Asp Leu Cys Arg Tyr Thr Gly Cys Lys Ile Thr Phe Tyr Arg His  
 85 90 95

Pro Thr Thr Asp Phe Ile Val Xaa Tyr Ser Arg Gln Pro Pro Phe Glu  
 100 105 110

Ile Asp Lys Xaa Thr Tyr Met Xaa Xaa His Pro Gln Xaa Leu Leu Leu  
 115 120 125

Arg Lys His Lys Lys Ile Ile Leu Ser Lys Ala Thr Asn Pro Lys Gly  
 130 135 140

Lys Leu Lys Lys Lys Ile Lys Ile Lys Pro Pro Lys Gln Met Leu Asn  
 145 150 155 160

Lys Trp Phe Phe Gln Lys Gln Phe Ala Xaa Tyr Gly Leu Val Gln Leu  
 165 170 175

Gln Ala Ala Ala Cys Asx Leu Arg Tyr Pro Arg Leu Gly Cys Cys Asn  
 180 185 190

Glu Asn Arg Leu Ile Thr Leu Tyr Tyr Leu Asn  
 195 200

<210> 234

<211> 162

<212> PRT

<213> Artificial Sequence

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

polypeptide

<220><221> MOD\_RES

<222> (12)..(12)

<223> Any amino acid

<220><221> MOD\_RES

<222> (20)..(20)

<223> Any amino acid

<220><221> MOD\_RES

<222> (23)..(23)

<223> Any amino acid

<220><221> MOD\_RES

<222> (30)..(30)

<223> Any amino acid

<220><221> MOD\_RES

<222> (58)..(58)

<223> I or L

<220><221> MOD\_RES

<222> (84)..(84)

<223> Any amino acid

<220><221> MOD\_RES

<222

> (90)..(90)

<223> Any amino acid

<220><221> MOD\_RES

<222> (95)..(95)

<223> Any amino acid

<220><221> MOD\_RES

<222> (105)..(105)

<223> Any amino acid

<220><221> MOD\_RES

<222> (111)..(111)

<223> I or L

<220><221> MOD\_RES

<222> (113)..(113)

<223> Any amino acid

<220><221> MOD\_RES

<222> (154)..(154)

<223> Any amino acid

<220><221> MOD\_RES

<222> (156)..(156)

<223> Any amino acid

<400> 234

Leu Pro Ile Val Val Ala Arg Tyr Asn Pro Ala Xaa Asp Thr Gly Lys

1                    5                    10                    15

Gly Asn Lys Xaa Trp Leu Xaa Ser Thr Leu Asn Gly Ser Xaa Trp Ala

                  20                    25                    30

Pro Pro Thr Thr Asp Lys Asp Leu Ile Ile Glu Gly Leu Pro Leu Trp

                  35                    40                    45

Leu Ala Leu Tyr Gly Tyr Trp Ser Tyr Xaa Lys Lys Val Lys Lys Asp

                  50                    55                    60

Lys Gly Ile Leu Gln Ser His Met Phe Val Val Lys Ser Pro Ala Ile

65                    70                    75                    80

Gln Pro Leu Xaa Thr Ala Thr Thr Gln Xaa Thr Phe Tyr Pro Xaa Ile

                  85                    90                    95

Asp Asn Ser Phe Ile Gln Gly Lys Xaa Pro Tyr Asp Glu Pro Xaa Thr

                  100                    105                    110

Xaa Asn Gln Lys Lys Leu Trp Tyr Pro Thr Leu Glu His Gln Gln Glu

                  115                    120                    125

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

                  130                    135                    140

Asn Gln Lys Asn Ser Thr Trp Glu Leu Xaa Tyr Xaa Tyr Thr Phe Tyr

145                    150                    155                    160

Phe Lys

<210> 235

<211> 177

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD\_RES

<222> (16)..(16)

<223> Any amino acid

<220><221> MOD\_RES

<222> (26)..(26)

<223> Any amino acid

<220><221> MOD\_RES

<222> (33)..(33)

<223> Any amino acid

<220><221> MOD\_RES

<222> (73)..(73)

<223> Any amino acid

<220><221> MOD\_RES

<222> (81)..(82)

<223> Any amino acid

<220><221> SITE

<222> (81)..(82)

<223> This region may encompass 0-2 residues

<220><221> MOD\_RES

<222> (90)..(90)

<223> Any amino acid

<220><221> MOD\_RES

<222> (94)..(94)

<223> Any amino acid

<220><221> MOD\_RES

<222> (119)..(124)

<223> Any amino acid

<220><221> SITE

<222> (119)..(124)

<223> This region may encompass 1-6 residues

<220><221> MOD\_RES

<222> (168)..(177)

<223> Any amino acid

<

220><221> SITE

<222> (168)..(177)

<223> This region may encompass 1-10 residues

<400> 235

Trp Gly Gly Pro Gln Ile Pro Asp Gln Pro Val Glu Asp Pro Lys Xaa

1                    5                    10                    15

Gln Gly Thr Tyr Pro Val Pro Asp Thr Xaa Gln Gln Thr Ile Gln Ile

                  20                    25                    30

Xaa Asn Pro Leu Lys Gln Lys Pro Glu Thr Met Phe His Asp Trp Asp

                  35                    40                    45

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

                  50                    55                    60

Asn Leu Glu Thr Asp Ser Ser Phe Xaa Ser Asp Ser Glu Glu Thr Pro

65                    70                    75                    80

Xaa Xaa Lys Lys Lys Lys Arg Leu Thr Xaa Glu Leu Pro Xaa Pro Gln

                  85                    90                    95

Glu Glu Thr Glu Glu Ile Gln Ser Cys Leu Leu Ser Leu Cys Glu Glu

                  100                    105                    110

Ser Thr Cys Gln Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa Glu Asn Leu Gln

                  115                    120                    125

Gln Leu Ile His Gln Gln Gln Gln Gln Gln Gln Leu Lys His Asn

                  130                    135                    140

Ile Leu Lys Leu Leu Ser Asp Leu Lys Glx Lys Gln Arg Leu Leu Gln

145                    150                    155                    160

Leu Gln Thr Gly Ile Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

                  165                    170                    175

Xaa

<210> 236

<400> 236

000

<210>

237

<400> 237

000

<210> 238

<400> 238

000

<210> 239

<400> 239

000

<210> 240

<400> 240

000

<210> 241

<400> 241

000

<210> 242

<400> 242

000

<210> 243

<400> 243

000

<210> 244

<400> 244

000

<210> 245

<400> 245

000

<210> 246

<400> 246

000

<210> 247

<400> 247

000

<210> 248  
<400> 248  
000  
<210> 249  
<400> 249  
000  
<210> 250  
<400> 250  
000  
<210> 251  
<400> 251  
000  
<210> 252  
<400> 252  
000  
<210> 253  
<400> 253  
000  
<210> 254  
  
<400> 254  
000  
<210> 255  
<400> 255  
000  
<210> 256  
<400> 256  
000  
<210> 257  
<400> 257  
000  
<210> 258  
<400> 258  
000  
<210> 259

<400> 259  
000  
<210> 260  
<400> 260  
000  
<210> 261  
<400> 261  
000  
<210> 262  
<400> 262  
000  
<210> 263  
<400> 263  
000  
<210> 264  
<400> 264  
000  
<210> 265  
<400> 265  
000  
<210> 266  
<400> 266  
000  
<210> 267  
<400> 267  
000  
<210> 268  
<400> 268  
000  
<210> 269  
<400> 269  
000  
<210> 270  
<400> 270  
000  
<210> 271

<400

> 271

000

<210> 272

<400> 272

000

<210> 273

<400> 273

000

<210> 274

<400> 274

000

<210> 275

<400> 275

000

<210> 276

<400> 276

000

<210> 277

<400> 277

000

<210> 278

<400> 278

000

<210> 279

<400> 279

000

<210> 280

<400> 280

000

<210> 281

<400> 281

000

<210> 282

<400> 282

000

<210> 283

<400> 283

000

<210> 284

<400> 284

000

<210> 285

<400> 285

000

<210> 286

<400> 286

000

<210> 287

<400> 287

000

<210> 288

<400> 288

000

<210> 289

<400> 289

000

<210> 290

<400> 290

000

<210> 291

<400> 291

000

<210> 292

<400> 292

000

<210> 293

<400> 293

000

<210> 294  
<400> 294  
000  
<210> 295  
<400> 295  
000  
<210> 296  
<400> 296  
000  
<210> 297  
<400> 297  
000  
<210> 298  
<400> 298  
000  
<210> 299  
<400> 299  
000  
<210> 300  
<400> 300  
000  
<210> 301  
<400> 301  
000  
<210> 302  
<400> 302  
000  
<210> 303  
<400> 303  
000  
<210> 304  
<400> 304  
000  
<210> 305  
<400> 305

000

<210> 306

<400> 306

000

<210> 307

<400> 307

000

<210> 308

<400> 308

000

<210> 309

<400> 309

000

<210> 310

<400> 310

000

<210> 311

<400> 311

000

<210> 312

<400> 312

000

<210> 313

<400> 313

000

<210> 314

<400> 314

000

<210> 315

<400> 315

000

<210> 316

<400> 316

000

<210> 317  
<400> 317  
000  
<210> 318  
<400> 318  
000  
<210> 319  
<400> 319  
000  
<210> 320  
<400> 320  
000  
<210> 321  
<400> 321  
000  
<210> 322  
<400> 322  
000  
<210  
> 323  
<400> 323  
000  
<210> 324  
<400> 324  
000  
<210> 325  
<400> 325  
000  
<210> 326  
<400> 326  
000  
<210> 327  
<400> 327  
000  
<210> 328

<400> 328  
000  
<210> 329  
<400> 329  
000  
<210> 330  
<400> 330  
000  
<210> 331  
<400> 331  
000  
<210> 332  
<400> 332  
000  
<210> 333  
<400> 333  
000  
<210> 334  
<400> 334  
000  
<210> 335  
<400> 335  
000  
<210> 336  
<400> 336  
000  
<210> 337  
<400> 337  
000  
<210> 338  
<400> 338  
000  
<210> 339  
<400> 339  
000

<210> 340  
  
<400> 340  
000  
<210> 341  
<400> 341  
000  
<210> 342  
<400> 342  
000  
<210> 343  
<400> 343  
000  
<210> 344  
<400> 344  
000  
<210> 345  
<400> 345  
000  
<210> 346  
<400> 346  
000  
<210> 347  
<400> 347  
000  
<210> 348  
<400> 348  
000  
<210> 349  
<400> 349  
000  
<210> 350  
<400> 350  
000  
<210> 351

<400> 351  
000  
<210> 352  
<400> 352  
000  
<210> 353  
<400> 353  
000  
<210> 354  
<400> 354  
000  
<210> 355  
<400> 355  
000  
<210> 356  
<400> 356  
000  
<210> 357  
<400>  
> 357  
000  
<210> 358  
<400> 358  
000  
<210> 359  
<400> 359  
000  
<210> 360  
<400> 360  
000  
<210> 361  
<400> 361  
000  
<210> 362  
<400> 362

000

<210> 363

<400> 363

000

<210> 364

<400> 364

000

<210> 365

<400> 365

000

<210> 366

<400> 366

000

<210> 367

<400> 367

000

<210> 368

<400> 368

000

<210> 369

<400> 369

000

<210> 370

<400> 370

000

<210> 371

<400> 371

000

<210> 372

<400> 372

000

<210> 373

<400> 373

000

<210> 374

<400> 374

000

<210> 375

<400> 375

000

<210> 376

<400> 376

000

<210> 377

<400> 377

000

<210> 378

<400> 378

000

<210> 379

<400> 379

000

<210> 380

<400> 380

000

<210> 381

<400> 381

000

<210> 382

<400> 382

000

<210> 383

<400> 383

000

<210> 384

<400> 384

000

<210> 385

<400> 385

000

<210> 386

<400> 386

000

<210> 387

<400> 387

000

<210> 388

<400> 388

000

<210> 389

<400> 389

000

<210> 390

<400> 390

000

<210> 391

<400> 391

000

<210> 392

<400> 392

000

<210> 393

<400> 393

000

<210> 394

<400> 394

000

<210> 395

<400> 395

000

<210> 396

<400> 396

000

<210> 397  
<400> 397  
000  
<210> 398  
<400> 398  
000  
<210> 399  
<400> 399  
000  
<210> 400  
<400> 400  
000  
<210> 401  
<400> 401  
000  
<210> 402  
<400> 402  
000  
<210> 403  
<400> 403  
000  
<210> 404  
<400> 404  
000  
<210> 405  
<400> 405  
000  
<210> 406  
<400> 406  
000  
<210> 407  
<400> 407  
000  
<210> 408  
<400> 408  
000

<210

> 409

<400> 409

000

<210> 410

<400> 410

000

<210> 411

<400> 411

000

<210> 412

<400> 412

000

<210> 413

<400> 413

000

<210> 414

<400> 414

000

<210> 415

<400> 415

000

<210> 416

<400> 416

000

<210> 417

<400> 417

000

<210> 418

<400> 418

000

<210> 419

<400> 419

000

<210> 420

<400> 420  
000  
<210> 421  
<400> 421  
000  
<210> 422  
<400> 422  
000  
<210> 423  
<400> 423  
000  
<210> 424  
<400> 424  
000  
<210> 425  
<400> 425  
000  
<210> 426  
  
<400> 426  
000  
<210> 427  
<400> 427  
000  
<210> 428  
<400> 428  
000  
<210> 429  
<400> 429  
000  
<210> 430  
<400> 430  
000  
<210> 431  
<400> 431

000

<210> 432

<400> 432

000

<210> 433

<400> 433

000

<210> 434

<400> 434

000

<210> 435

<400> 435

000

<210> 436

<400> 436

000

<210> 437

<400> 437

000

<210> 438

<400> 438

000

<210> 439

<400> 439

000

<210> 440

<400> 440

000

<210> 441

<400> 441

000

<210> 442

<400> 442

000

<210> 443

<400

> 443

000

<210> 444

<400> 444

000

<210> 445

<400> 445

000

<210> 446

<400> 446

000

<210> 447

<400> 447

000

<210> 448

<400> 448

000

<210> 449

<400> 449

000

<210> 450

<400> 450

000

<210> 451

<400> 451

000

<210> 452

<400> 452

000

<210> 453

<400> 453

000

<210> 454

<400> 454

000

<210> 455

<400> 455

000

<210> 456

<400> 456

000

<210> 457

<400> 457

000

<210> 458

<400> 458

000

<210> 459

<400> 459

000

<210> 460

<400> 460

000

<210> 461

<400> 461

000

<210> 462

<400> 462

000

<210> 463

<400> 463

000

<210> 464

<400> 464

000

<210> 465

<400> 465

000

<210> 466  
<400> 466  
000  
<210> 467  
<400> 467  
000  
<210> 468  
<400> 468  
000  
<210> 469  
<400> 469  
000  
<210> 470  
<400> 470  
000  
<210> 471  
<400> 471  
000  
<210> 472  
<400> 472  
000  
<210> 473  
<400> 473  
000  
<210> 474  
<400> 474  
000  
<210> 475  
<400> 475  
000  
<210> 476  
<400> 476  
000  
<210> 477  
<400> 477

000

<210> 478

<400> 478

000

<210> 479

<400> 479

000

<210> 480

<400> 480

000

<210> 481

<400> 481

000

<210> 482

<400> 482

000

<210> 483

<400> 483

000

<210> 484

<400> 484

000

<210> 485

<400> 485

000

<210> 486

<400> 486

000

<210> 487

<400> 487

000

<210> 488

<400> 488

000

<210> 489  
<400> 489  
000  
<210> 490  
<400> 490  
000  
<210> 491  
<400> 491  
000  
<210> 492  
<400> 492  
000  
<210> 493  
<400> 493  
000  
<210> 494  
<400> 494  
000  
<210  
> 495  
<400> 495  
000  
<210> 496  
<400> 496  
000  
<210> 497  
<400> 497  
000  
<210> 498  
<400> 498  
000  
<210> 499  
<400> 499  
000  
<210> 500

<400> 500  
000  
<210> 501  
<400> 501  
000  
<210> 502  
<400> 502  
000  
<210> 503  
<400> 503  
000  
<210> 504  
<400> 504  
000  
<210> 505  
<400> 505  
000  
<210> 506  
<400> 506  
000  
<210> 507  
<400> 507  
000  
<210> 508  
<400> 508  
000  
<210> 509  
<400> 509  
000  
<210> 510  
<400> 510  
000  
<210> 511  
<400> 511  
000

<210> 512  
  
<400> 512  
000  
<210> 513  
<400> 513  
000  
<210> 514  
<400> 514  
000  
<210> 515  
<400> 515  
000  
<210> 516  
<400> 516  
000  
<210> 517  
<400> 517  
000  
<210> 518  
<400> 518  
000  
<210> 519  
<400> 519  
000  
<210> 520  
<400> 520  
000  
<210> 521  
<400> 521  
000  
<210> 522  
<400> 522  
000  
<210> 523

<400> 523  
000  
<210> 524  
<400> 524  
000  
<210> 525  
<400> 525  
000  
<210> 526  
<400> 526  
000  
<210> 527  
<400> 527  
000  
<210> 528  
<400> 528  
000  
<210> 529  
<400  
> 529  
000  
<210> 530  
<400> 530  
000  
<210> 531  
<400> 531  
000  
<210> 532  
<400> 532  
000  
<210> 533  
<400> 533  
000  
<210> 534  
<400> 534

000

<210> 535

<400> 535

000

<210> 536

<400> 536

000

<210> 537

<400> 537

000

<210> 538

<400> 538

000

<210> 539

<400> 539

000

<210> 540

<400> 540

000

<210> 541

<400> 541

000

<210> 542

<400> 542

000

<210> 543

<400> 543

000

<210> 544

<400> 544

000

<210> 545

<400> 545

000

<210> 546

<400> 546

000

<210> 547

<400> 547

000

<210> 548

<400> 548

000

<210> 549

<400> 549

000

<210> 550

<400> 550

000

<210> 551

<400> 551

000

<210> 552

<400> 552

000

<210> 553

<400> 553

000

<210> 554

<400> 554

000

<210> 555

<400> 555

000

<210> 556

<400> 556

000

<210> 557

<400> 557

000

<210> 558

<400> 558

000

<210> 559

<400> 559

000

<210> 560

<400> 560

000

<210> 561

<400> 561

000

<210> 562

<400> 562

000

<210> 563

<400> 563

000

<210> 564

<400> 564

000

<210> 565

<400> 565

000

<210> 566

<400> 566

000

<210> 567

<400> 567

000

<210> 568

<400> 568

000

<210> 569  
<400> 569  
000  
<210> 570  
<400> 570  
000  
<210> 571  
<400> 571  
000  
<210> 572  
<400> 572  
000  
<210> 573  
<400> 573  
000  
<210> 574  
<400> 574  
000  
<210> 575  
<400> 575  
000  
<210> 576  
<400> 576  
000  
<210> 577  
<400> 577  
000  
<210> 578  
<400> 578  
000  
<210> 579  
<400> 579  
000  
<210> 580  
<400> 580  
000

<210

> 581

<400> 581

000

<210> 582

<400> 582

000

<210> 583

<400> 583

000

<210> 584

<400> 584

000

<210> 585

<400> 585

000

<210> 586

<400> 586

000

<210> 587

<400> 587

000

<210> 588

<400> 588

000

<210> 589

<400> 589

000

<210> 590

<400> 590

000

<210> 591

<400> 591

000

<210> 592

<400> 592  
000  
<210> 593  
<400> 593  
000  
<210> 594  
<400> 594  
000  
<210> 595  
<400> 595  
000  
<210> 596  
<400> 596  
000  
<210> 597  
<400> 597  
000  
<210> 598  
  
<400> 598  
000  
<210> 599  
<400> 599  
000  
<210> 600  
<400> 600  
000  
<210> 601  
<400> 601  
000  
<210> 602  
<400> 602  
000  
<210> 603  
<400> 603

000

<210> 604

<400> 604

000

<210> 605

<400> 605

000

<210> 606

<400> 606

000

<210> 607

<400> 607

000

<210> 608

<400> 608

000

<210> 609

<400> 609

000

<210> 610

<400> 610

000

<210> 611

<400> 611

000

<210> 612

<400> 612

000

<210> 613

<400> 613

000

<210> 614

<400> 614

000

<210> 615

<400

> 615

000

<210> 616

<400> 616

000

<210> 617

<400> 617

000

<210> 618

<400> 618

000

<210> 619

<400> 619

000

<210> 620

<400> 620

000

<210> 621

<400> 621

000

<210> 622

<400> 622

000

<210> 623

<400> 623

000

<210> 624

<400> 624

000

<210> 625

<400> 625

000

<210> 626

<400> 626

000

<210> 627

<400> 627

000

<210> 628

<400> 628

000

<210> 629

<400> 629

000

<210> 630

<400> 630

000

<210> 631

<400> 631

000

<210> 632

<400> 632

000

<210> 633

<400> 633

000

<210> 634

<400> 634

000

<210> 635

<400> 635

000

<210> 636

<400> 636

000

<210> 637

<400> 637

000

<210> 638  
<400> 638  
000  
<210> 639  
<400> 639  
000  
<210> 640  
<400> 640  
000  
<210> 641  
<400> 641  
000  
<210> 642  
<400> 642  
000  
<210> 643  
<400> 643  
000  
<210> 644  
<400> 644  
000  
<210> 645  
<400> 645  
000  
<210> 646  
<400> 646  
000  
<210> 647  
<400> 647  
000  
<210> 648  
<400> 648  
000  
<210> 649  
<400> 649

000

<210> 650

<400> 650

000

<210> 651

<400> 651

000

<210> 652

<400> 652

000

<210> 653

<400> 653

000

<210> 654

<400> 654

000

<210> 655

<400> 655

000

<210> 656

<400> 656

000

<210> 657

<400> 657

000

<210> 658

<400> 658

000

<210> 659

<400> 659

000

<210> 660

<400> 660

000

<210> 661  
<400> 661  
000  
<210> 662  
<400> 662  
000  
<210> 663  
<400> 663  
000  
<210> 664  
<400> 664  
000  
<210> 665  
<400> 665  
000  
<210> 666  
<400> 666  
000  
<210  
> 667  
<400> 667  
000  
<210> 668  
<400> 668  
000  
<210> 669  
<400> 669  
000  
<210> 670  
<400> 670  
000  
<210> 671  
<400> 671  
000  
<210> 672

<400> 672  
000  
<210> 673  
<400> 673  
000  
<210> 674  
<400> 674  
000  
<210> 675  
<400> 675  
000  
<210> 676  
<400> 676  
000  
<210> 677  
<400> 677  
000  
<210> 678  
<400> 678  
000  
<210> 679  
<400> 679  
000  
<210> 680  
<400> 680  
000  
<210> 681  
<400> 681  
000  
<210> 682  
<400> 682  
000  
<210> 683  
<400> 683  
000

<210> 684

<400> 684

000

<210> 685

<400> 685

000

<210> 686

<400> 686

000

<210> 687

<400> 687

000

<210> 688

<400> 688

000

<210> 689

<400> 689

000

<210> 690

<400> 690

000

<210> 691

<400> 691

000

<210> 692

<400> 692

000

<210> 693

<400> 693

000

<210> 694

<400> 694

000

<210> 695

<400> 695  
000  
<210> 696  
<400> 696  
000  
<210> 697  
<400> 697  
000  
<210> 698  
<400> 698  
000  
<210> 699  
<400> 699  
000  
<210> 700  
<400> 700  
000  
<210> 701  
<400  
> 701  
000  
<210> 702  
<400> 702  
000  
<210> 703  
<400> 703  
000  
<210> 704  
<400> 704  
000  
<210> 705  
<400> 705  
000  
<210> 706  
<400> 706

000

<210> 707

<400> 707

000

<210> 708

<400> 708

000

<210> 709

<400> 709

000

<210> 710

<400> 710

000

<210> 711

<400> 711

000

<210> 712

<400> 712

000

<210> 713

<400> 713

000

<210> 714

<400> 714

000

<210> 715

<400> 715

000

<210> 716

<400> 716

000

<210> 717

<400> 717

000

<210> 718

<400> 718

000

<210> 719

<400> 719

000

<210> 720

<400> 720

000

<210> 721

<400> 721

000

<210> 722

<400> 722

000

<210> 723

<400> 723

000

<210> 724

<400> 724

000

<210> 725

<400> 725

000

<210> 726

<400> 726

000

<210> 727

<400> 727

000

<210> 728

<400> 728

000

<210> 729

<400> 729

000

<210> 730

<400> 730

000

<210> 731

<400> 731

000

<210> 732

<400> 732

000

<210> 733

<400> 733

000

<210> 734

<400> 734

000

<210> 735

<400> 735

000

<210> 736

<400> 736

000

<210> 737

<400> 737

000

<210> 738

<400> 738

000

<210> 739

<400> 739

000

<210> 740

<400> 740

000

<210> 741  
<400> 741  
000  
<210> 742  
<400> 742  
000  
<210> 743  
<400> 743  
000  
<210> 744  
<400> 744  
000  
<210> 745  
<400> 745  
000  
<210> 746  
<400> 746  
000  
<210> 747  
<400> 747  
000  
<210> 748  
<400> 748  
000  
<210> 749  
<400> 749  
000  
<210> 750  
<400> 750  
000  
<210> 751  
<400> 751  
000  
<210> 752  
<400> 752  
000

<210

> 753

<400> 753

000

<210> 754

<400> 754

000

<210> 755

<400> 755

000

<210> 756

<400> 756

000

<210> 757

<400> 757

000

<210> 758

<400> 758

000

<210> 759

<400> 759

000

<210> 760

<400> 760

000

<210> 761

<400> 761

000

<210> 762

<400> 762

000

<210> 763

<400> 763

000

<210> 764

<400> 764  
000  
<210> 765  
<400> 765  
000  
<210> 766  
<400> 766  
000  
<210> 767  
<400> 767  
000  
<210> 768  
<400> 768  
000  
<210> 769  
<400> 769  
000  
<210> 770  
  
<400> 770  
000  
<210> 771  
<400> 771  
000  
<210> 772  
<400> 772  
000  
<210> 773  
<400> 773  
000  
<210> 774  
<400> 774  
000  
<210> 775  
<400> 775

000

<210> 776

<400> 776

000

<210> 777

<400> 777

000

<210> 778

<400> 778

000

<210> 779

<400> 779

000

<210> 780

<400> 780

000

<210> 781

<400> 781

000

<210> 782

<400> 782

000

<210> 783

<400> 783

000

<210> 784

<400> 784

000

<210> 785

<400> 785

000

<210> 786

<400> 786

000

<210> 787

<400

> 787

000

<210> 788

<400> 788

000

<210> 789

<400> 789

000

<210> 790

<400> 790

000

<210> 791

<400> 791

000

<210> 792

<400> 792

000

<210> 793

<400> 793

000

<210> 794

<400> 794

000

<210> 795

<400> 795

000

<210> 796

<400> 796

000

<210> 797

<400> 797

000

<210> 798

<400> 798

000

<210> 799

<400> 799

000

<210> 800

<400> 800

000

<210> 801

<211> 156

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 801

gcggcggggg ggcggccgcg ttcgcgcgcc gccaccagg ggggtctgcg cgccccccc 60

cgcgcatgcg cggggcccc ccccgggggg gctccgccc cccggcccc ccccgtgcta 120

aaccaccgc gcatgcgca ccacgcccc gcgcgc 156

<210> 802

<211> 150

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 802

ccgagcgta gcgaggagt gcacacctacc cctgggccc acttcttcgg agccgcgcgc 60

tacgccttcg gctgcgcgcg gcacctcaga ccccgcctcg tgctgacacg cttgcgcgtg 120

tcagaccact tcgggctcgc gggggtcggg 150

<210> 803

<211> 122

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 803  
 gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60  
 ccccccccg cgcgatgcgcg gggccccccc ccgcgggggg ctccgcccc cggccccccc 120  
 cg 122

<210> 804

<211> 111

<212> DNA

<213> Artificial Sequence

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

<400> 804  
 cggcccagcg gcggcgcgcg cgcttcgcgc gcgcgccggg gggtccgcc ccccccgcg 60  
 catgcgcggg gcccccccc gcggggggct ccgcccccg gtccccccc g 111

<210> 805

<211> 115

<212> DNA

<213> Artificial Sequence

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

<400> 805  
 cggccgtgcg gcggcgcgcg cgcttcgcgc gcgcgccggg gggtccgcc ccccccgcg 60  
 catgcgcgcg gggccccccc ccgcgggggg ctccgcccc cggccccccc ccccg 115

<210> 806

<211> 104

<212> DNA

<213> Artificial Sequence

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

<400> 806  
 cggcggcggc gcgcgcgcta cgcgcgcgcg ccggggggct gccgcccc ccccgcgat 60  
 gcgcggggcc ccccccgcg gggggctecg cccccggcc cccc 104

<210> 807

<211> 108

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 807

ggcggcgggcg cgcgcgctac gcgcgcgcgc cggggagctc tgccccccc cgcgcatgcg 60

cgcggtccc cccccgcgg ggggctccgc cccccgtcc ccccccg 108

<210> 808

<400> 808

000

<210> 809

<400> 809

000

<210> 810

<400> 810

000

<210> 811

<400> 811

000

<210> 812

<400> 812

000

<210> 813

<400> 813

000

<210> 814

<400> 814

000

<210> 815

<400> 815

000

<210> 816

<400> 816

000

<210> 817  
<400> 817  
000  
<210> 818  
<400> 818  
000  
<210> 819  
<400> 819  
000  
<210> 820  
<400> 820  
000  
<210> 821  
<400> 821  
000  
<210> 822  
<400>  
> 822  
000  
<210> 823  
<400> 823  
000  
<210> 824  
<400> 824  
000  
<210> 825  
<400> 825  
000  
<210> 826  
<400> 826  
000  
<210> 827  
<400> 827  
000  
<210> 828

<400> 828

000

<210> 829

<211> 11

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD\_RES

<222> (4)..(5)

<223> Any amino acid

<220><221> MOD\_RES

<222> (7)..(7)

<223> Any amino acid

<220><221> MOD\_RES

<222> (9)..(10)

<223> Any amino acid

<400> 829

Tyr Asn Pro Xaa Xaa Asp Xaa Gly Xaa Xaa Asn

1

5

10

<210> 830

<400> 830

000

<210> 831

<400> 831

000

<210> 832

<400> 832

000

<210> 833

<400> 833

000

<210> 834

<400> 834

000

<210> 835

<400> 835

000

<210> 836

<400> 836

000

<210> 837

<400> 837

000

<210> 838

<400> 838

000

<210> 839

<400> 839

000

<210> 840

<400> 840

000

<210> 841

<400> 841

000

<210> 842

<400> 842

000

<210> 843

<400> 843

000

<210> 844

<400> 844

000

<210> 845

<400> 845

000

<210> 846  
<400> 846  
000  
<210> 847  
<400> 847  
000  
<210> 848  
<400> 848  
000  
<210> 849  
<400> 849  
000  
<210> 850  
<400> 850  
000  
<210> 851  
<400> 851  
000  
<210> 852  
<400> 852  
000  
<210> 853  
<400> 853  
000  
<210> 854  
<400> 854  
000  
<210> 855  
<400> 855  
000  
<210> 856  
<400> 856  
000  
<210> 857  
<400> 857  
000

<210> 858  
<400> 858  
000  
<210> 859  
<400> 859  
000  
<210> 860  
<400> 860  
000  
  
<210> 861  
<400> 861  
000  
<210> 862  
<400> 862  
000  
<210> 863  
<400> 863  
000  
<210> 864  
<400> 864  
000  
<210> 865  
<400> 865  
000  
<210> 866  
<400> 866  
000  
<210> 867  
<400> 867  
000  
<210> 868  
<400> 868  
000  
<210> 869

<400> 869  
000  
<210> 870  
<400> 870  
000  
<210> 871  
<400> 871  
000  
<210> 872  
<400> 872  
000  
<210> 873  
<400> 873  
000  
<210> 874  
<400> 874  
000  
<210> 875  
<400> 875  
000  
<210> 876  
<400> 876  
000  
<210> 877  
<400> 877  
000  
<210  
> 878  
<400> 878  
000  
<210> 879  
<400> 879  
000  
<210> 880  
<400> 880

000

<210> 881

<400> 881

000

<210> 882

<400> 882

000

<210> 883

<400> 883

000

<210> 884

<400> 884

000

<210> 885

<400> 885

000

<210> 886

<211> 3176

<212> DNA

<213> Gammatorquevirus sp.

<400> 886

taaaatggcg ggagccaatc attttatact ttcactttcc aattaaaaat gcccacgtca	60
caaacaagggt gtggagccat ttaaaactata taactaagtg ggggtggcgaa tggctgagtt	120
taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgccca	180
tgggcgggag ccgaggtgag tgaaaccacc gaggtctagg ggcaattcgg gctagggcag	240
tctagcggaa cgggcaagaa acttaaaaca atatttgttt tacagatggt tagtatatcc	300
tcaagtgatt ttttaagaa aacgaaattt aatgaggaga cgcagaacca agtatggatg	360
tctcaaattg ctgactctca tgataatata tgcagttgct ggcatccatt tgctcacctt	420
cttgcttcca tatttctccc tggccacaaa gatcgtgatc ttactattaa ccaaattctt	480
ctaagagatt ataaagaaaa atgccattct ggtggagaag aaggagaaaa ttctggacca	540
acaacagggt taattacacc aaaagaagaa gatatagaaa aagatggccc agaaggcgcc	600
gcagaagaag accatacaga cgccctgttc gccgccgccg tagaaaactt cgaaaggtaa	660
agagaaaaaa aaaatcttta attgtagac aatggcaacc agacagtata agaacttgta	720

aaattatagg acagtcagct atagttgttg gggctgaagg aaagcaaatg tactgttata 780  
 ctgtcaataa gtaattaat gtgccccaa aacaccata tgggggaggc tttggagtag 840  
 accaatacac actgaaatac ttatatgaag aatacagatt tgcacaaaac atttggacac 900  
 aatctaattg actgaaagac ttatgcagat acataaatgt taagctaata ttctacagag 960  
 acaacaaaac agactttgtc ctttctatg acagaaacc accttttcaa ctaacaaaat 1020  
  
 ttacatacce aggagcacac ccacaacaaa tcatgcttca aaaacaccac aaattcatac 1080  
 tatcacaat gacaaagcct aatggaagac taacaaaaa actcaaaatt aaacctccta 1140  
 aacaaatgct ttctaaatgg ttcttttcaa aacaattctg taaatacct ttactatctc 1200  
 ttaaagcttc tgcactagac cttaggcact cttacctagg ctgctgtaat gaaaatccac 1260  
 aggtattttt ttattattta aaccatggat actacacaat aacaaactgg ggagcacaat 1320  
 cctcaacagc atacagacct aactccaagg tgacagacac aacatactac agatacaaaa 1380  
 atgacagaaa aatatatac attaaaagcc atgaatcga aaaaagtata tcatatgaaa 1440  
  
 acggttattt tcaatctagt ttcttcaaaa cacagtgcac atataccagt gagcgtggtg 1500  
 aagcctgtat agcagaaaaa ccactaggaa tagctattta caatccagta aaagacaatg 1560  
 gagatggtaa taigatatac cttgtaagca ccttagcaaa cacttgggac cagcctcaa 1620  
 aagacagtgc tattttaata caaggagtac ccatatggct aggcttattt ggatatttag 1680  
 actactgtag acaaatataa gctgacaaaa catggctaga cagtcagta ctagtaattc 1740  
 aaagtctgc tatttttact taccctaatc caggagcagg caaatggtat tgtccactat 1800  
 cacaagttt tataaatggc aatggtccgt ttaatcaacc acctacactg ctacaaaaag 1860  
  
 caaagtgggt tccacaata caataccaac aagaaattat taatagcttt gtagaatcag 1920  
 gaccatttgt tcccaaatat gcaaatcaaa ctgaaagcaa ctgggaacta aaatataaat 1980  
 atgtttttac atttaagtgg ggtggaccac aattccatga accagaaatt getgacccta 2040  
 gcaacaaga gcagtatgat gtccccgata ctttctacca aacaataca attgaagatc 2100  
 cagaaggaca agaccccaga tctctcatcc atgattggga ctacagacga ggctttatta 2160  
 aagaaagatc tcttaaaaga atgtcaactt acttctcaac tcatacagat cagcaagcaa 2220  
 cttcagagga agacattccc aaaaagaaaa agagaattgg accccaactc acagtcccac 2280  
  
 aacaaaaaga agaggagaca ctgtcatgtc tctctctct ctgcaaaaaa gataccttcc 2340  
 aagaaacaga gacacaagaa gacctccagc agctcatcaa gcagcagcag gagcagcagc 2400  
 tctctctcaa gagaacatc ctccagctca tccacaaact aaaagagaat caacaaatgc 2460  
 ttcagcttca cacaggcatg ttaccttaac cagatttaaa cctggatttg aagagcaaac 2520  
 agagagagaa ttagcaatta tatttcatag gccccctaga acctacaaag aggaccttcc 2580

attctatccc tggctaccac ctgcaccctc tgtacaatth aaccttaact tcaaaggcta 2640  
 ggccaacaat gtacacttag taaagcatgt ttattaaagc acaaccccca aaataaatgt 2700

aaaaataaaa aaaaaaaaaa aaaaataaaa aattgcaaaa attcggcgct cgcgcgcatg 2760  
 tgcgcctctg gcgcaaatca cgcaacgctc gcgcgcccgc gtatgtctct ttaccacgca 2820  
 cctagattgg ggtgcgcgcg ctagecgcgcg caccccaatg cgccccgccc tcgttccgac 2880  
 ccgcttgccg gggtcggacc acttcgggct cggggggggc gcctgcggc gettttttac 2940  
 taaacagact ccgagccgcc atttggcccc ctaagctccg cccccctcat gaatattcat 3000  
 aaaggaacc acataatag aattgccgac cacaaactgc catatgctaa ttagttcccc 3060  
 ttttacaag taaaagggga agtgaacata gccccacacc cgcaggggca aggccccgca 3120

ccctacgtc actaaccacg cccccgccgc catcttgggt gcggcagggc gggggc 3176

<210> 887  
 <211> 124  
 <212> PRT  
 <213> Gammatorquevirus sp.  
 <400> 887

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

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

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg  
 115 120

<210> 888

<211> 271

<212> PRT

<213> Gammatorquevirus sp.

<400> 888

Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn  
1                   5                   10                   15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His  
                  20                   25                   30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser  
                  35                   40                   45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile  
                  50                   55                   60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly  
65                   70                   75                   80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp  
                  85                   90                   95

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

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Ser Gly Val Asp His  
                  115                   120                   125

Asn Ser Met Asn Gln Lys Leu Leu Thr Leu Ala Asn Lys Ser Ser Met  
                  130                   135                   140

Met Ser Pro Ile Leu Ser Thr Lys Gln Tyr Lys Leu Lys Ile Gln Lys  
145                   150                   155                   160

Asp Lys Thr Pro Asp Leu Ser Ser Met Ile Gly Thr Thr Asp Glu Ala  
                  165                   170                   175

Leu Leu Lys Lys Asp Leu Leu Lys Glu Cys Gln Leu Thr Ser Gln Leu  
                  180                   185                   190

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

Arg Glu Leu Asp Pro Asn Ser Gln Ser His Asn Lys Lys Lys Arg Arg

210 215 220  
 His Cys His Val Ser Ser Leu Ser Ala Lys Lys Ile Pro Ser Lys Lys  
 225 230 235 240

Gln Arg His Lys Lys Thr Ser Ser Ser Ser Ser Ser Ser Ser Arg Ser  
 245 250 255  
 Ser Ser Ser Ser Ser Arg Glu Thr Ser Ser Ser Ser Ser Thr Asn  
 260 265 270

<210> 889

<211> 267

<212> PRT

<213> Gammatorquevirus sp.

<400> 889

Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn  
 1 5 10 15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His

20 25 30  
 Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser

35 40 45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile  
 50 55 60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly  
 65 70 75 80

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

85 90 95  
 Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp

100 105 110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg Ser Ala Ser Asn  
 115 120 125

Phe Arg Gly Arg His Ser Gln Lys Glu Lys Glu Asn Trp Thr Pro Thr  
 130 135 140

His Ser Pro Thr Thr Lys Arg Arg Gly Asp Thr Val Met Ser Pro Leu

145                    150                    155                    160  
 Ser Leu Gln Lys Arg Tyr Leu Pro Arg Asn Arg Asp Thr Arg Arg Pro  
                          165                    170                    175  
 Pro Ala Ala His Gln Ala Ala Ala Gly Ala Ala Ala Pro Pro Gln Glu  
                          180                    185                    190  
 Lys His Pro Pro Ala His Pro Gln Thr Lys Arg Glu Ser Thr Asn Ala  
                          195                    200                    205  
 Ser Ala Ser His Arg His Val Thr Leu Thr Arg Phe Lys Pro Gly Phe

                         210                    215                    220  
 Glu Glu Gln Thr Glu Arg Glu Leu Ala Ile Ile Phe His Arg Pro Pro  
 225                    230                    235                    240  
 Arg Thr Tyr Lys Glu Asp Leu Pro Phe Tyr Pro Trp Leu Pro Pro Ala  
                          245                    250                    255  
 Pro Leu Val Gln Phe Asn Leu Asn Phe Lys Gly  
                          260                    265

<210> 890

<211> 50

<212> PRT

<213> Gammatorquevirus sp.

<400> 890

Met Arg Arg Arg Arg Thr Lys Tyr Gly Cys Leu Lys Leu Leu Thr Leu

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

50

<210> 891

<211> 662

<212> PRT

<213> Gammatorquevirus sp.

<400> 891

Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg  
 1 5 10 15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg  
 20 25 30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg  
 35 40 45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln  
 50 55 60

Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala  
 65 70 75 80

Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn  
 85 90 95

Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly  
 100 105 110

Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala  
 115 120 125

Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr  
 130 135 140

Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val  
 145 150 155 160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr  
 165 170 175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe  
 180 185 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu  
 195 200 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys  
 210 215 220

Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp  
 225 230 235 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe

	245		250		255
Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala					
	260		265		270
Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr					
	275		280		285
Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His					
	290		295		300
Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser					
305		310		315	320
Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys					
	325		330		335
Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp					
	340		345		350
Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr					
	355		360		365
Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro					
	370		375		380
Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys					
385		390		395	400
Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro					
	405		410		415
Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro					
	420		425		430
Leu Ser Gln Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro					
	435		440		445
Thr Leu Leu Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln					
	450		455		460
Glu Ile Ile Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr					
465		470		475	480
Ala Asn Gln Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe					
	485		490		495

Thr Phe Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp  
                   500                                  505                                  510  
 Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr  
                   515                                  520                                  525  
  
 Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His  
           530                                  535                                  540  
 Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg  
 545                                  550                                  555                                  560  
 Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu  
                                   565                                  570                                  575  
 Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val  
                   580                                  585                                  590  
  
 Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys  
                   595                                  600                                  605  
 Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln  
           610                                  615                                  620  
 Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile  
 625                                  630                                  635                                  640  
 Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu  
                                   645                                  650                                  655  
  
 His Thr Gly Met Leu Pro  
                   660  
  
 <210> 892  
 <211> 215  
 <212> PRT  
 <213> Gammatorquevirus sp.  
 <400> 892  
  
 Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg  
 1                  5                                  10                                  15  
 Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg  
                   20                                  25                                  30  
 Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg



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

<210> 894

<400> 894

000

<210> 895

<400> 895

000

<210> 896

<400> 896

000

<210> 897

<400> 897

000

<210> 898

<400> 898

000

<210> 899

<400> 899

000

<210> 900  
<400> 900  
000  
<210> 901  
<400> 901  
000  
<210> 902  
<400> 902  
000  
<210> 903  
<400> 903  
000  
<210> 904  
<400> 904  
000  
<210> 905  
<400> 905  
000  
<210> 906  
<400> 906  
000  
<210> 907  
<400> 907  
000  
<210> 908  
<400> 908  
000  
<210> 909  
<400> 909  
000  
<210> 910  
<400> 910  
000  
<210> 911  
<400> 911  
000

<210> 912  
<400> 912  
000  
<210> 913  
<400> 913  
000  
<210> 914  
<400>  
> 914  
000  
<210> 915  
<400> 915  
000  
<210> 916  
<400> 916  
000  
<210> 917  
<400> 917  
000  
<210> 918  
<400> 918  
000  
<210> 919  
<400> 919  
000  
<210> 920  
<400> 920  
000  
<210> 921  
<400> 921  
000  
<210> 922  
<400> 922  
000  
<210> 923

<400> 923

000

<210> 924

<400> 924

000

<210> 925

<211> 662

<212> PRT

<213> Gammatorquevirus sp.

<400> 925

Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg  
 1                    5                    10                    15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg  
                   20                    25                    30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg  
                   35                    40                    45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln  
                   50                    55                    60

Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala  
 65                    70                    75                    80

Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn  
                   85                    90                    95

Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly  
                   100                    105                    110

Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala  
                   115                    120                    125

Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr  
                   130                    135                    140

Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val  
 145                    150                    155                    160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr  
                   165                    170                    175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe  
 180 185 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu  
 195 200 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys  
 210 215 220

Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp  
 225 230 235 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe  
 245 250 255

Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala  
 260 265 270

Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr  
 275 280 285

Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His  
 290 295 300

Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser  
 305 310 315 320

Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys  
 325 330 335

Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp  
 340 345 350

Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr  
 355 360 365

Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro  
 370 375 380

Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys  
 385 390 395 400

Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro  
 405 410 415

Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro



<210> 926

<211> 58

<212> PRT

<213> Gammatorquevirus sp.

<400> 926

Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg  
 1                    5                    10                    15  
 Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg  
                   20                    25                    30  
 Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg  
                   35                    40                    45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys  
                   50                    55

<210> 927

<211> 202

<212> PRT

<213> Gammatorquevirus sp.

<400> 927

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

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

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

Lys Ala Ser Ala Leu Asp Leu Arg His Ser Tyr Leu Gly Cys Cys Asn  
 180 185 190  
 Glu Asn Pro Gln Val Phe Phe Tyr Tyr Leu  
 195 200

<210> 928

<211> 79

<212> PRT

<213> Gammatorquevirus sp.

<400> 928

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

Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His Glu Tyr Glu Lys  
 35 40 45  
 Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser Phe Leu Gln Thr  
 50 55 60  
 Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys Ile Ala Glu  
 65 70 75

<210> 929

<211> 160

<212> PRT

<213> Gammatorquevirus sp.

<400> 929

Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp Asn Gly Asp

1                    5                    10                    15

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

20                    25                    30

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

35                    40                    45

Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys Ala Asp Lys

50                    55                    60

Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro Ala Ile Phe

65                    70                    75                    80

Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro Leu Ser Gln

85                    90                    95

Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro Thr Leu Leu

100                    105                    110

Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln Glu Ile Ile

115                    120                    125

Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr Ala Asn Gln

130                    135                    140

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

145                    150                    155                    160

<210> 930

<211> 163

<212> PRT

<213> Gammatorquevirus sp.

<400> 930

Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp Pro Ser Lys

1                    5                    10                    15

Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr Ile Gln Ile

20                    25                    30

Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His Asp Trp Asp

35                    40                    45

Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg Met Ser Thr  
 50 55 60  
 Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu Glu Asp Ile  
 65 70 75 80  
 Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val Pro Gln Gln  
 85 90 95  
  
 Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys Lys Lys Asp  
 100 105 110  
 Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln Leu Ile Lys  
 115 120 125  
 Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile Leu Gln Leu  
 130 135 140  
 Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu His Thr Gly  
 145 150 155 160

Met Leu Pro

<210> 931

<400> 931

000

<210> 932

<400> 932

000

<210> 933

<400> 933

000

<210> 934

<400> 934

000

<210> 935

<400> 935

000

<210> 936

<400> 936

000

<210> 937

<400> 937

000

<210> 938

<400> 938

000

<210> 939

<400> 939

000

<210> 940

<400> 940

000

<210> 941

<400> 941

000

<210> 942

<400> 942

000

<210> 943

<400> 943

000

<210> 944

<400> 944

000

<210> 945

<400> 945

000

<210> 946

<400> 946

000

<210

> 947

<400> 947

000

<210> 948  
 <400> 948  
 000  
 <210> 949  
 <211> 21  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
           peptide  
 <220><221> MOD\_RES  
 <222> (1)..(1)  
 <223> W or F  
 <220><221> MOD\_RES  
 <222> (2)..(8)  
 <223> Any amino acid  
 <220><221> MOD\_RES  
 <222> (10)..(12)  
 <223> Any amino acid  
 <220><221> MOD\_RES  
 <222> (14)..(14)  
 <223> Any amino acid  
 <220><221> MOD\_RES  
 <222> (16)..(20)  
 <223> Any amino acid  
 <400> 949  
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Cys Xaa Cys Xaa

1                  5                  10                  15  
 Xaa Xaa Xaa Xaa His  
                   20

<210> 950  
 <211> 22  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic

peptide

<220><221> MOD\_RES

<222> (6)..(7)

<223> Any amino acid

<220><221> MOD\_RES

<222> (9)..(9)

<223> Any amino acid

<220><221> MOD\_RES

<222> (15)..(16)

<223> Any amino acid

<400> 950

Tyr Asn Cys Ser Pro Xaa Xaa Asp Xaa Gly Ala Ser Lys Arg Xaa Xaa

1                    5                    10                    15

Asn Thr Ser Val Ala Lys

20

<210> 951

<211> 51

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified\_base

<222> (45)..(45)

<223> a, c, t, g, unknown or other

<400> 951

aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agggncagtc t 51

<210> 952

<211> 50

<212> DNA

<213> Alphatorquevirus sp.

<400> 952

aggtgagtgtt acacaccgca gtcaaggggc aattcgggct cgggactggc 50

<210> 953

<211> 50  
 <212> DNA  
 <213> Betatorquevirus sp.  
 <400> 953  
 aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agatcagtct 50  
 <210> 954  
 <211> 50  
 <212> DNA  
 <213> Gammatorquevirus sp.  
 <400> 954  
 aggtgagtga aaccaccgag gtctaggggc aattcgggct agggcagtct 50  
 <210> 955  
 <211> 6  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
 6xHis tag  
  
 <400> 955  
 His His His His His His  
 1 5  
 <210> 956  
 <211> 237  
 <212> PRT  
 <213> Unknown  
 <220><223> Description of Unknown:  
 Anelloviridae family sequence  
 <400> 956  
 Arg Arg Lys Leu Arg Val Arg Lys Arg Phe Tyr Lys Arg Lys Leu Lys  
 1 5 10 15  
 Lys Ile Val Leu Lys Gln Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr  
 20 25 30  
 Ile Phe Gly Thr Ile Cys Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn  
 35 40 45

Asn Asn Tyr Ile Gln Thr Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro  
 50 55 60  
 Gly Gly Gly Gly Trp Thr Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp  
 65 70 75 80  
 Glu Asp Trp Glu His Leu Lys Asn Val Trp Thr Gln Ser Asn Ala Gly  
 85 90 95  
 Leu Pro Leu Val Arg Tyr Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser  
 100 105 110  
 Ala Tyr Thr Asp Tyr Ile Ala Gln Val Phe Asn Cys Tyr Pro Met Thr  
 115 120 125  
 Asp Thr Lys Tyr Thr His Ala Asp Ser Ala Pro Asn Arg Met Leu Leu  
 130 135 140  
 Lys Lys His Val Ile Arg Val Pro Ser Arg Glu Thr Arg Lys Lys Arg  
 145 150 155 160  
 Lys Pro Tyr Lys Arg Val Arg Val Gly Pro Pro Ser Gln Met Gln Asn  
 165 170 175  
 Lys Trp Tyr Phe Gln Arg Asp Ile Cys Glu Ile Pro Leu Ile Met Ile  
 180 185 190  
 Ala Ala Thr Ala Val Asp Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys  
 195 200 205  
 Ala Ser Asn Asn Leu Thr Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln  
 210 215 220  
 Asn Gln Asp Phe Asp His Pro Ser Asp Thr Gln Gly Tyr  
 225 230 235  
 <210> 957  
 <211> 232  
 <212> PRT  
 <213> Unknown  
 <220><223> Description of Unknown:  
 Anelloviridae family sequence  
 <400> 957  
 Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys

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

<210> 958

<211> 238

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Anelloviridae family sequence

<400> 958

Arg Arg Arg Arg Arg Trp Arg Lys Arg Pro Thr Val Arg Arg Lys Leu  
1                   5                   10                   15

Lys Lys Leu Thr Ile Gln Gln Trp Gln Pro Lys Thr Ile Arg Lys Cys  
                  20                   25                   30

Cys Ile Gln Gly Leu His Cys Leu Phe Leu Val Thr Glu Asp Thr Ile  
                  35                   40                   45

Ser Arg Asn Tyr Arg Met Tyr Glu His Ser Tyr Thr Gly Glu Tyr Tyr  
                  50                   55                   60

Pro Gly Gly Gly Gly Phe Ser Ile Thr Arg Tyr Ser Leu Asp Gly Leu  
65                   70                   75                   80

Tyr Glu Gln His Gln Leu Asp Arg Asn Trp Trp Thr Asn Ser Asn Thr  
                  85                   90                   95

Asn Leu Pro Leu Val Arg Tyr Thr Gly Cys Lys Ile Lys Phe Tyr Gln  
                  100                   105                   110

Ser Trp Ser Val Asp Tyr Ile Cys Asn Tyr Ser Leu Thr Trp Pro Met  
                  115                   120                   125

Val Ala Thr Gln Leu Leu Tyr Gln Ser Cys Gln Pro Ser Phe Met Met  
                  130                   135                   140

Met Asn Lys Asn Ser Ile Met Ile Pro Ser Lys Leu Thr Lys Pro Ile  
145                   150                   155                   160

Lys Lys Gly Tyr Lys Thr Ile Lys Glu Lys Pro Pro His Glu Met Leu  
                  165                   170                   175

Asn Arg Trp Tyr Phe Ala Lys Asp Leu Ser Lys Val Gly Leu Leu Met  
                  180                   185                   190

Leu Thr Ala Ala Ser Ala Ser Phe Asp His Tyr Tyr Gln Ala Thr Asp  
                  195                   200                   205

Ser Leu Ser Asn Asn Cys Thr Phe Glu Ser Leu Asn Pro Tyr Phe Tyr

210 215 220  
Met Arg His Asp Phe Ile Leu Phe Pro Val Thr Gly Tyr Val  
225 230 235

<210> 959  
<211> 238  
<212> PRT  
<213> Unknown  
<220><223> Description of Unknown:  
Anelloviridae family sequence

<400> 959  
Arg Arg Arg Arg Arg Trp Val Arg Arg Lys Pro Phe Tyr Lys Arg Lys  
1 5 10 15  
Ile Lys Arg Leu Asn Ile Val Glu Trp Gln Pro Lys Ser Ile Arg Lys  
20 25 30  
Cys Arg Ile Lys Gly Met Leu Cys Leu Phe Gln Thr Thr Glu Asp Arg  
35 40 45  
Leu Ser Tyr Asn Phe Asp Met Tyr Glu Glu Ser Ile Ile Pro Glu Lys  
50 55 60  
Leu Pro Gly Gly Gly Gly Phe Ser Ile Lys Asn Ile Ser Leu Tyr Ala  
65 70 75 80  
Leu Tyr Gln Glu His Ile His Ala His Asn Ile Phe Thr His Thr Asn  
85 90 95  
Thr Asp Arg Pro Leu Ala Arg Tyr Thr Gly Cys Ser Leu Lys Phe Tyr  
100 105 110  
Gln Ser Lys Asp Ile Asp Tyr Val Val Thr Tyr Ser Thr Ser Leu Pro  
115 120 125  
Leu Arg Ser Ser Met Gly Met Tyr Asn Ser Met Gln Pro Ser Ile His  
130 135 140  
Leu Met Gln Gln Asn Lys Leu Ile Val Pro Ser Lys Gln Thr Gln Lys  
145 150 155 160  
Arg Arg Lys Pro Tyr Ile Lys Lys His Ile Ser Pro Pro Thr Gln Met

165 170 175

Lys Ser Gln Trp Tyr Phe Gln His Asn Ile Ala Asn Ile Pro Leu Leu

180 185 190

Met Ile Arg Thr Thr Ala Leu Thr Leu Asp Asn Tyr Tyr Ile Gly Ser

195 200 205

Arg Gln Leu Ser Thr Asn Val Thr Ile His Thr Leu Asn Thr Thr Tyr

210 215 220

Ile Gln Asn Arg Asp Trp Gly Asp Arg Asn Lys Thr Tyr Tyr

225 230 235

<210> 960

<211> 240

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Anelloviridae family sequence

<400> 960

Arg Arg Pro Val Arg Arg Arg Arg Arg Lys Leu Arg Lys Val Lys Arg

1 5 10 15

Lys Lys Lys Ser Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Arg

20 25 30

Thr Cys Lys Ile Ile Gly Gln Ser Ala Ile Val Val Gly Ala Glu Gly

35 40 45

Lys Gln Met Tyr Cys Tyr Thr Val Asn Lys Leu Ile Asn Val Pro Pro

50 55 60

Lys Thr Pro Tyr Gly Gly Gly Phe Gly Val Asp Gln Tyr Thr Leu Lys

65 70 75 80

Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala Gln Asn Ile Trp Thr Gln Ser

85 90 95

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

100 105 110

Tyr Arg Asp Asn Lys Thr Asp Phe Val Leu Ser Tyr Asp Arg Asn Pro



Trp Leu Tyr Gln Glu Tyr Thr Ala His Arg Asn Ile Trp Thr Lys Thr  
 85 90 95

Asn Glu Tyr Thr Asp Leu Cys Arg Tyr Thr Gly Ala Gln Ile Ile Leu  
 100 105 110

Tyr Arg His Pro Asp Val Asp Phe Ile Val Ser Trp Asp Asn Gln Pro  
 115 120 125

Pro Phe Leu Leu Asn Lys Tyr Thr Tyr Pro Glu Leu Gln Pro Gln Asn  
 130 135 140

Leu Leu Leu Ala Arg Arg Lys Arg Ile Ile Leu Ser Gln Lys Ser Asn  
 145 150 155 160

Pro Lys Gly Lys Leu Arg Ile Lys Leu Arg Ile Pro Pro Pro Lys Gln  
 165 170 175

Met Ile Thr Lys Trp Phe Phe Gln Arg Asp Phe Cys Asp Val Asn Leu  
 180 185 190

Phe Lys Leu Cys Ala Ser Ala Ala Ser Phe Arg Tyr Pro Gly Ile Ser  
 195 200 205

His Gly Ala Gln Ser Thr Ile Phe Ser Ala Tyr Ala Leu Asn Thr Asp  
 210 215 220

Phe Tyr Gln Cys Ser Asp Trp Cys Gln Thr Asn Thr Glu Thr Gly Tyr  
 225 230 235 240

<210> 962

<211> 239

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Anelloviridae family sequence

<400> 962

Gly Arg Arg Thr Tyr Thr Arg Arg Ala Val Arg Arg Arg Arg Arg Pro  
 1 5 10 15

Arg Lys Arg Leu Val Leu Thr Gln Trp Ser Pro Gln Thr Val Arg Asn  
 20 25 30

Cys Ser Ile Arg Gly Ile Val Pro Met Val Ile Cys Gly His Thr Lys



<400> 963

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg  
 1                    5                    10                    15  
 His Lys Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg  
                   20                    25                    30  
 His Cys Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly  
                   35                    40                    45

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

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

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

<210> 964

<211> 233

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Anelloviridae family sequence

<400> 964

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

1                    5                    10                    15

Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys Arg Ile Lys Gly

                  20                    25                    30

Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe Ala Thr Asn Phe

                  35                    40                    45

Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro Phe Gly Gly Gly

                  50                    55                    60

His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe Glu Glu His Leu

65                    70                    75                    80

Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn Leu Glu Leu Thr

                  85                    90                    95

Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His Pro Asp Gln Asp

                  100                    105                    110

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

                  115                    120                    125

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

                  130                    135                    140

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

145                    150                    155                    160

Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp Lys Trp Tyr Phe

                  165                    170                    175

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

                  180                    185                    190

Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr Asp Asn Thr Cys



Lys Ile Lys Val Gly Ala Pro Arg Met Phe Thr Asp Lys Trp Tyr Pro  
 165 170 175  
 Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Phe Ala Ser Ala  
 180 185 190  
 Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp Thr Ile Val  
 195 200 205  
 Val Ser Phe Gln Val Leu Gln Ser Met Tyr Asn Asp Cys Leu Ser Val  
 210 215 220  
 Leu Pro Asp Asn Phe Ala Glu Thr Ser  
 225 230  
 <210> 966  
 <211> 232  
 <212> PRT  
 <213> Unknown  
 <220><223> Description of Unknown:  
 Anelloviridae family sequence  
 <400> 966  
 Arg Arg Trp Lys Arg Lys Gly Arg Arg Arg Arg Lys Ala Lys Ile Ile  
 1 5 10 15  
 Ile Arg Gln Trp Gln Pro Asn Tyr Arg Arg Arg Cys Asn Ile Val Gly  
 20 25 30  
 Tyr Leu Pro Ile Leu Ile Cys Gly Gly Asn Thr Val Ser Arg Asn Tyr  
 35 40 45  
 Ala Thr His Ser Asp Asp Thr Asn Tyr Pro Gly Pro Phe Gly Gly Gly  
 50 55 60  
 Met Thr Thr Asp Lys Phe Ser Leu Arg Ile Leu Tyr Asp Glu Tyr Lys  
 65 70 75 80  
 Arg Phe Met Asn Tyr Trp Thr Ala Ser Asn Glu Asp Leu Asp Leu Cys  
 85 90 95  
 Arg Tyr Leu Gly Cys Thr Phe Tyr Phe Phe Arg His Pro Glu Val Asp  
 100 105 110  
 Phe Ile Ile Lys Ile Asn Thr Met Pro Pro Phe Leu Asp Thr Thr Ile



85 90 95  
 Pro Leu Ala Pro Phe Asp Gly Ala Lys Lys Trp Phe Val Ser Arg Gly  
 100 105 110  
 Phe Lys Arg Leu Leu Arg Pro Lys Pro Gln Ile Thr Ile Glu Gly Pro  
 115 120 125  
 Asn Ser Ala Gly Thr Lys Val Arg His Tyr Gly Ile Ala Phe Ser Phe  
 130 135 140  
  
 Pro Gln Pro Glu Gln Thr Val Thr Lys Leu Thr Leu Tyr Val Gln Phe  
 145 150 155 160  
 Arg Gln Phe Ala Pro Asn Asn Pro Ser Thr  
 165 170  
  
 <210> 968  
 <211> 179  
 <212> PRT  
 <213> Orthohepevirus A  
 <400> 968  
 Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr  
 1 5 10 15  
 Ser Ser Val Ala Ser Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu  
 20 25 30  
  
 Asn Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala  
 35 40 45  
 Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Val Arg Ala Thr Ile  
 50 55 60  
 Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Ser Ile Ser Phe Trp Pro  
 65 70 75 80  
 Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met Asn Ser Ile Thr Ser  
 85 90 95  
  
 Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile Ala Ser Glu Leu Val  
 100 105 110  
 Ile Pro Ser Glu Arg Leu His Tyr Arg Thr Arg Thr Gly Val Ala Glu  
 115 120 125

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