(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0146560
(43) 공개일자 2023년10월19일

(51) 국제특허분류(Int. Cl.)
C12N 15/86 (2006.01) *A61K 48/00* (2006.01)
C07K 14/005 (2006.01) *C07K 14/61* (2006.01)
(52) CPC특허분류
C12N 15/86 (2013.01)
A61K 48/005 (2013.01)
(21) 출원번호 10-2023-7030345
(22) 출원일자(국제) 2022년02월07일
심사청구일자 없음
(85) 번역문제출일자 2023년09월06일
(86) 국제출원번호 PCT/US2022/015499
(87) 국제공개번호 WO 2022/170195
국제공개일자 2022년08월11일
(30) 우선권주장
63/147,102 2021년02월08일 미국(US)

(71) 출원인
플래그쉽 파이어니어링 이노베이션스 브이, 인크.
미국 02142 매사추세츠주 캠브리지 8쓰 플로어 캠브리지 파크웨이 55
(72) 발명자
델라그레이브, 사이먼
미국 02142 매사추세츠 캠브리지 8쓰 플로어 캠브리지 파크웨이 55
레보, 케빈 제임스
미국 02142 매사추세츠 캠브리지 8쓰 플로어 캠브리지 파크웨이 55
(뒷면에 계속)
(74) 대리인
양영준, 이상남

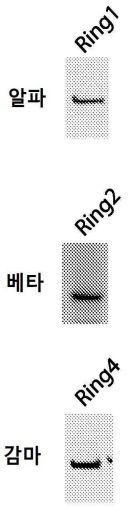
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **하이브리드 AAV-아넬로벡터**

(57) 요 약

본 발명은 일반적으로 아넬로벡터를 제조하고 투여하기 위한 조성물 및 이의 용도에 관한 것이다.

대 표 도 - 도1



(52) CPC특허분류
C07K 14/005 (2013.01)
C07K 14/61 (2013.01)
C12N 2750/00022 (2013.01)
C12N 2750/00043 (2013.01)
C12N 2750/00044 (2013.01)
C12N 2750/14122 (2013.01)
C12N 2750/14143 (2013.01)
C12N 2750/14144 (2013.01)
(72) 발명자
디비아시오-화이트, 마이클 제임스
미국 02142 매사추세츠 캠브리지 8쓰 플로어 캠브리지 파크웨이 55
나완다, 다난자이 마니크랄
미국 02142 매사추세츠 캠브리지 8쓰 플로어 캠브리지 파크웨이 55

명 세 서

청구범위

청구항 1

바이러스 입자로서,

(i) AAV 복제 원점, (ii) 치료용 RNA 또는 폴리펩티드를 인코딩하는 서열에 작동 가능하게 연결된 프로모터, 및 (iii) 아넬로바이러스 ORF1 분자에 결합하는 서열을 포함하는 환형 DNA를 포함하되, 환형 DNA는 아넬로바이러스 ORF1 분자를 포함하는 캡시드에 의해 캡슐화된, 바이러스 입자.

청구항 2

벡터로서,

a) 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부; 및

b) 비-아넬로바이러스 복제 원점을 포함하는 유전 요소

를 포함하며, 선택적으로 유전 요소는 (i) 외인성 이펙터를 인코딩하는 핵산 서열, 및/또는 (ii) 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소를 추가로 포함하는, 벡터.

청구항 3

유전 요소로서,

아넬로바이러스 ORF1 분자(예를 들어, 5' UTR)에 특이적으로 결합하는 단백질 결합 서열; 및

예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점;

선택적으로, 외인성 이펙터(예를 들어, 치료용 외인성 이펙터)를 인코딩하는 핵산 서열; 및

선택적으로, 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소

를 포함하는, 유전 요소.

청구항 4

시스템으로서,

a) 유전 요소 또는 유전 요소 작제물인 제1 핵산으로서,

예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점;

선택적으로, 외인성 이펙터(예를 들어, 치료용 외인성 이펙터)를 인코딩하는 핵산 서열; 및

선택적으로, 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소

를 포함하는, 제1 핵산;

b) 아넬로바이러스 ORF1 분자를 인코딩하는 제2 핵산

을 포함하는 시스템.

청구항 5

제2항의 벡터를 세포 내로 도입하는 단계를 포함하는, 외인성 이펙터를 표적 세포(예를 들어, 척추동물 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포)로 운반하는 방법.

청구항 6

제2항의 벡터를 대상체 내로 도입하는 단계를 포함하는, 질환 또는 장애의 치료 또는 예방을 필요로 하는 대상

체에서 질환 또는 장애를 치료 또는 예방하는 방법.

청구항 7

치료용 조성물을 제조하는 방법으로서,

(a) (i) AAV 복제 원점,

(ii) 치료용 이펙터(예를 들어, 치료용 RNA 또는 폴리펩티드)를 인코딩하는 서열에 작동 가능하게 연결된 프로모터,

(iii) 아넬로바이러스 ORF1 분자를 인코딩하는 서열,

(iv) 선택적으로 아넬로바이러스 ORF2 분자를 인코딩하는 서열,

(v) 선택적으로 AAV REP2 서열을 인코딩하는 서열

(vi) 선택적으로 하나 또는 복수의 헬퍼 단백질, 예를 들어, 아데노바이러스 헬퍼 단백질, 예를 들어, E2A 분자, 아데노바이러스 E4 분자, 및/또는 아데노바이러스 VARNA 분자를 인코딩하는 서열

을 포함하는 외인성 DNA를 포함하는 하나 또는 복수의 숙주 세포를 제공하는 단계;

(b) 유전 요소를 봉입(예를 들어, 캡시드화)하기에 충분한 수의 ORF1 분자를 포함하는 단백질성 외부(예를 들어, 캡시드)를 포함하는 벡터(예를 들어, 아넬로벡터, 예를 들어, 바이러스 입자)의 형성에 적합한 조건 하에서 하나 또는 복수의 숙주 세포를 배양하는 단계;

(c) 세포 배양물로부터 단계 (b)에서 생성된 벡터를 정제하는 단계

를 포함하며, 이를 통해 치료용 조성물을 제조하는, 방법.

발명의 설명

기 술 분 야

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 2021년 2월 8일에 출원된 미국 가출원 63/147,102의 이익을 주장한다. 전술한 출원의 내용은 본 명세서에 전문이 참조로 포함된다.

배 경 기 술

[0003] 환자에게 치료용 이펙터를 운반하기에 적합한 바이러스 벡터를 제조하기 위한 조성물 및 방법을 개발하는 것에 대한 필요성이 계속 존재한다.

발명의 내용

[0004] 본 개시내용은 진핵 세포(예를 들어, 인간 세포 또는 인간 조직의 세포)에, 예를 들어, 유전 물질을 운반하기 위한, 이펙터, 예를 들어, 페이로드(payload)를 운반하기 위한, 또는 치료제 또는 치료용 이펙터를 운반하기 위한 운반 비히클로서 사용될 수 있는 아넬로벡터, 예를 들어, 합성 아넬로벡터를 제공한다. 일반적으로, 아넬로벡터는 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질에 대하여, 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 캡시드 단백질) 및 단백질성 외부 내에 봉입된 유전 요소를 포함하며, 여기서 유전 요소는 아넬로바이러스 이외의 바이러스 유래의 적어도 하나의 핵산 서열(예를 들어, 길이가 적어도 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, 또는 4000개 뉴클레오티드인 인접 핵산 서열), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 아넬로바이러스 이외의 바이러스 유래의 핵산 서열은 아데노-연관 바이러스(AAV)(예를 들어, 본 명세서에 기재된 바와 같음) 유래이다. 일부 구현예에서, 이펙터(예를 들어, 페이로드), 또는 이펙터를 인코딩하는 서열은 비-아넬로바이러스 서열과 별개이다. 일부 구현예에서, 단백질성 외부는 유전 요소를 표적 세포(예를 들어, 포유동물 세포, 예를 들어, 인간 세포) 내로 도입할 수 있다. 본 개시내용은 진핵 세포(예를 들어, 인간 세포 또는 인간 조직)에, 예를 들어, 유전 물질을 운반하기 위한, 이펙터, 예를 들어, 페이로드를 운반하

기 위한, 또는 치료제 또는 치료용 이펙터를 운반하기 위한 운반 비히클로서 사용될 수 있는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터(예를 들어, 합성 아넬로벡터)를 투여하기 위한 조성물 및 방법을 추가로 제공한다.

[0005] 본 명세서에 기재된 이펙터(예를 들어, 본 명세서에 기재된 바와 같은 조성물 또는 방법을 사용하여 생성됨)를 운반하기 위한 방법에 사용될 수 있는 아넬로벡터 및 이의 구성성분은 일반적으로 유전 요소를 세포(예를 들어, 포유동물 세포, 예를 들어, 인간 세포) 내로 도입할 수 있는 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 캡시드 단백질, 예를 들어, 아넬로바이러스 ORF1 분자, 예를 들어, 아넬로바이러스 ORF1 단백질 또는 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 폴리펩티드를 포함하는 단백질성 외부)에 캡슐화된 유전 요소(예를 들어, 이펙터, 예를 들어, 외인성 또는 내인성 이펙터, 예를 들어, 치료용 이펙터를 포함하거나 인코딩하는 유전 요소)를 포함한다. 유전 요소는 일반적으로 아넬로바이러스 이외의 바이러스 유래의(예를 들어, AAV, 예를 들어, AAV1, AAV2, 또는 AAV5 유래의) 적어도 하나의 핵산 서열(예를 들어, 길이가 적어도 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, 또는 4000개 뉴클레오티드인 인접 핵산 서열), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 비-아넬로바이러스 서열은, 예를 들어, 모노드나바이러스(Monodnavirus), 예를 들어, 쇼토쿠바이러스(Shotokuvirus)(예를 들어, 크레스드나비리코타(Cressdnaviricota)[예를 들어, 레돈도바이러스(redondovirus), 서코바이러스(circovirus){예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스(beak-and-feather disease virus)}, 제미니바이러스(geminivirus){예를 들어, 토마토 골든 모자이크 바이러스(tomato golden mosaic virus)}, 또는 나노바이러스(nanovirus)(예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(Parvovirus)(예를 들어, 디펜도파라바이러스(dependoparavirus), 예를 들어, 보카바이러스(bocavirus) 또는 아데노-연관 바이러스(AAV))로부터 유래된, 비-아넬로바이러스 복제 원점을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은 AAV(예를 들어, AAV1, AAV2, 또는 AAV5)로부터 유래된다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은 AAV Rep-결합 모티프(RBM), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은 AAV 말단 분해 부위(terminal resolution site; TRS), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은, 역위 말단 반복부(ITR), 예를 들어, AAV ITR에 포함된다.

[0006] 일부 구현예에서, 아넬로벡터는 아넬로바이러스 ORF1 핵산(예를 들어, 본 명세서에 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, 또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 폴리펩티드)에 의해 인코딩된 폴리펩티드를 포함하는 단백질성 외부(예를 들어, 캡시드)를 포함하는 감염성 비히클 또는 입자이다. 구현예에서, 본 명세서에 기재된 아넬로벡터는, 예를 들어, 표 A1, B1, B3, C1, E1, F1, F3, 또는 F5 중 임의의 것에 기재된 바와 같은 서열, 또는 이에 대하여 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 갖는, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함한다. 구현예에서, 본 명세서에 기재된 아넬로벡터는, 예를 들어, 표 A2, B2, B4, C2, E2, F2, F4, 또는 F6 중 임의의 것에 기재된 바와 같은 서열을 갖는, ORF1 단백질의 서열을 갖는 폴리펩티드, 또는 이에 대하여 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 폴리펩티드를 포함한다. 구현예에서, 본 명세서에 기재된 아넬로벡터는, 예를 들어, 비-아넬로바이러스 게놈을 캡슐화하는 아넬로바이러스 캡시드를 포함하는, 감염성 비히클 또는 입자이다. 아넬로바이러스 캡시드의 생성은, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 ORF1 분자의 시험관 내 생성 또는 숙주 세포 발현을 포함할 수 있다.

[0007] 일부 구현예에서, 본 개시내용의 아넬로벡터의 유전 요소는 환형 및/또는 단일-가닥 DNA 분자(예를 들어, 환형 및 단일 가닥)이다. 일부 구현예에서, 본 개시내용의 아넬로벡터의 유전 요소는 선형 및/또는 단일-가닥 DNA 분자(예를 들어, 선형 및 단일 가닥)이다. 일부 구현예에서, 유전 요소는 이를 봉입하는 단백질성 외부, 또는 이에 부착된 폴리펩티드에 결합하는 단백질 결합 서열을 포함하며, 이는 단백질성 외부 내의 유전 요소의 봉입 및

/또는 단백질성 외부 내에서의 다른 핵산에 비한 유전 요소의 농축을 용이하게 할 수 있다. 일부 구현예에서, 아넬로벡터의 유전 요소는, 본 명세서에 기재된 바와 같은 조성물 또는 방법을 사용하여 생성된다.

[0008] 일부 예에서, 본 명세서에 기재된 이펙터의 운반 방법에 사용될 수 있는 아넬로벡터는, 예를 들어, 세포에서 발현될 수 있는, 이펙터(예를 들어, 핵산 이펙터, 예컨대 비-코딩 RNA, 또는 폴리펩티드 이펙터, 예를 들어, 단백질)를 포함하거나 인코딩하는 유전 요소를 포함한다. 일부 구현예에서, 이펙터는, 예를 들어, 본 명세서에 기재된 바와 같은, 치료제 또는 치료용 이펙터이다. 일부 구현예에서, 이펙터는, 예를 들어, 야생형 아넬로바이러스 또는 표적 세포에 대하여, 내인성 이펙터 또는 외인성 이펙터이다. 일부 구현예에서, 이펙터는 야생형 아넬로바이러스 또는 표적 세포에 대하여 외인성이다. 일부 구현예에서, 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 제공됨). 다른 예에서, 이펙터는 외인성 이펙터이다. 이펙터는 일부 예에서, 세포의 기능을 조절하거나, 세포 내의 표적 분자의 활성 또는 수준을 조절할 수 있다. 예를 들어, 이펙터는 세포 내의 표적 단백질의 수준을 감소시킬 수 있다. 또 다른 예에서, 아넬로벡터는 생체 내에서 이펙터, 예를 들어, 외인성 단백질을 운반하고 발현할 수 있다. 아넬로벡터는, 예를 들어, 유전 물질을 표적 세포, 조직 또는 대상체로 운반하기 위해; 이펙터를 표적 세포, 조직 또는 대상체로 운반하기 위해; 생물학적 반응, 예를 들어, 세포 또는 분자 반응을 조절하기 위해; 또는, 예를 들어, 요망되는 세포, 조직, 또는 대상체에 대하여 조절제 및/또는 치료제로서 작용할 수 있는 이펙터를 운반함으로써, 질환 또는 장애와 같은 병태를 치료하기 위해 사용될 수 있다.

[0009] 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은, 예를 들어, 숙주 세포에서, 본 명세서에 기재된 아넬로벡터를 투여하는 방법에 사용되는 합성 아넬로벡터의 유전 요소를 생성하는 데 사용될 수 있다. 합성 아넬로벡터는 야생형 바이러스(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스)와 비교하여 적어도 하나의 구조적 차이, 예를 들어, 야생형 바이러스에 비하여 결실, 삽입, 치환, 변형(예를 들어, 효소적 변형)을 갖는다. 일부 구현예에서, 구조적 차이는, 예를 들어, 본 명세서에 기재된 바와 같은, 유전 요소의 비-아넬로바이러스 서열을 포함한다. 일반적으로, 합성 아넬로벡터는 단백질성 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 진핵(예를 들어, 인간) 세포 내로 운반하기 위해 사용될 수 있다. 구현예에서, 아넬로벡터는 검출 가능하고/가능하거나 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% 초과의 증가를 야기하지 않으며, 예를 들어, 아넬로벡터는 표적 세포, 조직 또는 대상체에 대하여 실질적으로 비-면역원성일 수 있다.

[0010] 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은 아넬로벡터, 예를 들어, 본 명세서에 기재된 이펙터의 운반 방법에 사용될 수 있는 아넬로벡터의 유전 요소를 생성하는 데 사용될 수 있으며, 이는 (i) 프로모터 요소, 및 이펙터(예를 들어, 내인성 또는 외인성 이펙터)를 인코딩하는 서열, 및 단백질 결합 서열(예를 들어, 외부 단백질 결합 서열, 예를 들어, 패키징 신호)을 포함하는 유전 요소; 및 (ii) 단백질성 외부를 포함하고; 유전 요소는 단백질성 외부(예를 들어, 캡시드) 내에 봉입되고, 아넬로벡터는 유전 요소를 진핵(예를 들어, 포유동물, 예를 들어, 인간) 세포 내로 운반할 수 있다. 일부 구현예에서, 유전 요소는 단일-가닥 및/또는 환형 DNA이다. 대안적으로 또는 조합하여, 유전 요소는 하기 특성 중 1, 2, 3개, 또는 전부를 갖는다: 환형이고/이거나, 단일-가닥이고/이거나, 세포에 유입되는 유전 요소의 약 0.0001%, 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, 또는 2% 미만의 빈도로 세포의 게놈 내로 통합되고/되거나, 게놈당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 또는 30개 카피 미만으로 표적 세포의 게놈 내로 통합됨. 일부 구현예에서, 통합 빈도는, 예를 들어, 문헌[Wang et al., 2004, *Gene Therapy* 11: 711-721](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같이, 자유 벡터에서 분리된 게놈 DNA의 정량적 겔 정제 분석에 의해 결정된다. 일부 구현예에서, 유전 요소는 단백질성 외부 내에 봉입된다. 일부 구현예에서, 아넬로벡터는 유전 요소를 진핵 세포 내로 운반할 수 있다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스의 서열(예를 들어, 야생형 토르크 테노 바이러스(TTV), 토르크 테노 미니 바이러스(TTMV) 또는 TTMDV 서열, 예를 들어, 본 명세서에 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 핵산 서열(예를 들어, 300 내지 4000개 뉴클레오티드, 예를 들어, 300 내지 3500개 뉴클레오티드, 300 내지 3000개 뉴클레오티드, 300 내지 2500개 뉴클레오티드, 300 내지 2000개 뉴클레오티드, 300 내지 1500개 뉴클레오티드의 핵산 서열)을 포함한다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스의 서열(예를 들어, 본 명세서에 기재된 바와 같은 야생형 아넬로바이러스 서열)의

서열에 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 핵산 서열(예를 들어, 적어도 300개 뉴클레오티드, 500개 뉴클레오티드, 1000개 뉴클레오티드, 1500개 뉴클레오티드, 2000개 뉴클레오티드, 2500개 뉴클레오티드, 3000개 뉴클레오티드 또는 그 이상의 핵산 서열)을 포함한다. 일부 구현예에서, 핵산 서열은, 예를 들어, 포유동물(예를 들어, 인간) 세포에서의 발현을 위하여 코돈-최적화된다. 일부 구현예에서, 핵산 서열 내의 적어도 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 코돈은, 예를 들어, 포유동물(예를 들어, 인간) 세포에서의 발현을 위하여 코돈-최적화된다.

[0011] 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은 본 명세서에 기재된 아넬로벡터를 투여하는 방법에 사용될 수 있는 치료용 이펙터를 인코딩하는 (아넬로바이러스에 대한) 이종 서열 및 캡시드에 결합하는 단백질 결합 서열을 포함하는 유전 요소를 캡슐화하는 캡시드(예를 들어, 아넬로바이러스 ORF, 예를 들어, ORF1, 폴리펩티드를 포함하는 캡시드)를 포함하는 (예를 들어, 인간 세포에 대한) 감염성 아넬로벡터, 비히클, 또는 입자의 유전 요소를 생성하는 데 사용될 수 있다. 구현예에서, 아넬로벡터는 유전 요소를 포유동물, 예를 들어, 인간, 세포 내로 운반할 수 있다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스 게놈 서열에 대하여 약 6% 미만(예를 들어, 10%, 9.5%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5% 미만 또는 그 이하)의 동일성을 갖는다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스 게놈 서열에 대하여 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 5.5% 또는 6% 이하의 동일성을 갖는다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스에 대하여 적어도 약 2% 내지 적어도 약 5.5%(예를 들어, 2 내지 5%, 3% 내지 5%, 4% 내지 5%)의 동일성을 갖는다. 일부 구현예에서, 유전 요소는 약 2000, 3000, 4000, 4500 또는 5000개 초과의 비-바이러스 서열(예를 들어, 비 아넬로바이러스 게놈 서열)의 뉴클레오티드를 갖는다. 일부 구현예에서, 유전 요소는 약 2000 내지 5000, 2500 내지 4500, 3000 내지 4500, 2500 내지 4500, 3500 또는 4000, 4500개(예를 들어, 약 3000 내지 4500개) 초과의 비-바이러스 서열(예를 들어, 비 아넬로바이러스 게놈 서열)의 뉴클레오티드를 갖는다. 일부 구현예에서, 유전 요소는 단일-가닥, 환형 DNA이다. 대안적으로 또는 조합하여, 유전 요소는 하기 특성 중 1, 2, 또는 3개를 갖는다: 환형이거나, 단일 가닥이거나, 세포에 유입되는 유전 요소의 약 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, 또는 2% 미만의 빈도로 세포의 게놈 내로 통합되고/되거나, 게놈당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 또는 30개 카피 미만으로 표적 세포의 게놈 내로 통합되거나, (예를 들어, 세포 용해물로부터의 유전 요소 서열에 대한 게놈 DNA 내로의 통합 빈도를 비교함으로써) 세포에 유입되는 유전 요소의 약 0.0001%, 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, 또는 2% 미만의 빈도로 통합됨. 일부 구현예에서, 통합 빈도는, 예를 들어, 문헌[Wang et al., 2004, *Gene Therapy* 11: 711-721](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같이, 자유 벡터에서 분리된 게놈 DNA의 정량적 겔 정제 분석에 의해 결정된다.

[0012] 일부 구현예에서, 본 명세서에 기재된 방법에 따라 투여되는 아넬로바이러스 또는 아넬로벡터는, 본 명세서에 기재된 이펙터와 같은 작용제를 표적 세포, 예를 들어, 치료적 또는 예방적으로 치료될 대상체의 표적 세포에 도입하기 위한 효과적인 운반 비히클로 사용될 수 있다.

[0013] 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은 본 명세서에 기재된 투여 방법에 사용될 수 있는 아넬로벡터의 유전 요소를 생성하는 데 사용될 수 있으며, 유전 요소는 하기를 (예를 들어, 연속하여) 포함하는 폴리펩티드(예를 들어, 합성 폴리펩티드, 예를 들어, ORF1 분자)를 포함하는 단백질성 외부를 포함한다:

[0014] (i) 아르기닌-풍부 영역, 예를 들어, 적어도 60%, 70%, 또는 80%의 염기성 잔기(예를 들어, 아르기닌, 라이신, 또는 이의 조합)를 포함하는 적어도 약 40개의 아미노산의 서열을 포함하는 제1 영역,

[0015] (ii) 젤리-롤(jelly-roll) 도메인, 예를 들어, 적어도 6개의 베타 가닥을 포함하는 서열을 포함하는 제2 영역,

[0016] (iii) 본 명세서에 기재된 N22 도메인 서열을 포함하는 제3 영역,

[0017] (iv) 본 명세서에 기재된 아넬로바이러스 ORF1 C-말단 도메인(CTD) 서열을 포함하는 제4 영역, 및

[0018] (v) 선택적으로 폴리펩티드는, 예를 들어, 본 명세서에 기재된 바와 같은, 야생형 아넬로바이러스 ORF1 단백질에 대하여 100%, 99%, 98%, 95%, 90%, 85%, 80% 미만의 서열 동일성을 갖는 아미노산 서열을 가짐.

[0019] 일 양태에서, 본 발명은 이펙터, 예를 들어, 페이로드를 인코딩하는 서열에 작동 가능하게 연결된 프로모터 요소, 및 외부 단백질 결합 서열을 포함하는 유전 요소의 서열을 포함하는 단리된 핵산 분자(예를 들어, 핵산 작제물)를 특징으로 한다. 일부 구현예에서, 외부 단백질 결합 서열은, 예를 들어, 본 명세서에 개시된 바와 같은, 아넬로바이러스의 5'UTR 서열에 대하여 적어도 75%(적어도 80%, 85%, 90%, 95%, 97%, 100%) 동일한 서열

을 포함한다. 구현예에서, 유전 요소는 단일-가닥 DNA이고/이거나, 환형이고/이거나, 세포에 유입되는 유전 요소의 약 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, 또는 2% 미만의 빈도로 통합되고/되거나, 게놈당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 또는 30개 카피 미만으로 표적 세포의 게놈 내로 통합되거나, 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 구현예에서, 통합 빈도는, 예를 들어, 문헌[Wang et al., 2004, *Gene Therapy* 11: 711-721](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같이, 자유 벡터에서 분리된 게놈 DNA의 정량적 겔 정제 분석에 의해 결정된다. 구현예에서, 이펙터는 TTV로부터 기원하지 않으며, SV40-miR-S1이 아니다. 구현예에서, 핵산 분자는 TTMV-LY2의 폴리뉴클레오티드 서열을 포함하지 않는다. 구현예에서, 프로모터 요소는 진핵(예를 들어, 포유동물, 예를 들어, 인간) 세포에서 이펙터의 발현을 지시할 수 있다.

[0020] 일부 구현예에서, 핵산 분자는 환형이다. 일부 구현예에서, 핵산 분자는 선형이다. 일부 구현예에서, 본 명세서에 기재된 핵산 분자는 하나 이상의 변형된 뉴클레오티드(예를 들어, 염기 변형, 당 변형, 또는 백본 변형)를 포함한다.

[0021] 일부 구현예에서, 핵산 분자는 ORF1 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질)를 인코딩하는 서열을 포함한다. 일부 구현예에서, 핵산 분자는 ORF2 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF2 단백질)를 인코딩하는 서열을 포함한다. 일부 구현예에서, 핵산 분자는 ORF3 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF3 단백질)를 인코딩하는 서열을 포함한다. 일 양태에서, 본 발명은 (i) 프로모터 요소 및 이펙터, 예를 들어, 외인성 또는 내인성 이펙터를 인코딩하는 서열; (ii) 야생형 아넬로바이러스 서열에 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 적어도 72개의 인접 뉴클레오티드(예를 들어, 적어도 72, 73, 74, 75, 76, 77, 78, 79, 80, 90, 100 또는 150개 뉴클레오티드); 또는 야생형 아넬로바이러스 서열에 대하여 적어도 72%(예를 들어, 적어도 72, 73, 74, 75, 76, 77, 78, 79, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100%) 서열 동일성을 갖는 적어도 100개(예를 들어, 적어도 300, 500, 1000, 1500개)의 인접 뉴클레오티드; 및 (iii) 단백질 결합 서열, 예를 들어, 외부 단백질 결합 서열 중 1개, 2개 또는 3개를 포함하는 유전 요소를 특징으로 하며, 핵산 작제물은 단일-가닥 DNA이며; 핵산 작제물은 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 통합되고/통합되거나, 게놈당 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 또는 30개 카피 미만으로 표적 세포의 게놈 내로 통합된다. 일부 구현예에서, 이펙터(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 외인성 또는 내인성 이펙터)를 인코딩하는 유전 요소는 코돈 최적화된다. 일부 구현예에서, 유전 요소는 환형이다. 일부 구현예에서, 유전 요소는 선형이다. 일부 구현예에서, 본 명세서에 기재된 유전 요소는 하나 이상의 변형된 뉴클레오티드(예를 들어, 염기 변형, 당 변형, 또는 백본 변형)를 포함한다. 일부 구현예에서, 유전 요소는 ORF1 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질)를 인코딩하는 서열을 포함한다. 일부 구현예에서, 유전 요소는 ORF2 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF2 단백질)를 인코딩하는 서열을 포함한다. 일부 구현예에서, 유전 요소는 ORF3 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF3 단백질)를 인코딩하는 서열을 포함한다.

[0022] 일 양태에서, 본 발명은 하기를 포함하는 숙주 세포를 특징으로 한다: (a) ORF1 분자, ORF2 분자, 또는 ORF3 분자 중 하나 이상을 인코딩하는 서열(예를 들어, 본 명세서에 기재된 아넬로바이러스 ORF1 폴리펩티드를 인코딩하는 서열)을 포함하되, 예를 들어, 플라스미드이거나, 바이러스 핵산이거나, 염색체에 통합되는 하나 이상의 핵산 분자; 및 (b) (i) 이펙터(예를 들어, 외인성 이펙터 또는 내인성 이펙터)를 인코딩하는 핵산 서열(예를 들어, DNA 서열)에 작동 가능하게 연결된 프로모터 요소, 및 (ii) (a)의 ORF1 분자에 결합하는 단백질 결합 서열을 포함하는 유전 요소(여기서, (b)의 유전 요소는 ORF1 폴리펩티드(예를 들어, ORF1 단백질), ORF2 폴리펩티드(예를 들어, ORF2 단백질), 및/또는 ORF3 폴리펩티드(예를 들어, ORF3 단백질) 중 하나 이상을 인코딩하지 않음). 예를 들어, 숙주 세포는 (a) 및 (b)를 시스(둘 다 동일한 핵산 분자의 부분임) 또는 트랜스(각각 상이한 핵산 분자의 부분임)로 포함한다. 구현예에서, (a)의 하나 이상의 핵산은 환형, 단일-가닥 DNA일 수 있고; 다른 구현예에서, (a)의 하나 이상의 핵산은 선형 DNA일 수 있다. 구현예에서. (b)의 유전 요소는 환형, 단일-가닥 DNA이다. 일부 구현예에서, 숙주 세포는, 예를 들어, 본 명세서에 기재된 바와 같은, 제조(manufacturing) 세포주이다. 일부 구현예에서, 숙주 세포는 부착성이거나, 현탁액 중에 존재하거나, 둘 다이다. 일부 구현예에서, 숙주 세포 또는 헬퍼 세포는 마이크로캐리어에서 성장된다. 일부 구현예에서, 숙주 세포 또는 헬퍼 세포는 cGMP 제조 실무와 양립 가능하다. 일부 구현예에서, 숙주 세포 또는 헬퍼 세포는 세포 성장을 촉진하기에 적합한 배지 중에서 성장된다. 특정 구현예에서, 일단 숙주 세포 또는 헬퍼 세포가 (예를 들어, 적절한 세포 밀도로) 충

분히 성장하면, 배지를 숙주 세포 또는 헬퍼 세포에 의한 아넬로벡터의 생성에 적합한 배지로 교체할 수 있다.

[0023] 일 양태에서, 본 발명은 아넬로벡터(예를 들어, 합성 아넬로벡터), 예를 들어, 본 명세서에 기재된 방법에 의해 투여될 수 있는 아넬로벡터를 포함하는 약제학적 조성물을 특징으로 한다. 구현예에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체 또는 부형제를 추가로 포함한다. 구현예에서, 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 게놈 당량의 아넬로벡터를 포함하는 단위 용량을 포함한다. 일부 구현예에서, 제제를 포함하는 약제학적 조성물은 허용 가능한 기간 및 온도에 걸쳐 안정할 것이고/이거나, 요망되는 투여 경로 및/또는 투여 경로가 필요로 할 임의의 디바이스, 예를 들어, 니들(needle) 또는 주사기와 병용 가능할 것이다. 일부 구현예에서, 약제학적 조성물은 단회 용량 또는 다회 용량으로서의 투여를 위해 제형화된다. 일부 구현예에서, 약제학적 조성물은 투여 장소에서, 예를 들어, 의료 전문가(healthcare professional)에 의해 제형화된다. 일부 구현예에서, 약제학적 조성물은 (예를 들어, 부피당 게놈의 수에 의해 정의된 바와 같은) 요망되는 농도의 아넬로벡터 게놈 또는 게놈 당량을 포함한다.

[0024] 일 양태에서, 본 발명은 대상체에서의 질환 또는 장애의 치료 방법을 특징으로 하며, 방법은 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터, 예를 들어, 합성 아넬로벡터를 대상체에게 투여하는 단계를 포함한다.

[0025] 일 양태에서, 본 발명은 이펙터 또는 페이로드(예를 들어, 내인성 또는 외인성 이펙터)를 세포, 조직 또는 대상체에 운반하는 방법을 특징으로 하며, 방법은 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터, 예를 들어, 합성 아넬로벡터를 대상체에게 투여하는 단계를 포함하며, 아넬로벡터는 이펙터를 인코딩하는 핵산 서열을 포함한다. 구현예에서, 페이로드는 핵산이다. 구현예에서, 페이로드는 폴리펩티드이다.

[0026] 일 양태에서, 본 발명은 예를 들어, 본 명세서에 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.

[0027] 일 양태에서, 본 발명은 아넬로벡터, 예를 들어, 본 명세서에 기재된 아넬로벡터를 투여하는 방법에 사용될 수 있는 합성 아넬로벡터의 제조 방법을 특징으로 한다. 방법은 하기를 포함한다:

[0028] (a) (i) 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터의 유전 요소의 핵산 서열을 포함하는 제1 핵산 분자; 및

[0029] (ii) 아넬로바이러스 ORF1 폴리펩티드, 또는, 예를 들어, 본 명세서에 기재된 바와 같은 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2로부터 선택되는 아미노산 서열 중 하나 이상, 또는 이에 대하여 적어도 70%(예를 들어, 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%) 서열 동일성을 갖는 아미노산 서열을 인코딩하는 제2 핵산 분자

[0030] 를 포함하는 숙주 세포를 제공하는 단계; 및

[0031] (b) 숙주 세포를 유전 요소의 핵산 서열의 복제(예를 들어, 회전환 복제(rolling circle replication))에 적합한 조건 하에서 인큐베이션하는 단계; 및

[0032] 선택적으로 (c) 단백질성 외부(예를 들어, 제2 핵산 분자에 의해 인코딩된 폴리펩티드를 포함함)에 유전 요소를 봉입하기에 적합한 조건 하에서 숙주 세포를 인큐베이션하는 단계.

[0033] 또 다른 양태에서, 본 발명은 아넬로벡터 조성물, 예를 들어, 본 명세서에 기재된 투여 방법에 사용될 수 있는 아넬로벡터 조성물의 제조 방법을 특징으로 하며, 조성물은 (a), (b), 및 (c) 중 하나 이상(예를 들어, 이들 전부)을 포함한다:

[0034] a) 아넬로벡터, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 합성 아넬로벡터 중 하나 이상의 구성성분(예를 들어, 모든 구성성분)을 포함하는, 예를 들어, 발현하는 숙주 세포를 제공하는 단계;

[0035] b) 숙주 세포로부터 아넬로벡터의 제제를 생성하기에 적합한 조건 하에서 숙주 세포를 배양하되, 제제의 아넬로벡터는 유전 요소(예를 들어, 본 명세서에 기재된 바와 같음)를 캡슐화하는 단백질성 외부(예를 들어, 아넬로벡터 ORF1 폴리펩티드를 포함함)를 포함하여, 아넬로벡터의 제제를 제조하는 단계; 및

[0036] 선택적으로, c) 아넬로벡터의 제제를, 예를 들어, 대상체에게 투여하기에 적합한 약제학적 조성물로서 재형화하는 단계.

[0037] 예를 들어, 이 제조 방법에서 제공되는 숙주 세포는 (a) 본 명세서에 기재된 아넬로바이러스 ORF1 폴리펩티드를 인코딩하는 서열을 포함하되, 핵산은 플라스미드이거나, 바이러스 핵산 또는 게놈이거나, 헬퍼 세포 염색체에 통합되는 핵산; 및 (b) 유전 요소(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 유전 요소 서열 및/또는 유전 요소 영역을 포함함)를 생성할 수 있는 핵산 작제물을 포함하되, 예를 들어, 유전 요소는 (i) 이펙터(예를 들어, 외인성 이펙터 또는 내인성 이펙터)를 인코딩하는 핵산 서열(예를 들어, DNA 서열)에 작동 가능하게 연결된 프로모터 요소, 및 (i) (a)의 폴리펩티드에 결합하는 단백질 결합 서열(예를 들어, 패키징 서열)을 포함하고, 숙주 세포는 (a) 및 (b)를 시스 또는 트랜스로 포함한다. 구현예에서, (b)의 유전 요소는 환형, 단일-가닥 DNA이다. 일부 구현예에서, 숙주 세포는 제조 세포주이다.

[0038] 일부 구현예에서, 아넬로벡터의 구성성분은 생성 시에 (예를 들어, 일시적 형질감염에 의해) 숙주 세포 내로 도입된다. 일부 구현예에서, 숙주 세포는 아넬로벡터의 구성성분을 안정하게 발현한다(예를 들어, 아넬로벡터의 구성성분을 인코딩하는 하나 이상의 핵산은, 예를 들어, 안정한 형질감염에 의해 숙주 세포 또는 이의 자손 내로 도입됨).

[0039] 일 양태에서, 본 발명은 a) 본 명세서에 기재된 복수의 아넬로벡터, 또는 본 명세서에 기재된 아넬로벡터의 제제를 제공하는 단계; 및 b) 아넬로벡터 또는 이의 제제를, 예를 들어, 대상체에게 투여하기에 적합한 약제학적 조성물로서 제형화하는 단계를 포함하는, 아넬로벡터 조성물을 제조하는 방법을 특징으로 한다.

[0040] 일 양태에서, 본 발명은 아넬로벡터를 포함하는 숙주 세포, 예를 들어, 제1 숙주 세포 또는 생성자 세포(예를 들어, PCT/US19/65995의 도 12에 나타낸 바와 같음), 예를 들어, 제1 숙주 세포의 집단의 제조 방법을 특징으로 하며, 방법은 예를 들어, 본 명세서에 기재된 바와 같은 유전 요소를 생성하는 핵산 작제물을 숙주 세포 내로 도입하는 단계, 및 숙주 세포를 아넬로벡터의 생성에 적합한 조건 하에서 배양하는 단계를 포함한다. 구현예에서, 방법은 헬퍼, 예를 들어, 헬퍼 바이러스를, 숙주 세포에 도입하는 단계를 추가로 포함한다. 구현예에서, 도입하는 단계는 아넬로벡터를 이용한 숙주 세포의 형질감염(예를 들어, 화학적 형질감염) 또는 전기천공법을 포함한다.

[0041] 일 양태에서, 본 발명은, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터를 포함하는 숙주 세포, 예를 들어, 제1 숙주 세포 또는 생성자 세포(예를 들어, PCT/US19/65995의 도 12에 나타낸 바와 같음)를 제공하는 단계, 및 숙주 세포로부터 아넬로벡터를 정제하는 단계를 포함하는 아넬로벡터의 제조 방법을 특징으로 한다. 일부 구현예에서, 방법은 상기 제공하는 단계 이전에, 숙주 세포를, 예를 들어, 본 명세서에 기재된 바와 같은, 핵산 작제물 또는 아넬로벡터와 접촉시키는 단계, 및 숙주 세포를 아넬로벡터의 생성에 적합한 조건 하에서 인큐베이션하는 단계를 추가로 포함한다. 구현예에서, 숙주 세포는 상기 숙주 세포의 제조 방법에 기재된 제1 숙주 세포 또는 생성자 세포이다. 구현예에서, 숙주 세포로부터 아넬로벡터를 정제하는 단계는 숙주 세포를 용해시키는 것을 포함한다.

[0042] 일부 구현예에서, 방법은 제1 숙주 세포 또는 생성자 세포에 의해 생성되는 아넬로벡터를 제2 숙주 세포, 예를 들어, 허용 세포(예를 들어, PCT/US19/65995의 도 12에 나타낸 바와 같음), 예를 들어, 제2 숙주 세포의 집단과 접촉시키는 제2 단계를 추가로 포함한다. 일부 구현예에서, 방법은 제2 숙주 세포를 아넬로벡터의 생성에 적합한 조건 하에서 인큐베이션하는 단계를 추가로 포함한다. 일부 구현예에서, 방법은 제2 숙주 세포로부터 아넬로벡터를 정제함으로써, 예를 들어, 아넬로벡터 시드 집단을 생성하는 단계를 추가로 포함한다. 구현예에서, 제1 숙주 세포의 집단으로부터보다 제2 숙주 세포의 집단으로부터 적어도 약 2 내지 100배 더 많은 아넬로벡터가 생성된다. 구현예에서, 제2 숙주 세포로부터 아넬로벡터를 정제하는 단계는 제2 숙주 세포를 용해시키는 것을 포함한다. 일부 구현예에서, 방법은 제2 숙주 세포에 의해 생성되는 아넬로벡터를 제3 숙주 세포, 예를 들어, 허용 세포(예를 들어, PCT/US19/65995의 도 12에 나타낸 바와 같음), 예를 들어, 제3 숙주 세포의 집단과 접촉시키는 제2 단계를 추가로 포함한다. 일부 구현예에서, 방법은 제3 숙주 세포를 아넬로벡터의 생성에 적합한 조건 하에서 인큐베이션하는 단계를 추가로 포함한다. 일부 구현예에서, 방법은 제3 숙주 세포로부터 아넬로벡터를 정제함으로써, 예를 들어, 아넬로벡터 스톡(stock) 집단을 생성하는 단계를 추가로 포함한다. 구현예에서, 제3 숙주 세포로부터 아넬로벡터를 정제하는 단계는 제3 숙주 세포를 용해시키는 것을 포함한다. 구현예에서, 제2 숙주 세포의 집단으로부터보다 제3 숙주 세포의 집단으로부터 적어도 약 2 내지 100배 더 많은 아넬로벡터가 생성된다.

[0043] 일부 구현예에서, 숙주 세포는 세포 성장을 촉진하기에 적합한 배지 중에서 성장된다. 특정 구현예에서, 일단 숙주 세포가 (예를 들어, 적절한 세포 밀도로) 충분히 성장하면, 배지를 숙주 세포에 의한 아넬로벡터의 생성에 적합한 배지로 교체할 수 있다. 일부 구현예에서, 숙주 세포에 의해 생성되는 아넬로벡터는 제2 숙주 세포와의

접촉 이전에 (예를 들어, 숙주 세포를 용해시킴으로써) 숙주 세포로부터 분리된다. 일부 구현예에서, 숙주 세포에 의해 생성되는 아넬로벡터는 개재되는 정제 단계 없이 제2 숙주 세포와 접촉된다.

[0044] 일 양태에서, 본 발명은 약제학적 아넬로벡터 제제, 예를 들어, 본 명세서에 기재된 투여 방법에 사용되는 제제를 제조하는 방법을 특징으로 한다. 방법은 (a) 본 명세서에 기재된 바와 같은 아넬로벡터 제제를 제조하는 단계, (b) 제제(예를 들어, 약제학적 아넬로벡터 제제, 아넬로벡터 시드 집단 또는 아넬로벡터 스톡 집단)를 하나 이상의 약제학적 품질 관리 매개변수, 예를 들어, 아이덴티티, 순도, 역가, 효력(예를 들어, 아넬로벡터 입자당 게놈 당량) 및/또는 예를 들어, 아넬로벡터에 포함되는 유전 요소로부터의 핵산 서열에 대하여 평가하는 단계, 및 (c) 평가가 사전결정된 기준을 만족시키는, 예를 들어, 약제학적 규격을 만족시키는 약제학적 이용을 위한 제제를 제형화하는 단계를 포함한다. 일부 구현예에서, 아이덴티티를 평가하는 것은 아넬로벡터의 유전 요소의 서열, 예를 들어, 이펙터를 인코딩하는 서열을 평가하는 것(예를 들어, 확인하는 것)을 포함한다. 일부 구현예에서, 순도를 평가하는 것은 불순물, 예를 들어, 마이코플라스마, 내독소, 숙주 세포 핵산(예를 들어, 숙주 세포 DNA 및/또는 숙주 세포 RNA), 동물-유래 과정 불순물(예를 들어, 혈청 알부민 또는 트립신), 복제-적격 작용제(RCA), 예를 들어, 복제-적격 바이러스 또는 원치 않는 아넬로벡터(예를 들어, 요망되는 아넬로벡터, 예를 들어, 본 명세서에 기재된 바와 같은 합성 아넬로벡터 이외의 아넬로벡터), 유리 바이러스 캡시드 단백질, 외래성 물질(adventitious agent) 및 응집물의 양을 평가하는 것을 포함한다. 일부 구현예에서, 역가를 평가하는 것은 (예를 들어, HPLC에 의해 평가되는 바와 같은) 제제 내의 기능성 대 비-기능성(예를 들어, 감염성 대 비-감염성) 아넬로벡터의 비를 평가하는 것을 포함한다. 일부 구현예에서, 효력을 평가하는 것은 제제에서 검출 가능한 아넬로벡터 기능(예를 들어, 그 내에 인코딩된 이펙터의 발현 및/또는 기능 또는 게놈 당량)의 수준을 평가하는 것을 포함한다.

[0045] 구현예에서, 제형화된 제제는 병원체, 숙주 세포 오염물질 또는 불순물이 실질적으로 없거나; 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 갖는다. 일부 구현예에서, 다수의 아넬로벡터가 단일의 배치(batch)에서 생성될 수 있다. 구현예에서, 배치에서 생성되는 아넬로벡터의 수준이 (예를 들어, 개별적으로 또는 함께) 평가될 수 있다.

[0046] 일 양태에서, 본 발명은 하기를 포함하는 숙주 세포를 특징으로 한다:

[0047] (i) 본 명세서에 기재된 바와 같은 핵산 작제물을 포함하는 제1 핵산 분자, 및

[0048] (ii) 선택적으로, 예를 들어, 본 명세서에 기재된 바와 같은 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 또는 ORF1/2로부터 선택되는 아미노산 서열 또는 이에 대하여 적어도 약 70%(예를 들어, 적어도 약 70, 80, 90, 95, 96, 97, 98, 99 또는 100%) 서열 동일성을 갖는 아미노산 서열 중 하나 이상을 인코딩하는 제2 핵산 분자.

[0049] 일 양태에서, 본 발명은 본 명세서에 기재된 아넬로벡터 및 본 명세서에 기재된 투여 방법에 사용될 수 있는 헬퍼 바이러스를 포함하는 반응 혼합물을 특징으로 하며, 헬퍼 바이러스는 외부 단백질(예를 들어, 외부 단백질 결합 서열에 결합할 수 있는 외부 단백질, 및 선택적으로 지질 외피)을 인코딩하는 폴리뉴클레오티드, 복제 단백질(예를 들어, 중합효소)을 인코딩하는 폴리뉴클레오티드, 또는 이들의 임의의 조합을 포함한다.

[0050] 일부 구현예에서, 아넬로벡터(예를 들어, 합성 아넬로벡터)는 단리되며, 예를 들어, 숙주 세포로부터 단리되고/되거나 용액(예를 들어, 상청액) 중 다른 구성요소로부터 단리된다. 일부 구현예에서, 아넬로벡터(예를 들어, 합성 아넬로벡터)는 예를 들어, 용액(예를 들어, 상청액)으로부터 정제된다. 일부 구현예에서, 아넬로벡터는 용액 중 다른 구성요소에 비하여 용액 중에서 농축된다.

[0051] 상기 아넬로벡터, 조성물 또는 방법 중 임의의 것의 일부 구현예에서, 아넬로벡터를 제공하는 것은, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터-생성 세포를 포함하는 조성물로부터 아넬로벡터를 분리하는 것(예를 들어, 수확하는 것)을 포함한다. 다른 구현예에서, 아넬로벡터를 제공하는 것은, 예를 들어, 제3자로부터 아넬로벡터 또는 이의 제제를 수득하는 것을 포함한다.

[0052] 구현예에서, 유전 요소는 자가-복제할 수 없고/없거나 자가-증폭할 수 없다. 구현예에서, 유전 요소는, 예를 들어, 헬퍼, 예를 들어, 헬퍼 바이러스의 존재 하에 트랜스로 복제할 수 있고/있거나 증폭할 수 있다.

[0053] 상기 아넬로벡터, 조성물 또는 방법 중 임의의 것의 추가의 특징은 하기 열거된 구현예 중 하나 이상을 포함한다.

[0054] 당업자는 단지 일상적인 실험을 사용하여, 본 명세서에 기재된 본 발명의 특정 구현예의 다수의 등가물을 인식하거나 확인할 수 있을 것이다. 그러한 등가물은 하기 열거된 구현예에 의해 포함될 것이다.

[0055] **열거된 구현예**

[0056] 1. 바이러스 입자로서,

[0057] (i) AAV 복제 원점, (ii) 치료용 RNA 또는 폴리펩티드를 인코딩하는 서열에 작동 가능하게 연결된 프로모터, 및 (iii) 아넬로바이러스 ORF1 분자에 결합하는 서열을 포함하는 환형 DNA를 포함하되, 환형 DNA는 아넬로바이러스 ORF1 분자를 포함하는 캡시드에 의해 캡시드화된, 바이러스 입자.

[0058] 2. 바이러스 입자로서,

[0059] (i) AAV 복제 원점, 및 (ii) 치료용 RNA 또는 폴리펩티드를 인코딩하는 서열에 작동 가능하게 연결된 프로모터를 포함하는 환형 DNA를 포함하되, 환형 DNA는 아넬로바이러스 ORF1 분자를 포함하는 캡시드에 의해 캡시드화된, 바이러스 입자.

[0060] 3. 벡터로서,

[0061] a) 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부; 및

[0062] b) 비-아넬로바이러스 복제 원점을 포함하는 유전 요소

[0063] 를 포함하며,

[0064] 선택적으로 유전 요소는 (i) 외인성 이펙터를 인코딩하는 핵산 서열, 및/또는 (ii) 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소를 추가로 포함하는, 벡터.

[0065] 4. 구현예 3에 있어서, 비-아넬로바이러스 복제 원점은 DNA 바이러스, 예를 들어, 단일-가닥 DNA(ssDNA) 바이러스, 예를 들어, 선형 ssDNA 바이러스로부터 유래된, 벡터.

[0066] 5. 구현예 3 또는 4에 있어서, 비-아넬로바이러스 복제 원점은 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스(예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV)로부터 유래된, 벡터.

[0067] 6. 구현예 5에 있어서, 비-아넬로바이러스 복제 원점은 모노드나바이러스, 예를 들어, 쇼토쿠바이러스, 예를 들어, 코사비리코타(Cossaviricota), 예를 들어, 퀸토비리세테스(Quintoviricetes), 예를 들어, 피코비랄레스(Piccovirales), 예를 들어, 파르보비리대(Parvoviridae), 예를 들어, 파르보비리내(Parvovirinae), 예를 들어, 디펜도파르보바이러스(Dependoparvovirus), 예를 들어, 아데노-연관 바이러스(AAV)로부터 유래된, 벡터.

[0068] 7. 구현예 5에 있어서, 비-아넬로바이러스 복제 원점은 AAV(예를 들어, AAV1, AAV2, 또는 AAV5) 복제 원점인, 벡터.

[0069] 8. 구현예 5에 있어서, 비-아넬로바이러스 복제 원점은 회전환 복제에 의해 복제하는 바이러스로부터 유래된, 벡터.

[0070] 9. 구현예 5에 있어서, 비-아넬로바이러스 복제 원점은 회전 헤어핀 복제(rolling hairpin replication)에 의해 복제하는 바이러스로부터 유래된, 벡터.

[0071] 10. 구현예 5에 있어서, 비-아넬로바이러스 복제 원점은 동물(예를 들어, 포유동물, 예를 들어, 인간), 식물, 진균, 또는 박테리아를 감염시키는 바이러스로부터 유래된, 벡터.

[0072] 11. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 AAV Rep-결합 모티프(RBM), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함하는, 벡터.

[0073] 12. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 AAV 말단 분해 부위(TRS), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함하는, 벡터.

[0074] 13. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 역위 말단 반복부(ITR)를 포함하는, 벡터.

[0075] 14. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 아넬로바이러스 복제 원점, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하지 않는, 벡터.

[0076] 15. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 실질적으로 회전환 복제에 의해 복제하지 않는(예를 들어, 복제할 수 없는), 벡터.

[0077] 16. 선행하는 구현예 중 어느 하나에 있어서, 비-아넬로바이러스 복제 원점은 아넬로바이러스 게놈(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 또는 200개 뉴클레오티드의 인접 서열을 포함하지 않는, 벡터.

[0078] 17. 유전 요소로서,

[0079] 아넬로바이러스 ORF1 분자(예를 들어, 5' UTR)에 특이적으로 결합하는 단백질 결합 서열; 및

[0080] 예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점;

[0081] 선택적으로, 외인성 이펙터(예를 들어, 치료용 외인성 이펙터)를 인코딩하는 핵산 서열; 및

[0082] 선택적으로, 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소

[0083] 를 포함하는, 유전 요소.

[0084] 18. 유전 요소 작제물로서,

[0085] 아넬로바이러스 ORF1 분자(예를 들어, 5' UTR)에 특이적으로 결합하는 단백질 결합 서열; 및

[0086] 예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점;

[0087] 선택적으로, 외인성 이펙터(예를 들어, 치료용 외인성 이펙터)를 인코딩하는 핵산 서열; 및

[0088] 선택적으로, 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소

[0089] 를 포함하는, 유전 요소 작제물.

[0090] 19. 시스템으로서,

[0091] a) 유전 요소 또는 유전 요소 작제물인 제1 핵산으로서,

[0092] 예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점;

[0093] 선택적으로, 외인성 이펙터(예를 들어, 치료용 외인성 이펙터)를 인코딩하는 핵산; 및

[0094] 선택적으로, 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소

[0095] 를 포함하는, 제1 핵산;

[0096] b) 아넬로바이러스 ORF1 분자를 인코딩하는 제2 핵산

[0097] 을 포함하는, 시스템.

[0098] 20. 구현예 19에 있어서, 제1 핵산은 아넬로바이러스 ORF1 분자(예를 들어, 아넬로바이러스의 5' UTR 또는 GC-풍부 영역)에 특이적으로 결합하는 단백질 결합 서열을 추가로 포함하는, 시스템.

[0099] 21. 구현예 19 또는 20에 있어서, 아넬로바이러스 ORF2 분자를 인코딩하는 핵산 서열을 추가로 포함하는, 시스템.

[0100] 22. 구현예 21에 있어서, 아넬로바이러스 ORF2 분자를 인코딩하는 핵산 서열은 제3 핵산 상에 위치하는, 시스템.

[0101] 23. 구현예 19 내지 22 중 어느 하나에 있어서, AAV Rep2 분자(예를 들어, AAV Rep2 폴리펩티드, 예를 들어, AAV Rep2 단백질)를 인코딩하는 핵산 서열을 추가로 포함하는, 시스템.

[0102] 24. 구현예 23에 있어서, AAV REP2 분자를 인코딩하는 핵산 서열은 제4 핵산 상에 위치하는, 시스템.

[0103] 25. 구현예 19 내지 24 중 어느 하나에 있어서, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자 중 하나 이상(예를 들어, 전부)을 인코딩하는 하나 이상의 핵산 서열을 추가로 포함하는,

시스템.

[0104] 26. 구현예 25에 있어서, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자를 인코딩하는 핵산 서열은 제5 핵산 상에 위치하는, 시스템.

[0105] 27. 구현예 19 내지 26 중 어느 하나에 있어서, 제1, 제2, 제3, 제4, 및 제5 핵산 중 하나 이상(예를 들어, 전부)은 플라스미드인, 시스템.

[0106] 28. 구현예 19 내지 27 중 어느 하나에 있어서, 핵산은 혼합되거나 별도의 부피로 존재하는, 시스템.

[0107] 29. 구현예 19 내지 28 중 어느 하나에 있어서, 핵산은 세포, 예를 들어, 인간 세포, 예를 들어, 293 세포 또는 MOLT4 세포에 존재하는, 시스템.

[0108] 30. DNase-보호된 단백질성 복합체로서,

[0109] a) 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부; 및

[0110] b) 예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점을 포함하는 유전 요소

[0111] 를 포함하며,

[0112] 선택적으로 유전 요소는 (i) 외인성 이펙터를 인코딩하는 핵산 서열, 및/또는 (ii) 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소를 추가로 포함하는, DNase-보호된 단백질성 복합체.

[0113] 31. 구현예 30에 있어서,

[0114] 유전 요소에는 실질적으로 아넬로바이러스 서열이 없거나,

[0115] 유전 요소는 야생형 아넬로바이러스 게놈의 임의의 100개 뉴클레오티드 서열에 대하여 50% 초과 동일성을 갖는 100개 초과의 뉴클레오티드를 포함하지 않거나,

[0116] 유전 요소는 아넬로바이러스 5' UTR을 포함하지 않는, DNase-보호된 단백질성 복합체.

[0117] 32. DNase-보호된 단백질성 복합체로서,

[0118] a) 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부; 및

[0119] b) 유전 요소

[0120] 를 포함하며;

[0121] 유전 요소에는 실질적으로 아넬로바이러스 서열이 없고/없거나,

[0122] 유전 요소는 야생형 아넬로바이러스 게놈의 동일한 길이의 임의의 서열에 대하여 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 초과의 동일성을 갖는 10, 20, 30, 40, 50, 60, 70, 80, 90, 또는 100개 초과의 연속적인 뉴클레오티드를 포함하지 않고/않거나,

[0123] 유전 요소는 아넬로바이러스 5' UTR을 포함하지 않고;

[0124] 선택적으로 유전 요소는 (i) 외인성 이펙터를 인코딩하는 핵산 서열, 및/또는 (ii) 외인성 이펙터를 인코딩하는 핵산 서열에 작동적으로 연결된 프로모터 요소를 추가로 포함하는, DNase-보호된 단백질성 복합체.

[0125] 33. 구현예 32에 있어서, 유전 요소는 (iii) 제1 ITR, 예를 들어, 제1 AAV ITR을 추가로 포함하는, DNase-보호된 단백질성 복합체.

[0126] 34. 혼합물로서,

[0127] 아넬로바이러스 ORF1 분자, 및

[0128] 예를 들어, 제1 AAV 역위 말단 반복부(ITR)에 포함되는, AAV 복제 원점을 포함하는 핵산

[0129] 을 포함하는, 혼합물.

[0130] 35. 혼합물로서,

[0131] 아넬로바이러스 ORF1 분자, 및

[0132] 핵산(예를 들어, 유전 요소)

[0133] 를 포함하며,

[0134] 핵산에는 실질적으로 아넬로바이러스 서열이 없거나,

[0135] 핵산은 야생형 아넬로바이러스 게놈의 임의의 100개 뉴클레오티드 서열에 대하여 50% 초과 동일성을 갖는 100개 초과의 뉴클레오티드를 포함하지 않거나,

[0136] 핵산은 아넬로바이러스 5' UTR을 포함하지 않는, 혼합물.

[0137] 36. 구현예 34 또는 35에 있어서, 아넬로바이러스 ORF1 분자는 제1 AAV ITR을 포함하는 핵산에 결합되는, 혼합물.

[0138] 37. 구현예 34 내지 36 중 어느 하나에 있어서, 제1 AAV 복제 원점을 포함하는 핵산은 유전 요소, 예를 들어, 선행하는 구현예 중 어느 하나에 따른 유전 요소인, 혼합물.

[0139] 38. 복합체로서,

[0140] 선행하는 구현예 중 어느 하나에 따른 유전 요소, 및

[0141] 유전 요소에 결합된 캡시드 단백질(예를 들어, ORF1 분자)

[0142] 를 포함하는, 복합체.

[0143] 39. 구현예 34 내지 38 중 어느 하나에 있어서, 무세포 시스템 또는 실질적으로 무세포 조성물 중에 있는, 혼합물 또는 복합체.

[0144] 40. 구현예 38 또는 39에 있어서, 복합체는 세포, 예를 들어, 숙주 세포, 예를 들어, 헬퍼 세포 내에 있는, 복합체.

[0145] 41. 선행하는 구현예 중 어느 하나의 유전 요소 또는 유전 요소 작제물을 포함하는 세포.

[0146] 42. 구현예 41에 있어서, 인간 세포, 예를 들어, 293 세포, Expi293 세포, Expi293F 세포, 또는 MOLT-4 세포인, 세포.

[0147] 43. 선행하는 구현예 중 어느 하나의 벡터를 세포 내로 도입하는 단계를 포함하는, 외인성 이펙터를 표적 세포(예를 들어, 척추동물 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포)로 운반하는 방법.

[0148] 44. 선행하는 구현예 중 어느 하나의 벡터를 대상체 내로 도입하는 단계를 포함하는, 생물학적 활성 조절을 필요로 하는 대상체에서 생물학적 활성을 조절하는 방법.

[0149] 45. 선행하는 구현예 중 어느 하나의 벡터를 대상체 내로 도입하는 단계를 포함하는, 질환 또는 장애의 치료 또는 예방을 필요로 하는 대상체에서 질환 또는 장애를 치료 또는 예방하는 방법.

[0150] 46. 선행하는 구현예 중 어느 하나의 벡터를 대상체 내로 도입하는 단계를 포함하며, 외인성 이펙터는 감염원(infectious agent)(예를 들어, 바이러스 또는 박테리아) 유래의 항원을 포함하는, 백신접종을 필요로 하는 대상체에게 백신접종을 하는 방법.

[0151] 47. 구현예 43 내지 46 중 어느 하나에 있어서, 표적 세포는 인간 세포, 예를 들어, 293 세포, Expi293 세포, Expi293F 세포, 또는 MOLT-4 세포인, 방법.

[0152] 48. 구현예 43 내지 46 중 어느 하나에 있어서, 표적 세포는 동물(예를 들어, 농업 동물, 예를 들어, 소, 양, 돼지, 염소, 말, 들소, 또는 낙타) 유래의 세포인, 방법.

[0153] 49. 구현예 48에 있어서, 동물은 조류 동물(예를 들어, 칠면조, 닭, 메추라기, 에뮤, 또는 타조)인, 방법.

[0154] 50. 구현예 43내지 49 중 어느 하나에 있어서, 표적 세포는 생체 내 또는 시험관 내에 있는, 방법.

[0155] 51. 구현예 43 내지 50 중 어느 하나에 있어서, 벡터는 시험관 내, 생체 외, 또는 생체 내에서 세포와 접촉되는, 방법.

[0156] 52. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 DNAse I에 의한 분해로부터 실질적으로 보호되는, 벡터.

[0157] 53. 선행하는 구현예 중 어느 하나에 있어서, 외인성 이펙터가 mKate로 대체되는 경우, 벡터는 mKate를 시험관 내에서 복수의 표적 세포(예를 들어, MOLT4 세포)에 운반하여, 벡터와 접촉된 세포의 적어도 약 10%, 20%, 30%, 40%, 50%, 또는 60%가 배경 수준 초과의 형광을 가질 수 있게 되며, 배경 수준은, 예를 들어, 실시예 5의 유세포 분석 검정에서, ORF1이 결여된 다른 유사한 벡터와 접촉된 세포의 가장 형광성인 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 또는 10%를 제외한 전부를 배제한 수준인, 벡터.

[0158] 54. 선행하는 구현예 중 어느 하나에 있어서, 외인성 이펙터는 나노루시퍼라제(nanoLuciferase)로 대체되는 경우, 벡터는 나노루시퍼라제를 시험관 내에서 복수의 표적 세포(예를 들어, Vero 세포 또는 MCF7 세포)에 운반하여, 배경 수준의 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 또는 15배의 발광을 나타내는 벡터와 접촉되는 세포의 집단을 생성하며, 배경 수준은, 예를 들어, 실시예 4 또는 8의 발광 검정에서, 벡터와 접촉되지 않은 다른 유사한 세포의 발광인, 벡터.

[0159] 55. 선행하는 구현예 중 어느 하나에 있어서, 예를 들어, 실시예 5에 따라, CsCl 구배 상에서 약 1.2 내지 1.4 g/ml의 밀도로 침강하는, 벡터.

[0160] 56. 벡터를 제조하는 방법으로서,

[0161] (a) 상기 구현예 중 어느 하나의 유전 요소를 포함하는 숙주 세포를 제공하는 단계, 및

[0162] (b) 단백질성 외부(예를 들어, 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부)의 유전 요소의 봉입에 적합한 조건 하에서 숙주 세포를 인큐베이션하는 단계

[0163] 를 포함하며, 이를 통해 벡터를 제조하는, 방법.

[0164] 57. 벡터를 제조하는 방법으로서,

[0165] (a) 상기 구현예 중 어느 하나의 시스템을 포함하는 숙주 세포를 제공하는 단계, 및

[0166] (b) 단백질성 외부(예를 들어, 아넬로바이러스 ORF1 분자를 포함하는 단백질성 외부)의 유전 요소의 봉입에 적합한 조건 하에서 숙주 세포를 인큐베이션하는 단계

[0167] 를 포함하며, 이를 통해 벡터를 제조하는, 방법.

[0168] 58. 구현예 56 또는 57에 있어서, 숙주 세포의 용해를 포함하는, 방법.

[0169] 59. 구현예 56 내지 58 중 어느 하나에 있어서, 숙주 세포의 상청액으로부터 벡터를 수득하는 단계를 포함하는, 방법.

[0170] 60. 구현예 56 내지 59 중 어느 하나에 있어서, 숙주 세포는 아넬로바이러스 ORF2 분자, AAV REP2 분자, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자 중 하나 이상(예를 들어, 전부)을 인코딩하는 하나 이상의 추가적인 핵산을 추가로 포함하는, 방법.

[0171] 61. 치료용 조성물을 제조하는 방법으로서,

[0172] (a) (i) AAV 복제 원점,

[0173] (ii) 치료용 이펙터(예를 들어, 치료용 RNA 또는 폴리펩티드)를 인코딩하는 서열에 작동 가능하게 연결된 프로모터,

[0174] (iii) 선택적으로 아넬로바이러스 ORF1 분자를 인코딩하는 서열,

[0175] (iv) 선택적으로 아넬로바이러스 ORF2 분자를 인코딩하는 서열,

[0176] (v) 선택적으로 Rep 단백질(예를 들어, AAV Rep 단백질, 예를 들어, AAV Rep2 단백질)을 인코딩하는 서열, 및

[0177] (vi) 선택적으로 하나 또는 복수의 헬퍼 단백질, 예를 들어, 아데노바이러스 헬퍼 단백질, 예를 들어, E2A 분자, 아데노바이러스 E4 분자, 및/또는 아데노바이러스 VARNA 분자를 인코딩하는 서열

[0178] 을 포함하는 외인성 DNA를 포함하는 하나 또는 복수의 숙주 세포를 제공하는 단계;

[0179] (b) 치료용 이펙터를 인코딩하는 서열에 작동 가능하게 연결된 프로모터를 포함하는 유전 요소를 봉입(예를 들어, 캡시드화)하기에 충분한 수의 ORF1 분자를 포함하는 단백질성 외부(예를 들어, 캡시드)를 포함하는 벡터(예를 들어, 아넬로벡터, 예를 들어, 바이러스 입자)의 형성에 적합한 조건 하에서 하나 또는 복수의 숙주 세포를

배양하며; 선택적으로 유전 요소는 환형 또는 선형인, 단계;

[0180] (c) 세포 배양물로부터 단계 (b)에서 생성된 벡터를 농축, 예를 들어, 정제하는 단계

[0181] 를 포함하며, 이를 통해 치료용 조성물을 제조하는, 방법.

[0182] 62. 구현예 61에 있어서,

[0183] (d) 내독소, 마이코플라스마, 숙주 세포 핵산(예를 들어, 숙주 세포 DNA 및/또는 숙주 세포 RNA), 동물-유래 과정 불순물(예를 들어, 혈청 알부민 또는 트립신), 복제-적격 입자, 유리 바이러스 캡시드 단백질, 외래성 물질(adventitious agent), 및 응집물로부터 선택되는 하나 이상의 불순물에 대하여 정제된 바이러스 입자를 평가하는 단계;

[0184] (e) 선택적으로 단계 (d)에서 검출되는 경우 바이러스 입자로부터 하나 이상의 불순물을 감소시키거나 제거하는 단계; 및

[0185] (f) 선택적으로 인간에게 투여하기 위해 정제된 바이러스 입자를 제형화하는 단계

[0186] 를 추가로 포함하며, 이를 통해 치료용 조성물을 제조하는, 방법.

[0187] 63. 구현예 61 또는 62에 있어서, (a) (i) 내지 (vi)의 외인성 DNA는 하나의 숙주 세포에서 제공되는, 방법.

[0188] 64. 구현예 61 내지 63 중 어느 하나에 있어서, (a) (i) 내지 (vi)의 외인성 DNA는 복수의 숙주 세포에서 제공되는, 방법.

[0189] 65. 구현예 61 내지 64 중 어느 하나에 있어서, (a) (i) 및 (ii)의 외인성 DNA는 하나의 숙주 세포에서 제공되고 (a) (iii) 내지 (vi)의 외인성 DNA는 제2 숙주 세포에서 제공되는, 방법.

[0190] 66. 구현예 61 내지 65 중 어느 하나에 있어서, (a) (i) 및 (ii)의 외인성 DNA는 숙주 세포 염색체의 일부가 아닌, 방법.

[0191] 67. 구현예 61 내지 66 중 어느 하나에 있어서, (a) (i) 및 (ii)의 외인성 DNA는 동일한 핵산, 예를 들어, 환형 DNA 또는 선형 DNA의 일부인, 방법.

[0192] 68. 구현예 61 내지 67 중 어느 하나에 있어서, (a) (i) 및 (ii)의 외인성 DNA는 선행하는 구현예 중 어느 하나에 따른 유전 요소인, 방법.

[0193] 69. 구현예 61 내지 68 중 어느 하나에 있어서, (a) (iii)의 하나 이상의 외인성 DNA는 숙주 세포 염색체에 통합되는, 방법.

[0194] 70. 구현예 61 내지 69 중 어느 하나에 있어서, (a) (iv) 내지 (vi) 중 임의의 것의 하나 이상의 외인성 DNA는 존재하는 경우 숙주 세포 염색체에 통합되는, 방법.

[0195] 71. 구현예 61 내지 70 중 어느 하나에 있어서, (a) (iii)의 하나 이상의 외인성 DNA는 플라스미드의 일부인, 방법.

[0196] 72. 구현예 61 내지 71 중 어느 하나에 있어서, (a) (iv) 내지 (vi) 중 임의의 것의 하나 이상의 외인성 DNA는 존재하는 경우 플라스미드의 일부인, 방법.

[0197] 73. 구현예 61 내지 72 중 어느 하나에 있어서, 숙주 세포는 포유동물 세포(예를 들어, 인간 세포, 예를 들어, HEK293 세포)인, 방법.

[0198] 74. 구현예 61 내지 73 중 어느 하나에 있어서, 숙주 세포는 무한증식 세포인, 방법.

[0199] 75. 치료용 조성물을 제조하는 방법으로서,

[0200] (a) (i) AAV 복제 원점 및 치료용 이펙터(예를 들어, 치료용 RNA 또는 폴리펩티드)를 인코딩하는 서열에 작동가능하게 연결된 프로모터를 포함하는 유전 요소, 및

[0201] (ii) 복수의 ORF1 분자(예를 들어, 동일한 ORF1 분자의 복수의 카피)

[0202] 를 포함하는 용액을 제공하는 단계;

[0203] (b) 유전 요소를 봉입(예를 들어, 캡시드화)하기에 충분한 수의 ORF1 분자를 포함하는 단백질성 외부(예를 들어, 캡시드)를 포함하는 벡터(예를 들어, 아넬로벡터, 예를 들어, 바이러스 입자)의 형성에 적합한 조건 하에

서 용액을 인큐베이션하는 단계; 및

[0204] (c) 선택적으로 용액으로부터 단계 (b)에서 생성된 벡터를 농축, 예를 들어, 정제하는 단계

[0205] 를 포함하며, 이를 통해 치료용 조성물을 제조하는, 방법.

[0206] 76. 구현예 75에 있어서, 유전 요소는

[0207] (iii) 선택적으로 아넬로바이러스 ORF1 분자를 인코딩하는 서열,

[0208] (iv) 선택적으로 아넬로바이러스 ORF2 분자를 인코딩하는 서열,

[0209] (v) 선택적으로 AAV REP2 서열을 인코딩하는 서열

[0210] (vi) 선택적으로 하나 또는 복수의 헬퍼 단백질, 예를 들어, 아데노바이러스 헬퍼 단백질, 예를 들어, E2A 분자, 아데노바이러스 E4 분자, 및/또는 아데노바이러스 VARNA 분자를 인코딩하는 서열

[0211] 을 사용하여 제조된, 방법.

[0212] 77. 구현예 61 내지 76 중 어느 하나에 있어서, 단계 (b)에서 생성된 벡터는 선행하는 구현예 중 어느 하나의 벡터인, 방법.

[0213] 78. 선행하는 구현예 중 어느 하나의 유전 요소 또는 유전 요소 작제물을 포함하는 숙주 세포(예를 들어, 척추동물 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포).

[0214] 79. 구현예 78에 있어서, 아넬로바이러스 ORF1 분자 또는 아넬로바이러스 ORF1 분자를 인코딩하는 핵산을 추가로 포함하는, 숙주 세포.

[0215] 80. 구현예 78 또는 79에 있어서, 아넬로바이러스 ORF2 분자, AAV REP2 분자, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자 중 하나 이상(예를 들어, 전부)을 추가로 포함하는, 숙주 세포.

[0216] 81. 구현예 78 내지 80 중 어느 하나에 있어서, 아넬로바이러스 ORF2 분자, AAV REP2 분자, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자 중 하나 이상(예를 들어, 전부)을 인코딩하는 하나 이상의 핵산을 추가로 포함하는, 숙주 세포.

[0217] 82. 선행하는 구현예 중 어느 하나의 벡터를 포함하는, 숙주 세포.

[0218] 83. 구현예 78 내지 82 중 어느 하나의 숙주 세포를 제조하는 방법으로서, 유전 요소를 세포 내로 도입하는 단계를 포함하고, 예를 들어, 유전 요소를 도입하는 단계는 유전 요소의 생성을 가능하게 하는 조건 하에서 유전 요소 작제물을 세포 내로 도입하는 것을 포함하는, 방법.

[0219] 84. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는, 예를 들어, 제2 AAV 역위 말단 반복부(ITR)에 포함되는, 제2 AAV 복제 원점을 추가로 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0220] 85. 구현예 84에 있어서, 제2 ITR은 제1 ITR에 대해 역방향으로 배향된, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0221] 86. 구현예 84에 있어서, 제2 ITR은 제1 ITR에 대해 동일한 배향을 갖는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0222] 87. 구현예 84 내지 86 중 어느 하나에 있어서, 제2 ITR은 제1 ITR과 동일한 서열을 갖는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0223] 88. 구현예 84 내지 86 중 어느 하나에 있어서, 제2 ITR은 제1 ITR에 대해 하나 이상의 서열 차이를 갖는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0224] 89. 구현예 84 내지 88 중 어느 하나에 있어서, 외인성 이펙터를 인코딩하는 핵산 서열은 제1 ITR과 제2 ITR 사이에 위치하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0225] 90. 선행하는 구현예 중 어느 하나에 있어서, 제1 AAV ITR은 서열번호 1051 내지 1059 중 임의의 것의 서열, 또는 이에 대하여 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0226] 91. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 선형인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0227] 92. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 환형인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0228] 93. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소 작제물은 환형인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0229] 94. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소 작제물은 선형인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0230] 95. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 길이가 약 500 내지 1000, 1000 내지 1500, 1500 내지 2000, 2000 내지 2500, 2500 내지 3000, 3000 내지 3500, 3500 내지 4000, 4000 내지 4100, 4100 내지 4200, 4200 내지 4300, 4300 내지 4400, 4400 내지 4500, 4500 내지 4600, 4600 내지 4700, 4700 내지 4800, 4800 내지 4900, 4900 내지 5000, 5000 내지 5500, 5500 내지 6000, 또는 6000 내지 7000개 뉴클레오티드인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0231] 96. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 길이가 적어도 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 또는 6000개 뉴클레오티드인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0232] 97. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 DNA를 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0233] 98. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 DNA로 이루어지는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0234] 99. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% DNA로 이루어지는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0235] 100. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 단일 가닥 DNA 또는 이중 가닥 DNA인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0236] 101. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소 작제물은 단일 가닥 DNA 또는 이중 가닥 DNA인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0237] 102. 선행하는 구현예 중 어느 하나에 있어서, 환화된 이중-가닥 DNA를 사용하여 생성되며, 예를 들어, 환화된 DNA는 시험관 내 환화에 의해 생성된, 유전 요소.

[0238] 103. 선행하는 구현예 중 어느 하나에 있어서, 프로모터 요소는 아넬로바이러스에 대하여 내인성인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0239] 104. 선행하는 구현예 중 어느 하나에 있어서, 프로모터 요소는 AAV에 대하여 내인성인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0240] 105. 선행하는 구현예 중 어느 하나에 있어서, 프로모터 요소는 아넬로바이러스에 대하여 외인성인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0241] 106. 선행하는 구현예 중 어느 하나에 있어서, 프로모터 요소는 AAV에 대하여 외인성인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0242] 107. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소 작제물의 복제에, 예를 들어, 박테리아 세포에서의 복제에 적합한 백본 영역을 포함하는, 유전 요소 작제물.

[0243] 108. 선행하는 구현예 중 어느 하나에 있어서, 백본 영역은 복제 원점 및 선택 가능한 마커 중 하나 또는 둘 다를 포함하는, 유전 요소 작제물.

[0244] 109. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 아넬로바이러스 5' UTR, 아넬로바이러스 GC-풍부 영역, 및 아넬로바이러스 3' UTR, 또는 이들의 임의의 조합을 추가로 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0245] 110. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 표 A1, B1, B3, C1, E1, F1, F3, 또는 F5 중 임의의 것의 아넬로바이러스 5' UTR을 추가로 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0246] 111. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 표 A1, B1, B3, C1, E1, F1, F3, 또는 F5 중 임의의 것의 아넬로바이러스 GC-풍부 영역을 추가로 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0247] 112. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 표 A1, B1, B3, C1, E1, F1, F3, 또는 F5 중 임의의 것의 아넬로바이러스 3' UTR을 추가로 포함하는, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0248] 113. 선행하는 구현예 중 어느 하나에 있어서, 외인성 이펙터를 인코딩하는 핵산 서열은 길이가 약 20 내지 50, 50 내지 100, 100 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 1,000개 뉴클레오티드인, 유전 요소, 유전 요소 작제물, 시스템, 세포, 방법, 또는 벡터.

[0249] 114. 선행하는 구현예 중 어느 하나에 있어서, 이펙터는 miRNA를 포함하는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0250] 115. 선행하는 구현예 중 어느 하나에 있어서, 이펙터, 예를 들어, miRNA는 숙주 유전자를 표적화하고, 예를 들어, 유전자의 발현을 조절하고, 예를 들어, 유전자의 발현을 증가시키거나 감소시키는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0251] 116. 선행하는 구현예 중 어느 하나에 있어서, 이펙터는 miRNA를 포함하고, 숙주 유전자의 발현을 감소시키는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0252] 117. 선행하는 구현예 중 중 어느 하나에 있어서, 이펙터는 길이가 약 20 내지 200, 30 내지 180, 40 내지 160, 50 내지 140, 또는 60 내지 120개 뉴클레오티드인 핵산 서열을 포함하는, 유전 요소, 핵산 작제물, CA벡터, 복합체, 방법, 또는 숙주 세포.

[0253] 118. 선행하는 구현예 중 어느 하나에 있어서, 이펙터를 인코딩하는 핵산 서열은 길이가 약 20 내지 200, 30 내지 180, 40 내지 160, 50 내지 140, 또는 60 내지 120개 뉴클레오티드인, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0254] 119. 선행하는 구현예 중 어느 하나에 있어서, 이펙터를 인코딩하는 서열은 적어도 약 100개 뉴클레오티드의 크기를 갖는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0255] 120. 선행하는 구현예 중 어느 하나에 있어서, 이펙터를 인코딩하는 서열은 약 100 내지 약 5000개 뉴클레오티드의 크기를 갖는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0256] 121. 선행하는 구현예 중 어느 하나에 있어서, 이펙터를 인코딩하는 서열은 약 100 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, 또는 1500 내지 2000개 뉴클레오티드의 크기를 갖는, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0257] 122. 선행하는 구현예 중 어느 하나에 있어서, 유전 요소는 DNA인, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0258] 123. 선행하는 구현예 중 어느 하나에 있어서, 벡터는 복제-결핍인, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0259] 124. 선행하는 구현예 중 어느 하나에 있어서,

[0260] (i) 유전 요소에는 실질적으로 아넬로바이러스 서열이 없고/없거나,

[0261] (ii) 유전 요소는 야생형 아넬로바이러스 게놈의 동일한 길이의 임의의 서열에 대하여 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 초과의 동일성을 갖는 10, 20, 30, 40, 50, 60, 70, 80, 90, 또는 100개 초과의 연속적인 뉴클레오티드를 포함하지 않고/않거나,

[0262] (iii) 유전 요소는 아넬로바이러스 5' UTR을 포함하지 않는,

[0263] 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0264] 125. 선행하는 구현예 중 어느 하나에 있어서, 벡터는 바이러스 입자인, 유전 요소, 유전 요소 작제물, 벡터, 혼합물, 복합체, 방법, 또는 숙주 세포.

[0265] 126. 선행하는 구현예 중 어느 하나의 벡터, 및 약제학적으로 허용 가능한 담체 및/또는 부형제를 포함하는 약제학적 조성물.

[0266] 본 발명의 다른 특징, 목적 및 이점이 설명 및 도면, 및 청구범위로부터 명백해질 것이다.

[0267] 달리 정의되지 않는 한, 본 명세서에 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 분야의 당업자에 의해 일반적으로 이해되는 바와 동일한 의미를 갖는다. 본 명세서에 언급된 모든 간행물, 특허 출원, 특허 및 다른 참고문헌은 전문이 참조로 포함된다. 추가적으로, 물질, 방법 및 실시예는 단지 예시적인 것일 뿐이며, 제한하려는 것이 아니다.

도면의 간단한 설명

[0268] 도 1은 N-말단 3xFlag-태깅된 아넬로바이러스 ORF1 단백질의 발현을 나타내는 웨스턴 블롯이다. 상단, 알파토르크바이러스 Ring1 ORF1(91 kda). 중간, 베타토르크바이러스 Ring2 ORF1(79 kda). 하단, 감마토르크바이러스 Ring4 ORF1(82 kda).

도 2는 Cap-무함유 AAV-Rep 발현 작제물에 의한 ITR-측접 페이로드의 복제를 나타내는 일련의 다이어그램이다. hrGFP 및 pHelper에 대해 조사된 서던 블록이 도시되어 있다. 레인 1 내지 3은 형질감염되지 않은 대조군 DNA를 함유하고, 레인 4 내지 6은 상이한 Rep 작제물로 형질감염된 세포 유래의 총 DNA를 함유한다. 화살표는 pHelper 플라스미드, pITR-hrGFP 플라스미드, 및 복제된 ITR-hrGFP DNA에 대한 밴드 위치를 나타낸다.

도 3a 및 3b는 CsCl 선형 구배로부터 nLuc 이식유전자를 포함하는 R2 아넬로벡터의 정제를 나타내는 일련의 그래프이다. nLuc 리포터 유전자에 대한 qPCR을 통해 벡터를 정량화하였다. (A) 아넬로바이러스 ORF1, ORF2 단백질 및 nLuc 이식유전자를 함유하는 입자 모두의 트랜스-발현을 통해 벡터를 생성하였다. (B) 아넬로바이러스 ORF1 및 ORF2가 트랜스로 발현되지 않았을 때 nLuc 이식유전자의 정량화.

도 4는 R2-nLuc 아넬로벡터를 이용한 비-인간 영장류 세포의 형질도입을 나타내는 그래프이다. Vero 세포를 24웰 플레이트에서 웰당 1e5개 세포로 시딩하였다. MOI 0.4(qPCR 역가에 기반함)에서 벡터의 첨가를 통해 형질도입을 수행하였다. 2일 후 루시퍼라제 검정을 수행하였다.

도 5는 R2-nLuc 아넬로벡터를 이용한 인간 세포의 형질도입을 나타내는 그래프이다. IGR-OV1 세포를 24웰 플레이트에서 웰당 1e5개 세포로 시딩하였다. MOI 0.4(qPCR 역가에 기반함)에서 벡터의 첨가를 통해 형질도입을 수행하였다. 2일 후 루시퍼라제 검정을 수행하였다.

도 6은 아넬로바이러스/AAV 벡터의 생성 및 MOLT4 세포에서 성공적인 형질도입을 나타내는 일련의 다이어그램이다. 상단 패널은 Expi-293 세포에서 mKate 페이로드를 변화시키는 아넬로/AAV 하이브리드 벡터를 생성하고 벡터를 MOLT4 세포로 형질도입한 다음, mKate 형광에 대한 유세포 분석을 위한 예시적인 작업 흐름을 나타낸다. 하단 좌측 패널은 역위 말단 반복부(ITR)가 측접한 mKate-인코딩 유전자를 포함하는 유전 요소를 봉입하는 ORF1 단백질 캡시드를 포함하는 아넬로/AAV 하이브리드 벡터의 다이어그램을 나타낸다. 하단 우측 패널은 표시된 플라스미드를 사용하여 생성된 벡터로 형질도입된 MOLT4 세포의 유세포 분석 결과를 나타낸다.

도 7a 및 7b는 조작된 Ring2 아넬로바이러스 DNA가 AAV Rep 단백질을 통해 복제함을 나타내는 일련의 다이어그램이다. (A) Rep 결합 모티프(RBM) 및 말단 분해 부위(TRS)를 포함하여, AAV 복제에 필요한 최소 영역을 포함하는 Ring2 dsDNA 게놈을 나타내는 다이어그램. (B) AAV-Rep 플라스미드 및 WT Ring2 게놈 또는 Ring2 + RBM/TRS DNA(도 7a에 나타낸 바와 같음)의 표시된 조합으로 형질감염된 Expi-293 세포로부터 수득된 DNA 샘플로부터의 선형 플라스미드 및 Dpn1 분해 생성물을 나타내는 서던 블롯.

도 8a 및 8b는 페이로드로서 인간 성장 호르몬(hGH)을 인코딩하는 아넬로벡터에 의한 포유동물 세포주의 형질도입을 나타내는 일련의 그래프이다. (A) IGR-OV1 세포를 AAV Rep 벡터, pHelper 벡터, 및 (i) hGH를 인코딩하는 Ring2 캡시드 아넬로벡터, (ii) hGH를 인코딩하는, hGH를 인코딩하는 Ring9 캡시드 아넬로벡터, (iii) hGH를 인코딩하는 AAV2 캡시드 바이러스 벡터(양성 대조군), 또는 (iv) 캡시드가 없는 음성 대조군 중 하나 이상으로 형

질감염시켰다. 제0일, 제2일, 및 제3일에 ELISA에 의해 hGH 수준을 정량화하였다. (A) Vero 세포를 AAV Rep 벡터, pHelper 벡터, 및 (i) hGH를 인코딩하는 Ring2 캡시드 아넬로벡터, (ii) hGH를 인코딩하는, hGH를 인코딩하는 Ring9 캡시드 아넬로벡터, (iii) hGH를 인코딩하는 AAV2 캡시드 바이러스 벡터(양성 대조군), 또는 (iv) 캡시드가 없는 음성 대조군 중 하나 이상으로 형질감염시켰다. 제0일, 제2일, 및 제3일에 ELISA에 의해 hGH 수준을 정량화하였다.

도 9는 AAV Rep의 존재 또는 부재 하에(각각 +AAV Rep 또는 -AAV Rep) 생성된 Ring2-AAV ITR-nLuc 아넬로벡터로 형질감염된 293F 세포로부터의 세포 용해물에서 나노-루시퍼라제 발광을 나타내는 그래프이다.

도 10a 내지 10l은 본 명세서에 기재된 바와 같은 아넬로벡터에 대한 유전 요소를 생성하는 데 사용될 수 있는 예시적인 유전 요소 작제물의 개략도를 나타내는 일련의 다이어그램이다. 개별 개략도는 하기 표 61에 표시된 플라스미드에 해당한다. 검정색 = Ring2 게놈 서열(예를 들어, 본 명세서에 기재된 바와 같음); 녹색 = 외인성 이펙터 서열; 파란색 = AAV 복제 원점.

본 발명의 구현예에 대한 하기의 상세한 설명은 첨부된 도면과 함께 읽으면 더 잘 이해될 것이다. 본 발명의 예시 목적을 위하여, 본 명세서에 예시된 구현예가 도면에 나타나 있다. 그러나, 본 발명이 도면에 나타낸 구현예의 정확한 배열 및 수단으로 제한되지 않음을 이해하여야 한다. 특허 또는 출원 파일은 컬러로 작성된 적어도 하나의 도면을 포함한다. 컬러 도면(들)이 포함된 본 특허 또는 특허 출원 공개의 사본은 요청 및 필요한 수수료의 지불 시 사무국이 제공할 것이다.

발명을 실시하기 위한 구체적인 내용

[0269] 정의

[0270] 본 발명은 특정 구현예에 대해 그리고 특정 도면을 참조하여 설명될 것이지만 본 발명은 이에 제한되지 않고, 청구범위에 의해서만 제한된다. 이하에 제시된 용어는 일반적으로 달리 지시되지 않는 한 그의 일반적인 의미로 이해되어야 한다.

[0271] 발명을 실시하기 위한 구체적인 내용 및 청구범위에서 "포함하는"이라는 용어가 사용되는 경우, 이는 다른 요소를 배제하지 않는다. 본 발명의 목적을 위하여, "로 이루어지는"이라는 용어는 "로 구성되는"이라는 용어의 바람직한 구현예인 것으로 간주된다. 이하에서 그룹이 적어도 특정 개수의 구현예를 포함하는 것으로 정의되는 경우, 이는 또한 바람직하게는 이들 구현예만으로 이루어지는 그룹을 개시하는 것으로 이해되어야 한다.

[0272] 단수 명사를 언급할 때, 부정 관사 또는 정관사, 예를 들어, 하나의("a", "an") 또는 상기("the")가 사용되는 경우, 이는 그 밖의 어떤 것이 특별히 언급되지 않는 한, 그 명사의 복수형을 포함한다.

[0273] 용어 "치료, 조절 등을 위한 화합물, 조성물, 생성물 등"은 표시된 치료, 조절 등의 목적에 적합한 화합물, 조성물, 생성물 등 그 자체를 지칭하는 것으로 이해되어야 한다. 용어 "치료, 조절 등을 위한 화합물, 조성물, 생성물 등"은 추가로 일 구현예로서, 이러한 화합물, 조성물, 생성물 등이 치료, 조절 등에 사용하기 위한 것임을 개시한다.

[0274] 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 구현예 및 청구범위의 동등한 개시로서 간주된다. 따라서, 구현예 또는 청구범위가 "질환을 앓는 것으로 의심되는 인간 또는 동물을 치료하는 데 사용하기 위한 화합물"을 지칭한다면, 이는 또한 "질환을 앓는 것으로 의심되는 인간 또는 동물을 치료하기 위한 의약의 제조에서의 화합물의 용도" 또는 "질환을 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 표시된 치료, 조절 등의 목적에 적합한 화합물, 조성물, 생성물 등 그 자체를 지칭하는 것으로 이해되어야 한다.

[0275] 기간, 값, 수 등의 하기의 예가 괄호 안에 제공된다면, 이는 괄호 안에 언급된 예가 구현예를 구성할 수 있다는 표시로서 이해되어야 한다. 예를 들어, "구현예에서, 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 구현예는 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.

[0276] 본 명세서에 사용되는 바와 같이, 용어 "증폭"은 핵산 분자 또는 이의 일부(예를 들어, 유전 요소 또는 유전 요소 영역)의 하나 이상의 추가의 카피를 생성하기 위한 핵산 분자 또는 이의 일부의 복제를 지칭한다. 일부 구현예에서, 증폭은 핵산 서열의 부분 복제를 초래한다. 일부 구현예에서, 증폭은 회전환 복제를 통해 발생한다.

[0277] 본 명세서에 사용되는 바와 같이, 용어 "아넬로벡터"는 단백질성 외부 내에 봉입된 유전 요소, 예를 들어, 환형 DNA를 포함하는 비히클을 지칭하며, 예를 들어, 유전 요소는 단백질성 외부에 의해 DNAse I을 이용한 분해로부터 실질적으로 보호된다. 본 명세서에 사용되는 바와 같이, "합성 아넬로벡터"는 일반적으로, 천연 발생이 아닌, 예를 들어, 야생형 바이러스(예를 들어, 본 명세서에 기재된 바와 같은 야생형 아넬로바이러스)에 비하여 상이한 서열을 갖는 아넬로벡터를 지칭한다. 일부 구현예에서, 합성 아넬로벡터는 조작되거나 또는 재조합이며, 예를 들어, 야생형 바이러스 게놈(예를 들어, 본 명세서에 기재된 바와 같은 야생형 아넬로바이러스 게놈)에 비하여 차이 또는 변형을 포함하는 유전 요소를 포함한다. 일부 구현예에서, 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 포함한다. 유전 요소가, 예를 들어, 숙주 세포로의 유입 이전에 단백질성 외부에 보유되거나 DNAse I을 이용한 분해로부터 보호되는 한, 예를 들어, (예를 들어, 단백질성 외부를 물, 이온, 퓁티드 또는 소분자에 대해 투과 가능하게 만드는) 갭 또는 불연속부가 단백질성 외부에 존재할 수 있다. 일부 구현예에서, 아넬로벡터는 정제되며, 예를 들어, 아넬로벡터는 이의 원래의 공급원으로부터 분리되고/분리되거나 다른 성분이 실질적으로 없다(50% 초과, 60% 초과, 70% 초과, 80% 초과, 90% 초과). 일부 구현예에서, 아넬로벡터는 유전 요소를 (예를 들어, 감염을 통해) 표적 세포로 도입할 수 있다. 일부 구현예에서, 아넬로벡터는 감염성 합성 아넬로바이러스 바이러스 입자이다.

[0278] 본 명세서에 사용되는 바와 같이, 용어 "항체 분자"는 단백질, 예를 들어, 적어도 하나의 면역글로불린 가변 도메인 서열을 포함하는 면역글로불린 쇄 또는 이의 단편을 지칭한다. 용어 "항체 분자"는 전장 항체 및 항체 단편(예를 들어, scFv)을 포함한다. 일부 구현예에서, 항체 분자는 다중특이적 항체 분자이고, 예를 들어, 항체 분자는 복수의 면역글로불린 가변 도메인 서열을 포함하며, 복수의 것들 중 제1 면역글로불린 가변 도메인 서열은 제1 에피토프에 대한 결합 특이성을 가지고, 복수의 것들 중 제2 면역글로불린 가변 도메인 서열은 제2 에피토프에 대한 결합 특이성을 갖는다. 구현예에서, 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.

[0279] 본 명세서에 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.

[0280] 본 명세서에 사용되는 바와 같은 "외인성" 작용제(예를 들어, 이펙터, 핵산(예를 들어, RNA), 유전자, 페이로드, 단백질)는 상응하는 야생형 바이러스, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스에 의해 포함되지 않거나, 이에 의해 인코딩되지 않는 작용제를 지칭한다. 일부 구현예에서, 외인성 작용제, 예컨대 천연 발생 단백질 또는 핵산에 비하여 (예를 들어, 삽입, 결실 또는 치환에 의해) 변경된 서열을 갖는 단백질 또는 핵산은 자연적으로 존재하지 않는다. 일부 구현예에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지 않는다. 일부 구현예에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지만, 바이러스에 대하여 외인성이다. 일부 구현예에서, 외인성 작용제는 숙주 세포에서 자연적으로 존재하지만, 요망되는 수준으로 또는 요망되는 시간에 존재하지 않는다.

[0281] 또 다른 작용제 또는 요소(예를 들어, 이펙터, 핵산 서열, 아미노산 서열)에 관하여 본 명세서에 사용되는 바와 같은, "이종" 작용제 또는 요소(예를 들어, 이펙터, 핵산 서열, 아미노산 서열)는 천연적으로, 예를 들어, 야생형 바이러스, 예를 들어, 아넬로바이러스에서 함께 발견되지 않는 작용제 또는 요소를 지칭한다. 일부 구현예에서, 이종 핵산 서열은 천연 발생 핵산 서열(예를 들어, 아넬로바이러스에서 천연적으로 발생하는 서열)과 동일한 핵산에 존재할 수 있다. 일부 구현예에서, 이종 작용제 또는 요소는 아넬로벡터의 다른(예를 들어, 나머지) 요소가 기반으로 하는 아넬로바이러스에 비하여 외인성이다.

[0282] 본 명세서에 사용되는 바와 같이, 용어 "유전 요소"는, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터를 형성하기 위해, 단백질성 외부이거나 단백질성 외부 내에 봉입될 수 있는(예를 들어, 단백질성 외부에 의해 DNAse I로부터 보호될 수 있는) 핵산 분자를 지칭한다. 유전 요소가 네이키드(naked) DNA로서 생성되고, 선택적으로 추가로 단백질성 외부 내로 조립될 수 있는 것이 이해된다. 또한, 아넬로벡터는 이의 유전 요소를 세포 내로 삽입하여, 유전 요소가 세포 내에 존재하게 하고, 단백질성 외부가 반드시 세포에 유입되는 것은 아니라는

점이 이해된다.

[0283] 본 명세서에 사용되는 바와 같이, "유전 요소 작제물"은 적어도 하나(예를 들어, 2개)의 유전 요소 서열(들), 또는 이의 단편을 포함하는 핵산 작제물(예를 들어, 플라스미드, 박미드, 코스미드, 또는 미니서클)을 지칭한다. 일부 구현예에서, 유전 요소 작제물은 적어도 하나의 전장 유전 요소 서열을 포함한다. 일부 구현예에서, 유전 요소는 전장 유전 요소 서열 및 부분 유전 요소 서열을 포함한다. 일부 구현예에서, 유전 요소는 2개 이상의 부분 유전 요소 서열(예를 들어, 5'에서 3' 순서로, 예를 들어, 도 27c에 나타낸 바와 같이, 3'-절단된 유전 요소 서열과 탠덤으로 배열된 5'-절단된 유전 요소 서열)을 포함한다.

[0284] 본 명세서에 사용되는 바와 같이, 용어 "유전 요소 영역"은 유전 요소의 서열을 포함하는 작제물의 영역을 지칭한다. 일부 구현예에서, 유전 요소 영역은 야생형 아넬로바이러스 서열, 또는 이의 단편에 대해 충분한 동일성을 갖는 서열을 포함하여 단백질성 외부에 의해 봉입됨으로써 아넬로벡터(예를 들어, 야생형 아넬로바이러스 서열 또는 이의 단편에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열)를 형성한다. 구현예에서, 유전 요소 영역은, 예를 들어, 본 명세서에 기재된 바와 같은 단백질 결합 서열(예를 들어, 본 명세서에 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 구현예에서, 유전 요소 영역은 회전환 복제를 겪을 수 있다. 일부 구현예에서, 유전 요소는 Rep 단백질 결합 부위를 포함한다. 일부 구현예에서, 유전 요소는 Rep 단백질 변위 부위를 포함한다. 일부 구현예에서, 유전 요소 영역을 포함하는 작제물은 단백질성 외부로 봉입되지 않지만, 작제물로부터 생성된 유전 요소는 단백질성 외부로 봉입될 수 있다. 일부 구현예에서, 유전 요소 영역을 포함하는 작제물은 벡터 백본을 추가로 포함한다.

[0285] 본 명세서에 사용되는 바와 같이, 용어 "역위 말단 반복부"("ITR")는 바이러스 Rep 분자(예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 단백질), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 폴리펩티드에 의해 주변 핵산 서열(또는 이의 일부)의 복제에 적합한 복제 원점을 포함하는 핵산 서열을 지칭한다. 일반적으로, ITR(또는 ITR이 유래된 바이러스 서열)은 이의 역 보체가 뒤따르는(예를 들어, 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 또는 100개 뉴클레오티드에 직접 인접하거나, 이에 의해 분리된) 뉴클레오티드의 인접 서열을 포함한다. 일부 예에서 ITR의 카피는 단일-가닥 바이러스 게놈(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 비-아넬로바이러스의 게놈, 예를 들어, AAV)의 게놈의 한쪽 또는 양쪽 종말단에 포함될 수 있다. ITR 서열은 헤어핀을 형성할 수 있다. ITR은, 예를 들어, 본 명세서에 기재된 바와 같은, Rep-결합 모티프(RBM) 및/또는 말단 분해 부위(TRS)를 포함할 수 있다. 일부 예에서, ITR 서열은, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터의 유전 요소에 존재한다. 일부 예에서, 아넬로벡터의 유전 요소에 존재하는 ITR은 유전 요소의 종말단(예를 들어, 5' 종말단 또는 3' 종말단)에 위치할 수 있다. 일부 예에서, 아넬로벡터의 유전 요소에 존재하는 ITR은 유전 요소의 종말단(예를 들어, 5' 종말단 또는 3' 종말단)에 위치하지 않을 수 있으며, 예를 들어, (예를 들어, 환형 유전 요소 또는 선형 유전 요소에서) 유전 요소의 5' 및 3' 말단에서 핵산 서열이 측접될 수 있다.

[0286] 본 명세서에 사용되는 바와 같이, 용어 "돌연변이체"는 게놈(예를 들어, 아넬로바이러스 게놈) 또는 이의 단편과 관련하여 사용될 때 상응하는 야생형 아넬로바이러스 서열에 대해 적어도 하나의 변화를 갖는 서열을 지칭한다. 일부 구현예에서, 돌연변이체 게놈 또는 이의 단편은 상응하는 야생형 아넬로바이러스 서열에 대해 적어도 하나의 단일 뉴클레오티드 다형성, 추가, 결실, 또는 프레임시프트를 포함한다. 일부 구현예에서, 돌연변이체 게놈 또는 이의 단편은 상응하는 야생형 아넬로바이러스 서열에 대해 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 구현예에서, 돌연변이체 게놈 또는 이의 단편은 상응하는 야생형 아넬로바이러스 서열에 대해 모든 아넬로바이러스 ORF(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 및 ORF1/2 전부)의 결실을 포함한다. 일부 구현예에서, 돌연변이체 게놈 또는 이의 단편은 상응하는 야생형 아넬로바이러스 서열에 대해 적어도 하나의 아넬로바이러스 비코딩 영역(예를 들어, 5' UTR, 3' UTR, 및/또는 GC-풍부 영역 중 하나 이상)의 결실을 포함한다. 일부 구현예에서, 돌연변이체 게놈 또는 이의 단편은 외인성 이펙터를 포함하거나 인코딩한다.

[0287] 본 명세서에 사용되는 바와 같이, 용어 "비-아넬로바이러스" 서열은 아넬로비리대 과(family *Anelloviridae*)로 분류되지 않는 바이러스 유래의 서열을 지칭한다. 비-아넬로바이러스 서열은 일반적으로, (i) 아넬로비리대 과로 분류되는 바이러스(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 알파토르크바이러스, 베타토르크바이러스, 또는 감마토르크바이러스)의 게놈, 유전자, 또는 비-코딩 기능적 요소(예를 들어, 복제 원점)와 동일한 핵산 서열을 포함하지 않고/않거나; 아넬로비리대 과로 분류되지 않는 바이러스 유래의 하나 이상의 단백질(예를 들어, 캡시드 단백질 또는 Rep 단백질)을 인코딩하지 않는다. 일부 예에서, 비-아넬로바이러스 서열은 아

넬로비리대 과로 분류되는 임의의 바이러스(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 알파토르크바이러스, 베타토르크바이러스, 또는 감마토르크바이러스)의 게놈, 유전자, 또는 비-코딩 기능적 요소(예를 들어, 복제 원점)에 대하여 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 또는 90% 이하의 서열 동일성을 갖는다. 일부 구현예에서, 비-아넬로바이러스 서열은 아넬로비리대 과로 분류되지 않는 바이러스 유래의 야생형 서열이다. 다른 구현예에서, 아넬로비리대 과로 분류되지 않는 바이러스 유래의 비-아넬로바이러스 서열은 바이러스의 게놈 유래의 하나 이상의 비-천연 발생 돌연변이를 포함한다. 일부 예에서, 비-아넬로바이러스 서열은 비-인간 유기체(예를 들어, 비-인간 영장류, 비-인간 포유동물, 또는 조류)를 감염시키는 바이러스 유래이다. 일부 예에서, 비-아넬로바이러스 서열은 인간을 감염시키는 바이러스 유래이다. 일부 예에서, 비-아넬로바이러스 서열은 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 및 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV)로 이루어진 군으로부터 선택되는 바이러스 유래이다.

[0288] "ORF 분자"는 아넬로바이러스 ORF 단백질(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 및/또는 ORF1/2 단백질), 또는 이의 기능적 단편의 활성 및/또는 구조적 특징을 갖는 폴리펩티드를 지칭한다. 총칭적으로 사용되는 경우(즉, "ORF 분자"), 폴리펩티드는 본 명세서에 기재된 임의의 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1 및/또는 ORF1/2), 또는 이의 기능적 단편의 활성 및/또는 구조적 특징을 포함할 수 있다. 특정 오픈-리딩 프레임을 표시하는 데 수식어와 함께 사용되는 경우(예를 들어, "ORF1 분자", "ORF2 분자", "ORF2/2 분자", "ORF2/3 분자", "ORF1/1 분자", 또는 "ORF1/2 분자"), 이는 일반적으로 폴리펩티드가 상응하는 아넬로바이러스 ORF 단백질 또는 이의 기능적 단편(예를 들어, "ORF1 분자"에 대해 하기 정의된 바와 같음)의 활성 및/또는 구조적 특징을 포함함을 의미한다. 예를 들어, "ORF2 분자"는 아넬로바이러스 ORF2 단백질 또는 이의 기능적 단편의 활성 및/또는 구조적 특징을 포함한다.

[0289] 본 명세서에 사용되는 바와 같이, 용어 "ORF1 분자"는 아넬로바이러스 ORF1 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)의 활성 및/또는 구조적 특징을 갖는 폴리펩티드 또는 이의 기능적 단편을 지칭한다. ORF1 분자는 일부 예에서, 하기 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, 본 명세서에 기재된 바와 같음, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 N22 도메인)의 구조 또는 활성을 포함하는 제3 영역, 및/또는 아넬로바이러스 C-말단 도메인(CTD)(예를 들어, 본 명세서에 기재된 바와 같음, 예를 들어, 본 명세서에 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, 아넬로벡터는 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, 이종 서열, 예를 들어, 초가변 영역(HVR), 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 HVR을 추가로 포함할 수 있다. 본 명세서에 사용되는 바와 같이, "아넬로바이러스 ORF1 단백질"은 아넬로바이러스 게놈(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 게놈)에 의해 인코딩된 ORF1 단백질을 지칭한다.

[0290] 본 명세서에 사용되는 바와 같이, 용어 "ORF2 분자"는 아넬로바이러스 ORF2 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF2 단백질)의 활성 및/또는 구조적 특징을 갖는 폴리펩티드 또는 이의 기능적 단편을 지칭한다. 본 명세서에 사용되는 바와 같이, "아넬로바이러스 ORF2 단백질"은 아넬로바이러스 게놈(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 게놈)에 의해 인코딩된 ORF2 단백질을 지칭한다.

[0291] 본 명세서에 사용되는 바와 같이, "복제 원점"은 Rep 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 바이러스 Rep 단백질, 예를 들어, 비-아넬로바이러스 Rep 단백질, 예를 들어, AAV Rep 단백질)의 존재 하에 DNA 복제를 촉진하는 서열을 포함하는 핵산 서열을 지칭한다. 일부 예에서, 핵산 분자(예를 들어, 본 명세서에 기재된 바와 같은 유전 요소) 내에 위치한 복제 원점은 Rep 분자의 존재 하에 복제 원점이 결여된 다른 유사한 핵산 분자보다 더 큰 정도로 유전 요소, 또는 이의 일부의 복제를 촉진한다. 일부 예에서, 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 비-아넬로바이러스 게놈, 예를 들어, AAV 게놈의, 역위 말단

반복부(ITR) 서열에 포함된다. 일부 예에서, 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 비-아넬로바이러스(예를 들어, AAV) 유래의, Rep-결합 모티프(RBM) 및/또는 말단 분해 부위(TRS) 중 하나 또는 둘 다를 포함한다. 다른 예에서, 복제 원점은 아넬로바이러스 복제 원점을 포함한다. 본 명세서에 사용되는 바와 같이, "AAV 복제 원점"은 AAV Rep 분자(예를 들어, AAV Rep 단백질)의 존재 하에 DNA 복제를 촉진하는 서열을 포함하는 핵산 서열을 지칭한다. 일부 예에서, AAV 복제 원점은 AAV Rep 분자(예를 들어, AAV Rep 단백질)에 의해 인식되고 결합된다. 일부 예에서, AAV 복제 원점은 말단 분해 부위(TRS)(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, AAV TRS) 및/또는 Rep-결합 모티프(RBM)(예를 들어, 본 명세서에 기재된 바와 같은, AAV RBM)를 포함한다. 일부 구현예에서, AAV 복제 원점은 AAV ITR에 위치한다.

[0292] 본 명세서에 사용되는 바와 같이, 용어 "단백질성 외부"는 주로(예를 들어, 50% 초과, 60% 초과, 70% 초과, 80% 초과, 90% 초과, 95% 초과, 96% 초과, 97% 초과, 98% 초과, 또는 99% 초과) 단백질인 외부 구성성분을 지칭한다.

[0293] 본 명세서에 사용되는 바와 같이, 용어 "조절 핵산"은 발현 생성물을 인코딩하는 DNA 서열의 발현, 예를 들어, 전사 및/또는 번역을 변형시키는 핵산 서열을 지칭한다. 구현예에서, 발현 생성물은 RNA 또는 단백질을 포함한다.

[0294] 본 명세서에 사용되는 바와 같이, 용어 "조절 서열"은 표적 유전자 생성물의 전사를 변형시키는 핵산 서열을 지칭한다. 일부 구현예에서, 조절 서열은 프로모터 또는 인핸서이다.

[0295] 본 명세서에 사용되는 바와 같이, 용어 "Rep 분자"는 바이러스 게놈 복제를 촉진하는 단백질, 예를 들어, 바이러스 단백질을 지칭한다. 일부 구현예에서, Rep 분자는, 예를 들어, 본 명세서에 기재된 바와 같은, 비-아넬로바이러스 Rep 단백질(예를 들어, AAV Rep 단백질)이다. 일부 구현예에서, Rep 분자는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 Rep 분자, 예를 들어, 아넬로바이러스 ORF2 분자이다. 본 명세서에 사용되는 바와 같이, "AAV Rep 분자"는 일반적으로 야생형 AAV Rep 단백질의 기능성을 갖는, 예를 들어, AAV RBM(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 야생형 AAV RBM, 또는 본 명세서에 기재된 바와 같은 RBM 컨센서스 서열을 갖는 RBM)에 결합하는 능력을 갖고 AAV RBM을 포함하는 핵산 분자의 복제를 유도하는 단백질을 지칭한다.

[0296] 본 명세서에 사용되는 바와 같이, 용어 "Rep-결합 모티프"("RBM")는 Rep 분자에 결합하는, 바이러스 게놈(예를 들어, 비-아넬로바이러스 게놈, 예를 들어, AAV 게놈) 유래의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 지칭한다. 일반적으로, RBM은 본 명세서에 기재된 바와 같은 RBM 서열(예를 들어, 본 명세서에 기재된 바와 같은 AAV RBM 서열)에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는다. 일부 예에서, RBM은 복제 원점에, 예를 들어, 아넬로벡터의 유전 요소에 포함된다. 일부 예에서, RBM은, 예를 들어, 본 명세서에 기재된 바와 같은, 말단 분해 부위(TRS)의 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 또는 30개 뉴클레오티드 내에 위치한다. 일부 예에서, RBM은 TRS로부터 약 13개 뉴클레오티드에 위치한다. 일부 예에서, RBM은 TRS에 대해 3'에 위치한다. 일부 예에서, RBM은 복제 원점으로 Rep 분자를 동원한다.

[0297] 본 명세서에 사용되는 바와 같이, "실질적으로 비-병원성인" 유기체, 입자 또는 구성성분은, 예를 들어, 숙주 유기체, 예를 들어, 포유동물, 예를 들어, 인간에서 허용 불가능한 질환 또는 병원성 병태를 야기하거나 유도하지 않는 유기체, 입자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로벡터) 또는 이의 구성성분을 지칭한다. 일부 구현예에서, 대상체에 대한 아넬로벡터의 투여는 치료 기준의 일부로서 허용 가능한 부차 반응 또는 부작용을 초래할 수 있다.

[0298] 본 명세서에 사용되는 바와 같이, 용어 "비-병원성"은 숙주 유기체, 예를 들어, 포유동물, 예를 들어, 인간에서 허용 불가능한 질환 또는 병원성 병태를 야기하거나 유도하지 않는 유기체 또는 이의 구성성분을 지칭한다.

[0299] 본 명세서에 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, 본 명세서에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로벡터 내의 유전 요소를 지칭한다. 일부 구현예에서, 유전 요소는, 예를 들어, 숙주 세포의 게놈 내로 검출 가능하게 통합되지 않는다. 일부 구현예에서, 게놈 내로의 유전 요소의 통합은 본 명세서에 기재된 바와 같은 기법, 예를 들어, 핵산 서열결정, PCR 검출 및/또는 핵산 혼성화를 사용하여 검출될 수 있다. 일부 구현예에서, 통합 빈도는, 예를 들어, 문헌[Wang et al., 2004, *Gene Therapy* 11: 711-721](본 명세서에 전문

이 참조로 포함됨)에 기재된 바와 같이, 자유 벡터에서 분리된 게놈 DNA의 정량적 겔 정제 분석에 의해 결정된다.

[0300] 본 명세서에 사용되는 바와 같이, "실질적으로 비-면역원성인" 유기체, 입자 또는 성분은 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, 본 명세서에 기재된 바와 같은, 예를 들어, 바이러스 또는 아넬로벡터) 또는 이의 성분을 지칭한다. 구현예에서, 실질적으로 비-면역원성인 유기체, 입자 또는 성분은 임상적으로 유의한 면역 반응을 생성하지 않는다. 구현예에서, 실질적으로 비-면역원성인 아넬로벡터는 아넬로바이러스 또는 아넬로벡터 유전 요소의 아미노산 서열을 포함하거나, 핵산 서열에 의해 인코딩된 단백질에 대하여 임상적으로 유의한 면역 반응을 생성하지 않는다. 구현예에서, 면역 반응(예를 들어, 요망되지 않는 또는 비표적화된 면역 반응)은 예를 들어, 문헌[Tsuda et al., 1999; *J. Virol. Methods* 77: 199-206](본 명세서에 참조로 포함됨)에 기재된 항-TTV 항체 검출 방법 및/또는 문헌[Kakkola et al., 2008; *Virology* 382: 182-189](본 명세서에 참조로 포함됨)에 기재된 항-TTV IgG 수준의 결정 방법에 따라 대상체에서 항체(예를 들어, 중화 항체) 존재 또는 수준(예를 들어, 항-아넬로벡터 항체의 존재 또는 수준, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터에 대한 항체의 존재 또는 수준)을 검정함으로써 검출된다. 아넬로바이러스 또는 이에 기초한 아넬로벡터에 대한 항체(예를 들어, 중화 항체)는 또한 항-바이러스 항체를 검출하기 위한 당업계의 방법, 예를 들어, 문헌[Calcedo et al., 2013; *Front. Immunol.* 4(341): 1-7](본 명세서에 참조로 포함됨)에 기재된 바와 같은 예를 들어, 항-AAV 항체의 검출 방법에 의해 검출될 수 있다.

[0301] 본 명세서에 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 이종인 서열을 포함하는 하위서열, 예를 들어, 상이한 아넬로바이러스 유래의 상응하는 하위서열)에 의해 대체될 수 있다.

[0302] 본 명세서에 사용되는 바와 같이, 용어 "말단 분해 부위"("TRS")는, 예를 들어, 본 명세서에 기재된 바와 같은, 바이러스의 게놈의 TRS 서열(예를 들어, 본 명세서에 기재된 바와 같은 AAV TRS 서열)에 대하여 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 지칭한다. 일부 예에서, TRS는 Rep 분자에 의해(예를 들어, rep 분자의 엔도뉴클레아제 활성을 통해) 절단된다. 일부 예에서, Rep 분자에 의한 TRS의 절단은 TRS를 포함하는 핵산 분자의 복제를 위한 3' 하이드록실 말단을 생성한다. 일부 예에서, TRS는 복제 원점에, 예를 들어, 아넬로벡터의 유전 요소에 포함된다. 일부 예에서, TRS는, 예를 들어, 본 명세서에 기재된 바와 같은, Rep-결합 모티프(RBM)의 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 또는 30개 뉴클레오티드 내에 위치한다. 일부 예에서, TRS는 RBM으로부터 약 13개 뉴클레오티드에 위치한다. 일부 예에서, TRS는 RBM에 대해 5'에 위치한다.

[0303] 본 명세서에 사용되는 바와 같은, "치료", "치료하는" 및 이의 동족어(cognate)는 질환, 병리학적 병태, 또는 장애를 개선하거나, 호전시키거나, 안정화시키거나, 예방하거나 치유하기 위한 의도를 갖는 대상체의 의학적 관리를 지칭한다. 이 용어는 적극 치료(질환, 병리학적 병태, 또는 장애를 개선하고자 하는 치료), 원인 치료(연관 질환, 병리학적 병태, 또는 장애의 원인을 대상으로 하는 치료), 완화 치료(증상의 경감을 위해 설계된 치료), 예방적 치료(연관 질환, 병리학적 병태, 또는 장애의 예방, 최소화, 또는 이의 발생을 부분적으로 또는 완전히 억제하고자 하는 치료); 및 지지 치료(또 다른 치료법을 보충하기 위해 이용되는 치료)를 포함한다.

[0304] 본 발명은 일반적으로 아넬로벡터, 예를 들어, 합성 아넬로벡터, 아넬로벡터의 투여 방법, 및 이의 용도에 관한 것이다. 본 개시내용은 아넬로벡터, 아넬로벡터를 포함하는 조성물, 및 아넬로벡터의 제조 또는 사용 방법을 제공한다. 아넬로벡터는 일반적으로, 예를 들어, 치료제를 진핵 세포로 운반하기 위한 운반 비히클로서 유용하다. 일반적으로, 아넬로벡터는 단백질성 외부 내에 봉입된 (예를 들어, 이펙터, 예를 들어, 외인성 이펙터 또는 내인성 이펙터를 인코딩하는) 핵산 서열을 포함하는 유전 요소를 포함할 것이다. 아넬로벡터는 아넬로바이러스 서열(예를 들어, 본 명세서에 기재된 바와 같음)에 비한 서열(예를 들어, 본 명세서에 기재된 바와 같은 영역 또는 도메인)의 하나 이상의 결실을 포함할 수 있다. 아넬로벡터는, 예를 들어, 세포를 포함하는 대상체에서 질환 또는 장애를 치료하기 위하여, 유전 요소 또는 그 내에 인코딩된 이펙터(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 폴리펩티드 또는 핵산 이펙터)를 진핵 세포 내로 운반하기 위한 실질적으로 비-면역원성인 비

히클로서 사용될 수 있다.

[0305] **목차**

[0306] I. 아넬로벡터를 제조하기 위한 조성물 및 방법

[0307] A. 아넬로벡터의 구성성분 및 조립

[0308] i. 아넬로벡터의 조립을 위한 ORF1 분자

[0309] ii. 아넬로벡터의 조립을 위한 ORF2 분자

[0310] iii. 단백질 구성성분의 생성

[0311] B. 유전 요소 작제물

[0312] i. 비-아넬로바이러스 서열(예를 들어, AAV 서열)

[0313] ii. 플라스미드

[0314] iii. 환형 핵산 작제물

[0315] iv. 시험관 내 환화

[0316] v. 탠덤 작제물

[0317] vi. 시스/트랜스 작제물

[0318] vii. 발현 카세트

[0319] viii. 유전 요소 작제물의 설계 및 생성

[0320] C. 이펙터

[0321] D. 숙주 세포

[0322] i. 숙주 세포 내로의 유전 요소의 도입

[0323] ii. 시스 또는 트랜스로 단백질(들)을 제공하는 방법

[0324] iii. 헬퍼, 예를 들어, 비-아넬로바이러스 헬퍼

[0325] iv. 예시적인 세포 유형

[0326] E. 배양 조건

[0327] F. 수확

[0328] G. 시험관 내 조립 방법

[0329] H. 농축 및 정제

[0330] II. 아넬로벡터

[0331] A. 아넬로바이러스

[0332] B. ORF1 분자

[0333] C. ORF2 분자

[0334] D. 유전 요소, 예를 들어, 비-아넬로바이러스 서열을 포함하는 유전 요소

[0335] E. 단백질 결합 서열

[0336] F. 5' UTR 영역

[0337] G. GC-풍부 영역

[0338] H. 이펙터

[0339] I. 조절 서열

[0340] J. 복제 단백질

[0341] K. 기타 서열

[0342] L. 단백질성 외부

[0343] III. 핵산 작제물

[0344] IV. 조성물

[0345] V. 사용 방법

[0346] VI. 투여/운반

[0347] I. 아넬로벡터를 제조하기 위한 조성물 및 방법

[0348] 본 개시내용은, 일부 양태에서, 이펙터를 운반하기 위한 아넬로벡터 및 이의 방법을 제공한다. 일부 구현예에서, 아넬로벡터 또는 이의 구성성분은 하기 기재된 바와 같이 제조될 수 있다. 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은 유전 요소 또는 유전 요소 작제물을 생성하는 데 사용될 수 있다. 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은 하나 이상의 아넬로바이러스 ORF 분자(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2 분자, 또는 이의 기능적 단편 또는 스플라이스 변이체)를 생성하는 데 사용될 수 있다. 일부 구현예에서, 본 명세서에 기재된 조성물 및 방법은, 예를 들어, 숙주 세포에서 단백질성 외부 또는 이의 구성성분(예를 들어, ORF1 분자)을 생성하는 데 사용될 수 있다. 일부 구현예에서, 아넬로벡터 또는 이의 구성성분은, 예를 들어, 미국 가출원 63/038,483에 기재된 바와 같이 탠덤 작제물을 사용하여 제조될 수 있으며, 이는 본 명세서에 전문이 참조로 포함된다. 일부 구현예에서, 아넬로벡터 또는 이의 구성성분은, 예를 들어, 미국 가출원 63/038,603에 기재된 바와 같이 기재된 바와 같이, 박미드/곤충 세포 시스템을 사용하여 제조될 수 있으며, 이는 본 명세서에 전문이 참조로 포함된다.

[0349] 이론에 구속되고자 하지 않지만, 회전환 증폭은 유전 요소 영역에 대해 5'에(또는 유전 요소 영역의 5' 영역 내에) 위치한 Rep 결합 부위(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 5' UTR을 포함함, 예를 들어, 헤어핀 루프 및/또는 복제 원점을 포함함)에 대한 Rep 단백질 결합을 통해 발생할 수 있다. 그 다음 Rep 단백질은 유전 요소 영역을 통해 진행되어, 유전 요소의 합성을 초래할 수 있다. 그 다음 유전 요소는 환화된 다음, 단백질성 외부 내에 봉입되어 아넬로벡터를 형성할 수 있다.

[0350] ***아넬로벡터의 구성성분 및 조립***

[0351] 본 명세서의 조성물 및 방법은 아넬로벡터를 생성하는 데 사용될 수 있다. 본 명세서에 기재된 바와 같이, 아넬로벡터는 일반적으로 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함함) 내에 봉입된 유전 요소(예를 들어, 본 명세서에 기재된 바와 같은 5' UTR 영역을 포함하는, 예를 들어, 단일-가닥 환형 DNA 분자)를 포함한다. 일부 구현예에서, 유전 요소는 아넬로바이러스 ORF(예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2 중 하나 이상)를 인코딩하는 하나 이상의 서열을 포함한다. 본 명세서에 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/US2018/037379 또는 PCT/US19/65995(이들 각각은 본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같이, 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. 구현예에서, 유전 요소는 아넬로바이러스 ORF1, 또는 이의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 핵산(예를 들어, 아넬로바이러스 ORF1 분자 또는 이의 스플라이스 변이체 또는 기능적 단편)에 의해 인코딩된 폴리펩티드를 포함한다.

[0352] 일부 구현예에서, 아넬로벡터는 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같음) 내에 유전 요소(예를 들어, 본 명세서에 기재된 바와 같음)를 봉입함으로써 조립된다. 일부 구현예에서, 유전 요소는 숙주 세포(예를 들어, 본 명세서에 기재된 바와 같음)에서 단백질성 외부 내에 봉입된다. 일부 구현예에서, 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 구현예에서, 숙주 세포는 아넬로바이러스 ORF1 분자, 예를 들어, 아넬로바이러스 ORF1 폴리펩티드의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 야생형 아넬로바이러스 ORF1 단백질 또는 야생형 아넬로바이러스 ORF1 핵산에 의해 인코

딩된 폴리펩티드)을 인코딩하는 핵산 서열을 포함한다. 구현예에서, 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열은 숙주 세포에 포함된 핵산 작제물(예를 들어, 플라스미드, 바이러스 벡터, 바이러스, 미니서클, 박미드, 또는 인공 염색체)에 포함된다. 구현예에서, 아넬로바이러스 ORF1 분자를 인코딩하는 핵산 서열은 숙주 세포의 게놈에 통합된다.

[0353] 일부 구현예에서, 숙주 세포는 유전 요소 및/또는 유전 요소의 서열을 포함하는 핵산 작제물을 포함한다. 일부 구현예에서, 핵산 작제물은 플라스미드, 바이러스 핵산, 미니서클, 박미드, 또는 인공 염색체로부터 선택된다. 일부 구현예에서, 유전 요소는 핵산 작제물로부터 절제되고, 선택적으로, 이중-가닥 형태에서 단일-가닥 형태로 (예를 들어, 변성에 의해) 전환된다. 일부 구현예에서, 유전 요소는 핵산 작제물의 주형 서열에 기반한 중합효소에 의해 생성된다. 일부 구현예에서, 중합효소는 유전 요소 서열의 단일-가닥 카피를 생성하며, 이는 본 명세서에 기재된 바와 같은 유전 요소를 형성하기 위해 선택적으로 환화될 수 있다. 다른 구현예에서, 핵산 작제물은 시험관 내에서 유전 요소의 핵산 서열의 환화에 의해 생성된 이중-가닥 미니서클이다. 구현예에서, 시험관 내-환화(IVC) 미니서클은 숙주 세포 내로 도입되고, 여기서 본 명세서에 기재된 바와 같은 단백질성 외부에 봉입하기에 적합한 단일-가닥 유전 요소로 전환된다.

[0354] ***예를 들어, 아넬로벡터의 조립을 위한 ORF1 분자***

[0355] 아넬로벡터는, 예를 들어, 단백질성 외부 내에 유전 요소를 봉입함으로써 제조될 수 있다. 아넬로벡터의 단백질성 외부는 일반적으로 아넬로바이러스 ORF1 핵산(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 분자 또는 이의 스플라이스 변이체 또는 기능적 단편)에 의해 인코딩된 폴리펩티드를 포함한다. ORF1 분자는 일부 구현예에서 하기 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. 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 아르기닌-풍부 영역, 젤리-롤 영역, N22 도메인, 초가변 영역, 및/또는 C-말단 도메인 중 하나 이상(예를 들어, 1, 2, 3, 4개, 또는 5개 전부)을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 젤리-롤 영역(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 아르기닌-풍부 영역(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 N22 도메인(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 초가변 영역(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다. 일부 구현예에서, 단백질성 외부는 아넬로바이러스 ORF1 C-말단 도메인(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다.

[0356] 일부 구현예에서, 아넬로벡터는 ORF1 분자 및/또는 ORF1 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)의 구조적 특징 및/또는 활성을 갖는 폴리펩티드 또는 이의 기능적 단편을 포함한다. 일부 구현예에서, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)에 비해 절단을 포함한다. 일부 구현예에서, 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 700개 아미노산만큼 절단된다. 일부 구현예에서, ORF1 분자는, 예를 들어, 본 명세서에 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, 본 명세서에 기재된 바와 같은, 예를 들어, 유전 요소)에 결합할 수 있다. 일부 구현예에서, ORF1 분자는 세포의 핵에 국소화된다. 특정 구현예에서, ORF1 분자는 세포의 핵소체에 국소화된다.

[0357] 이론에 구속되고자 하지 않지만, ORF1 분자는 다른 ORF1 분자에 결합하여, 예를 들어, (예를 들어, 본 명세서에 기재된 바와 같은) 단백질성 외부를 형성할 수 있다. 이러한 ORF1 분자는 캡시드를 형성하는 능력을 갖는 것으로서 설명될 수 있다. 일부 구현예에서, 단백질성 외부는 핵산 분자(예를 들어, 본 명세서에 기재된 바와 같은 조성물 또는 작제물을 사용하여 생성된, 예를 들어, 본 명세서에 기재된 바와 같은 유전 요소)를 봉입할 수 있다. 일부 구현예에서, 복수의 ORF1 분자는, 예를 들어, 단백질성 외부를 생성하기 위해, 다량체를 형성할 수 있다. 일부 구현예에서, 다량체는 동종다량체일 수 있다. 다른 구현예에서, 다량체는 이종다량체일 수 있다.

[0358] 일부 구현예에서, 본 명세서에 기재된 바와 같은 ORF1 분자를 포함하는 제1 복수의 아넬로벡터가 대상체에게 투

여된다. 일부 구현예에서, 본 명세서에 기재된 ORF1 분자를 포함하는 제2 복수의 아넬로벡터가 제1 복수의 투여 후 대상체에게 후속적으로 투여된다. 일부 구현예에서 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터에 의해 포함되는 ORF1 분자와 동일한 아미노산 서열을 갖는 ORF1 분자를 포함한다. 일부 구현예에서 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터에 의해 포함되는 ORF1 분자에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 아미노산 서열 동일성을 갖는 ORF1 분자를 포함한다.

[0359] ***예를 들어, 아넬로벡터의 조립을 위한 ORF2 분자***

[0360] 본 명세서에 기재된 조성물 또는 방법을 사용하여 아넬로벡터를 생성하는 것은 아넬로바이러스 ORF2 분자(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이의 스플라이스 변이체 또는 기능적 단편의 발현을 수반할 수 있다. 일부 구현예에서, 아넬로벡터는 ORF2 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편, 및/또는 ORF2 분자를 인코딩하는 핵산, 또는 이의 스플라이스 변이체 또는 기능적 단편을 포함한다. 일부 구현예에서, 아넬로벡터는 ORF2 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편, 및/또는 ORF2 분자를 인코딩하는 핵산, 또는 이의 스플라이스 변이체 또는 기능적 단편을 포함하지 않는다. 일부 구현예에서, 아넬로벡터를 생성하는 것은 ORF2 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편의 발현을 포함하지만, ORF2 분자는 아넬로벡터에 포함되지 않는다.

[0361] ***단백질 구성성분의 생성***

[0362] 아넬로벡터의 단백질 구성성분, 예를 들어, ORF1은, 예를 들어, 본 명세서에 기재된 바와 같은, 다양한 방식으로 생성될 수 있다. 일부 구현예에서, 예를 들어, 단백질성 외부를 포함하여 아넬로벡터의 단백질 구성성분은 유전 요소를 단백질성 외부에 패키징하여 아넬로벡터를 생성하는 동일한 숙주 세포에서 생성된다. 일부 구현예에서, 예를 들어, 단백질성 외부를 포함하여 아넬로벡터의 단백질 구성성분은 유전 요소 및/또는 유전 요소 작제물(예를 들어, 본 명세서에 기재된 바와 같음)을 포함하지 않는 세포에서 생성된다.

[0363] *바큘로바이러스 발현 시스템*

[0364] 바이러스 발현 시스템, 예를 들어, 바큘로바이러스 발현 시스템은, 예를 들어, 본 명세서에 기재된 바와 같은, (예를 들어, 아넬로벡터의 생성을 위해) 단백질을 발현하는 데 사용될 수 있다. 바큘로바이러스는 환형의 수퍼코일 이중-가닥 DNA 게놈을 갖는 막대 형상의 바이러스이다. 바큘로바이러스의 속은 알파바큘로바이러스(인시목(*Lepidoptera*)에서 단리된 뉴클레오폴리헤드로바이러스(NPV)), 베타바큘로바이러스(인시목에서 단리된 그래뉼로바이러스(GV)), 감마바큘로바이러스(막시목(*Hymenoptera*)에서 단리된 NPV) 및 델타바큘로바이러스(쌍시목(*Diptera*)에서 단리된 NPV)를 포함한다. GV는 통상적으로 외피당 단지 하나의 뉴클레오캡시드를 함유하는 반면, NPV는 통상적으로 외피당 단일(SNPV) 뉴클레오캡시드 또는 다중(MNPV) 뉴클레오캡시드를 함유한다. 외피가 있는 비리온은 GV의 과립소 매트릭스 및 NPV의 다면체에서 추가로 폐색된다. 바큘로바이러스는 통상적으로 용해 및 폐색 수명 주기를 둘 다 갖는다. 일부 구현예에서, 용해 및 폐색 수명 주기는 바이러스 복제의 3가지 단계, 즉 초기, 후기, 및 극후기 단계 전체에 걸쳐 독립적으로 나타난다. 일부 구현예에서, 초기 단계 동안, 바이러스 DNA 복제는 숙주 세포 내부로의 바이러스 진입, 초기 바이러스 유전자 발현 및 숙주 유전자 발현 기구의 차단 후에 발생한다. 일부 구현예에서, 후기 단계에서 바이러스 DNA 복제를 코딩하는 후기 유전자가 발현되고, 바이러스 입자가 조립되며, 세포외 바이러스(EV)가 숙주 세포에 의해 생성된다. 일부 구현예에서, 극후기 단계에서 다면체 및 p10 유전자가 발현되고, 폐색된 바이러스(OV)는 숙주 세포에 의해 생성되며, 숙주 세포가 용해된다. 바큘로바이러스는 곤충 종을 감염시키므로, 바큘로바이러스는 바큘로바이러스-허용 곤충 세포 또는 유충에서 외인성 단백질을 생성하는 데 생물학적 작용제로서 사용될 수 있다. 바큘로바이러스의 상이한 분리주, 예컨대, 아우토그라파 칼리포르니카 다중 핵다각체 바이러스(*Autographa californica* multiple nuclear polyhedrosis virus; AcMNPV) 및 봄빅스 모리(누에) 핵다각체 바이러스(*Bombyx mori* (silkworm) nuclear polyhedrosis virus; BmNPV)가 외인성 단백질 발현에 사용될 수 있다. 다양한 바큘로바이러스 발현 시스템은, 예를 들어, ThermoFisher로부터 구매 가능하다.

[0365] 일부 구현예에서, 본 명세서에 기재된 단백질(예를 들어, 아넬로바이러스 ORF 분자, 예를 들어, ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2, 또는 이의 기능적 단편 또는 스플라이스 변이체)은 본 명세서에 기재된 하나 이상의 구성성분을 포함하는 바큘로바이러스 발현 벡터(예를 들어, 박미드)를 사용하여 발현될 수 있다. 예를 들어, 바큘로바이러스 발현 벡터는 선택 가능한 마커(예를 들어, kanR), 복제 원점(예를 들어, 박테리아 복제 원점 및 곤충 세포 복제 원점 중 하나 또는 둘 다), 재조합효소 인식 부위(예를 들어, att 부위), 및 프로모터 중 하나 이상(예를 들어, 전부)을 포함할 수 있다. 일부 구현예에서, 바큘로바이러스 발현 벡터(예를 들어, 본 명세서에 기재된 바와 같은 박미드)는, 바큘로바이러스 폐색체를 인코딩하는 천연 발생 야생형 다면체

유전자를 본 명세서에 기재된 단백질을 인코딩하는 유전자로 대체함으로써 생성될 수 있다. 일부 구현예에서, 본 명세서에 기재된 단백질을 인코딩하는 유전자는 바큘로바이러스 프로모터를 함유하는 바큘로바이러스 발현 벡터(예를 들어, 본 명세서에 기재된 바와 같은 박미드)로 클로닝된다. 일부 구현예에서, 바큘로바이러스 벡터는 하나 이상의 비-바큘로바이러스 프로모터, 예를 들어, 포유동물 프로모터 또는 아넬로바이러스 프로모터를 포함한다. 일부 구현예에서, 본 명세서에 기재된 단백질을 인코딩하는 유전자는 공여자 벡터(예를 들어, 본 명세서에 기재된 바와 같음)로 클로닝된 다음, 빈 바큘로바이러스 발현 벡터(예를 들어, 빈 박미드)와 접촉되어, 본 명세서에 기재된 단백질을 인코딩하는 유전자는 공여자 벡터로부터 바큘로바이러스 발현 벡터(예를 들어, 박미드)로 (예를 들어, 상동 재조합 또는 트랜스포사제 활성에 의해) 전달된다. 일부 구현예에서, 바큘로바이러스 프로모터에는 비필수 다면체 유전자 좌 유래의 바큘로바이러스 DNA가 측접되어 있다. 일부 구현예에서, 본 명세서에 기재된 단백질은 바이러스 복제의 극후기 단계에서 AcNPV 다면체 프로모터의 전사 제어 하에 있다. 일부 구현예에서, 곤충 세포에서 바큘로바이러스 발현에 사용하기에 적합한 강력한 프로모터는 바큘로바이러스 p10 프로모터, 다면체(polh) 프로모터, p6.9 프로모터 및 캡시드 단백질 프로모터를 포함하지만 이에 제한되지 않는다. 곤충 세포에서 바큘로바이러스 발현에 사용하기에 적합한 약한 프로모터는 바큘로바이러스의 ie1, ie2, ie0, et1, 39K(pp31로도 알려짐) 및 gp64 프로모터를 포함한다.

[0366] 일부 구현예에서, 재조합 바큘로바이러스는 바큘로바이러스 게놈(예를 들어, 야생형 또는 돌연변이 바큘로바이러스 게놈)과 전달 벡터 사이의 상동 재조합에 의해 생성된다. 일부 구현예에서, 본 명세서에 기재된 단백질을 인코딩하는 하나 이상의 유전자는 전달 벡터로 클로닝된다. 일부 구현예에서, 전달 벡터는 비필수 유전자 좌, 예를 들어, 다면체 유전자 유래의 DNA가 측접된 바큘로바이러스 프로모터를 추가로 함유한다. 일부 구현예에서, 본 명세서에 기재된 단백질을 인코딩하는 하나 이상의 유전자는 바큘로바이러스 게놈과 전달 벡터 사이의 상동 재조합에 의해 바큘로바이러스 게놈에 삽입된다. 일부 구현예에서, 바큘로바이러스 게놈은 하나 이상의 독특한 부위에서 선형화된다. 일부 구현예에서, 선형화된 부위는 본 명세서에 기재된 단백질을 인코딩하는 유전자를 바큘로바이러스 게놈으로 삽입하기 위한 표적 부위 근처에 위치한다. 일부 구현예에서, 유전자, 예를 들어, 다면체 유전자로부터 하류의 바큘로바이러스 게놈의 단편이 누락된 선형화된 바큘로바이러스 게놈이 상동 재조합에 사용될 수 있다. 일부 구현예에서, 바큘로바이러스 게놈 및 전달 벡터는 곤충 세포로 동시-형질감염된다. 일부 구현예에서, 재조합 바큘로바이러스를 생성하는 방법은 본 명세서에 기재된 하나 이상의 단백질을 인코딩하는 유전자를 함유하는 전달 벡터와 상동 재조합을 수행하기 위한 바큘로바이러스 게놈을 준비하는 단계 및 전달 벡터 및 바큘로바이러스 게놈 DNA를 곤충 세포에 동시-형질감염시키는 단계를 포함한다. 일부 구현예에서, 바큘로바이러스 게놈은 전달 벡터의 영역에 대해 상동성인 영역을 포함한다. 이러한 상동성 영역은 바큘로바이러스 게놈과 전달 벡터 사이의 재조합 가능성을 향상시킬 수 있다. 일부 구현예에서, 전달 벡터의 상동성 영역은 프로모터의 상류 또는 하류에 위치한다. 일부 구현예에서, 상동 재조합을 유도하기 위해, 바큘로바이러스 게놈, 및 전달 벡터는 약 1:1 내지 10:1의 중량비로 혼합된다.

[0367] 일부 구현예에서, 재조합 바큘로바이러스는 Tn7을 이용한 부위-특이적 전위를 포함하는 방법에 의해 생성되며, 예를 들어, 이에 의해 본 명세서에 기재된 단백질을 인코딩하는 유전자는 박미드 DNA에 삽입되고, 예를 들어, 박테리아, 예를 들어, 이. 콜라이(*E. coli*)(예를 들어, DH 10Bac 세포)에서 증식된다. 일부 구현예에서, 본 명세서에 기재된 단백질을 인코딩하는 유전자는 pFASTBAC® 벡터로 클로닝되고, 미니-*att*Tn7 표적 부위를 갖는 박미드 DNA를 함유하는, 적격 세포, 예를 들어, DH10BAC® 적격 세포로 형질전환된다. 일부 구현예에서, 바큘로바이러스 발현 벡터, 예를 들어, pFASTBAC® 벡터는 프로모터, 예를 들어, 이중 프로모터(예를 들어, 다면체 프로모터, p10 프로모터)를 가질 수 있다. 구매 가능한 pFASTBAC® 공여자 플라스미드는 pFASTBAC 1, pFASTBAC HT, 및 pFASTBAC DUAL을 포함한다. 일부 구현예에서, 재조합 박미드 DNA-함유 콜로니를 확인하고 박미드 DNA를 단리하여 곤충 세포를 형질감염시킨다.

[0368] 일부 구현예에서, 바큘로바이러스 벡터는 헬퍼 핵산과 함께 곤충 세포 내로 도입된다. 도입은 동시적이거나 순차적일 수 있다. 일부 구현예에서, 헬퍼 핵산은 하나 이상의 바큘로바이러스 단백질을 제공하여, 예를 들어, 바큘로바이러스 벡터의 패키징을 촉진한다.

[0369] 일부 구현예에서, 곤충 세포에서 (예를 들어, 상동 재조합에 의해) 생성된 재조합 바큘로바이러스는 확장되고 재조합 단백질 발현을 위해 (예를 들어, 중간 대수 성장기에서) 곤충 세포를 감염시키는 데 사용된다. 일부 구현예에서, 박테리아, 예를 들어, 이. 콜라이에서 부위-특이적 전위에 의해 생성된 재조합 박미드 DNA는 곤충 세포를 형질감염제, 예를 들어, Cellfectin® II로 형질감염시키는 데 사용된다. 바큘로바이러스 발현 시스템에 대

한 추가 정보는 미국 특허 출원 14/447,341, 14/277,892, 및 12/278,916에 논의되어 있으며, 이들은 본 명세서에 참조로 포함된다.

[0370] *곤충 세포 시스템*

[0371] 본 명세서에 기재된 단백질은, 예를 들어, 상기 기재된 바와 같이, 재조합 바큘로바이러스 또는 박미드 DNA로 감염되거나 형질감염된 곤충 세포에서 발현될 수 있다. 일부 구현예에서, 곤충 세포는 스포돕테라 프루기페르다(*Spodoptera frugiperda*)로부터 유래된 Sf9 및 Sf21 세포 및 트리코플러시아 니(*Trichoplusia ni*)로부터 유래된 Tn-368과 High Five™ BTI-TN-5B1-4 세포(또한 Hi5 세포로도 지칭됨)를 포함한다. 일부 구현예에서, 번데기 밤나방(pupal fall army worm)인 스포돕테라 프루기페르다의 난소로부터 유래된 곤충 세포주 Sf21 및 Sf9는 바큘로바이러스 발현 시스템을 사용하여 재조합 단백질의 발현에 사용될 수 있다. 일부 구현예에서, Sf21 및 Sf9 곤충 세포는 구매 가능한 혈청-보충 또는 무혈청 배지에서 배양될 수 있다. 곤충 세포의 배양에 적합한 배지는 Grace's Supplemented(TNM-FH), IPL-41, TC-100, Schneider's Drosophila, SF-900 II SFM, 및 EXPRESS-FIVE™ SFM을 포함한다. 일부 구현예에서, 일부 무혈청 배지 제형은 인산염 완충 시스템을 이용하여 배양 및 재조합 단백질 생성 둘 다를 위해 배양물 pH를 6.0 내지 6.4의 범위로 유지한다(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 구현예에서, 다양한 곤충 세포주를 배양하기 위해 6.0 내지 6.8의 pH가 사용될 수 있다. 일부 구현예에서, 곤충 세포는 통기 하에 25° 내지 30℃의 온도에서 현탁액으로 또는 단층으로 배양된다. 곤충 세포에 대한 추가 정보는, 예를 들어, 미국 특허 출원 14/564,512 및 14/775,154에 논의되어 있으며, 이들 각각은 본 명세서에 참조로 포함된다.

[0372] *포유동물 세포 시스템*

[0373] 일부 구현예에서, 본 명세서에 기재된 단백질은, 예를 들어, 본 명세서에 기재된 바와 같이, 단백질을 인코딩하는 벡터로 감염되거나 형질감염된 동물 세포주에서 시험관 내에서 발현될 수 있다. 본 개시내용의 맥락에서 고려되는 동물 세포주는 돼지 세포주, 예를 들어, 돼지 신장 상피 세포주 PK-15 및 SK, 단골수성 세포주 3D4/31 및 고환 세포주 ST(이에 제한되지 않음)와 같은 무한증식 돼지 세포주를 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 구현예는 상피 세포주, 즉 상피 계통 세포의 세포주인 동물 세포주를 사용한다. 본 명세서에 기재된 단백질을 발현하기에 적합한 세포주는 인간 또는 영장류 기원의 세포주, 예컨대 인간 또는 영장류 신장 암종 세포주를 포함하지만 이에 제한되지 않는다.

[0374] ***예를 들어, 아넬로벡터의 조립을 위한 유전 요소 작제물***

[0375] 본 명세서에 기재된 바와 같은 아넬로벡터의 유전 요소는 유전 요소 영역 및 선택적으로 벡터 백본과 같은 다른 서열을 포함하는 유전 요소 작제물로부터 생성될 수 있다. 일반적으로, 유전 요소 작제물은 아넬로바이러스 5' UTR(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다. 유전 요소 작제물은 유전 요소가 단백질성 외부 내에 봉입될 수 있는 숙주 세포 내로 유전 요소의 서열을 운반하기에 적합한 임의의 핵산 작제물일 수 있다. 일부 구현예에서, 유전 요소 작제물은 프로모터를 포함한다. 일부 구현예에서, 유전 요소 작제물은 선형 핵산 분자이다. 일부 구현예에서, 유전 요소 작제물은 환형 핵산 분자(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 플라스미드, 박미드, 또는 미니서클)이다. 유전 요소 작제물은 일부 구현예에서 이중-가닥일 수 있다. 다른 구현예에서, 유전 요소는 단일-가닥이다. 일부 구현예에서, 유전 요소 작제물은 DNA를 포함한다. 일부 구현예에서, 유전 요소 작제물은 RNA를 포함한다. 일부 구현예에서, 유전 요소 작제물은 하나 이상의 변형된 뉴클레오티드를 포함한다.

[0376] 일부 양태에서, 본 개시내용은 (예를 들어, 세포 배양 시스템에서) 본 명세서에 기재된 바와 같은 아넬로벡터의 복제 및 증식을 위한 방법을 제공하며, 이는 하기의 단계 중 하나 이상을 포함할 수 있다: (a) 유전 요소(예를 들어, 선형화됨)를 아넬로벡터 감염에 민감한 세포주 내로 도입하는(예를 들어, 형질감염시키는) 단계; (b) 세포를 수확하고, 선택적으로 유전 요소의 존재를 나타내는 세포를 단리하는 단계; (c) 단계 (b)에서 수득되는 세포를 실험 조건 및 유전자 발현에 따라, (예를 들어, 적어도 3일, 예컨대 적어도 1주 이상 동안) 배양하는 단계; 및 (d) 예를 들어, 본 명세서에 기재된 바와 같이, 단계 (c)의 세포를 수확하는 단계.

[0377] *비-아넬로바이러스 서열*

[0378] 본 명세서에 기재된 바와 같은 유전 요소 작제물은 비-아넬로바이러스의 게놈 유래의 핵산 서열(예를 들어, 길

이가 적어도 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, 또는 4000개 뉴클레오티드인 서열), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함할 수 있다. 비-아넬로바이러스 서열이 유래될 수 있는 바이러스의 예는 제한 없이, 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 아데노-연관 바이러스(AAV)를 포함한다. 일부 예에서, 유전 요소 작제물은 모노드나바이러스, 예를 들어, 쇼토쿠바이러스, 예를 들어, 코사비리코타, 예를 들어, 퀸토비리세테스, 예를 들어, 피코비랄레스, 예를 들어, 파르보비리대, 예를 들어, 파르보비리내, 예를 들어, 디펜도파르보바이러스, 예를 들어, AAV 유래의 서열을 포함한다. 일부 예에서, 유전 요소는 AAV(예를 들어, AAV1, AAV2, 또는 AAV5) 유래의 서열을 포함한다.

[0379] 일부 예에서, 유전 요소 작제물은, 예를 들어, 본 명세서에 기재된 바와 같은, 비-아넬로바이러스 복제 원점을 포함한다. 비-아넬로바이러스 복제 원점은, 일부 예에서, 비-아넬로바이러스 유래의 ITR, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열에 포함될 수 있다. 비-아넬로바이러스 복제 원점은, 일부 예에서, 비-아넬로바이러스의 Rep-결합 모티프(RBM), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열에 포함될 수 있다. 비-아넬로바이러스 복제 원점은, 일부 예에서, 비-아넬로바이러스의 말단 분해 부위(TRS), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열에 포함될 수 있다.

[0380] *플라스미드*

[0381] 일부 구현예에서, 유전 요소 작제물은 플라스미드이다. 플라스미드는 일반적으로 본 명세서에 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 내성 유전자)를 포함할 것이다. 일부 구현예에서, 유전 요소의 서열은 플라스미드로부터 절제될 수 있다. 일부 구현예에서, 플라스미드는 박테리아 세포에서 복제할 수 있다. 일부 구현예에서, 플라스미드는 포유동물 세포(예를 들어, 인간 세포)에서 복제할 수 있다. 일부 구현예에서, 플라스미드는 길이가 적어도 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 또는 5000 bp이다. 일부 구현예에서, 플라스미드는 길이가 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 또는 10,000 bp 미만이다. 일부 구현예에서, 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. 일부 구현예에서, 유전 요소는, 예를 들어, (예를 들어, 시험관 내 환화에 의해) 플라스미드로부터 절제되어, 예를 들어, 본 명세서에 기재된 바와 같은 미니서클을 형성할 수 있다. 구현예에서, 유전 요소의 절제는 유전 요소 서열을 플라스미드 백본으로부터 분리한다(예를 들어, 유전 요소를 박테리아 백본으로부터 분리함).

[0382] *작은 환형 핵산 작제물*

[0383] 일부 구현예에서, 유전 요소 작제물은, 예를 들어, 백본이 결여된(예를 들어, 박테리아 복제 원점 및/또는 선택 가능한 마커가 결여된) 환형 핵산 작제물이다. 구현예에서, 유전 요소는 이중-가닥 환형 핵산 작제물이다. 구현예에서, 이중-가닥 환형 핵산 작제물은, 예를 들어, 본 명세서에 기재된 바와 같이, 시험관 내 환화(IVC)에 의해 생성된다. 구현예에서, 이중-가닥 환형 핵산 작제물은 숙주 세포 내로 도입될 수 있으며, 숙주 세포 내에서, 예를 들어, 본 명세서에 기재된 바와 같은 단일-가닥 환형 유전 요소를 생성하기 위한 주형으로 전환되거나 이러한 주형으로서 사용될 수 있다. 일부 구현예에서, 환형 핵산 작제물은 플라스미드 백본 또는 이의 기능적 단편을 포함하지 않는다. 일부 구현예에서, 환형 핵산 작제물은 길이가 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 bp이다. 일부 구현예에서, 환형 핵산 작제물은 길이가 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 미만이다. 일부 구현예에서, 환형 핵산 작제물은 길이가 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 bp이다. 일부 구현예에서, 환형 핵산 작제물은 미니서클이다.

[0384] *시험관 내 환화*

[0385] 일부 예에서, 단백질성 외부에 패키징되는 유전 요소는 단일 가닥 환형 DNA이다. 유전 요소는 일부 예에서, 단일 가닥 환형 DNA 이외의 형태를 갖는 유전 요소 작제물을 통해 숙주 세포 내로 도입될 수 있다. 예를 들어, 유전 요소 작제물은 이중-가닥 환형 DNA일 수 있다. 그 다음 이중-가닥 환형 DNA는 숙주 세포(예를 들어, 문헌 [Wawrzyniak et al. 2017, *Front. Microbiol.* 8: 2353](열거된 효소와 관련하여 본 명세서에 참조로 포함됨)에 기재된 바와 같은, 예를 들어, 회전환 복제에 적합한 효소, 예를 들어, 아넬로바이러스 Rep 단백질, 예를 들어, Rep68/78, Rep60, RepA, RepB, Pre, MobM, TraX, TrwC, Mob02281, Mob02282, NikB, ORF50240, NikK, TecH, OrfJ, 또는 TraI를 포함하는 숙주 세포)에서 단일-가닥 환형 DNA로 전환될 수 있다. 일부 구현예에서, 이중-가닥 환형 DNA는, 예를 들어, 실시예 15에 기재된 바와 같이, 시험관 내 환화(IVC)에 의해 생성된다.

[0386] 일반적으로, 시험관 내 환화된 DNA 작제물은 유전 요소 서열이 선형 DNA 분자로서 절제되도록, 패키징될 유전 요소 작제물(예를 들어, 유전 요소의 서열을 포함하는 플라스미드)을 분해함으로써 생성될 수 있다. 그 다음, 생성된 선형 DNA는, 예를 들어, DNA 리가제를 사용하여 결찰되어, 이중-가닥 환형 DNA를 형성할 수 있다. 일부 예에서, 시험관 내 환화에 의해 생성되는 이중-가닥 환형 DNA는, 예를 들어, 본 명세서에 기재된 바와 같이, 회전환 복제를 겪을 수 있다. 이론에 구속되고자 하지 않지만, 시험관 내 환화가 추가의 변형 없이 회전환 복제를 겪을 수 있는 이중-가닥 DNA 작제물을 생성함으로써, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로벡터에 패키징되기에 적합한 크기의 단일-가닥 환형 DNA를 생성할 수 있는 것으로 고려된다. 일부 구현예에서, 이중-가닥 DNA 작제물은 플라스미드(예를 들어, 박테리아 플라스미드)보다 더 작다. 일부 구현예에서, 이중-가닥 DNA 작제물은 플라스미드(예를 들어, 박테리아 플라스미드)로부터 절제된 다음, 예를 들어, 시험관 내 환화에 의해 환화된다.

[0387] *탠덤 작제물*

[0388] 일부 구현예에서, 유전 요소 작제물은 탠덤으로 배열된, 유전 요소 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 유전 요소의 핵산 서열)의 제1 카피 및 유전 요소 서열(예를 들어, 동일한 유전 요소의 핵산 서열, 또는 상이한 유전 요소의 핵산 서열)의 제2 카피의 적어도 일부를 포함한다. 이러한 구조를 갖는 유전 요소 작제물은 일반적으로 본 명세서에서 탠덤 작제물로 지칭된다. 이러한 탠덤 작제물은 아넬로벡터 유전 요소를 생성하는 데 사용된다. 유전 요소 서열의 제1 카피 및 유전 요소 서열의 제2 카피는, 일부 예에서 유전 산 작제물에서 서로 바로 인접할 수 있다. 다른 예에서, 유전 요소 서열의 제1 카피 및 유전 요소 서열의 제2 카피는, 예를 들어, 스페이서 서열에 의해 분리될 수 있다. 일부 구현예에서, 유전 요소 서열의 제2 카피, 또는 이의 일부는, 예를 들어, 본 명세서에 기재된 바와 같은, 상류 복제-촉진 서열(uRFS)을 포함한다. 일부 구현예에서, 유전 요소 서열의 제2 카피, 또는 이의 일부는, 예를 들어, 본 명세서에 기재된 바와 같은, 하류 복제-촉진 서열(dRFS)을 포함한다. 일부 구현예에서, uRFS 및/또는 dRFS는 복제 원점(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 포유동물 복제 원점, 곤충 복제 원점, 또는 바이러스 복제 원점, 예를 들어, 비-아넬로바이러스 복제 원점) 또는 이의 일부를 포함한다. 일부 구현예에서, uRFS 및/또는 dRFS는 복제 원점을 포함하지 않는다. 일부 구현예에서, uRFS 및/또는 dRFS는 (예를 들어, 5' UTR에) 헤어핀 루프를 포함한다. 일부 구현예에서, 탠덤 작제물은 유전 요소의 제2 카피 또는 이의 일부가 결여된 다른 유사한 작제물보다 더 높은 수준의 유전 요소를 생성한다. 이론에 구속되고자 하지 않지만, 본 명세서에 기재된 탠덤 작제물은, 일부 구현예에서 회전환 복제에 의해 복제할 수 있다. 일부 구현예에서, 탠덤 작제물은 플라스미드이다. 일부 구현예에서, 탠덤 작제물은 환형이다. 일부 구현예에서, 탠덤 작제물은 선형이다. 일부 구현예에서, 탠덤 작제물은 단일-가닥이다. 일부 구현예에서, 탠덤 작제물은 이중-가닥이다. 일부 구현예에서, 탠덤 작제물은 DNA이다.

[0389] 탠덤 작제물은, 일부 예에서, 유전 요소 서열의 제1 카피 및 유전 요소 서열의 제2 카피, 또는 이의 일부를 포함할 수 있다. 제2 카피는 제1 카피 또는 이의 일부의 동일한 카피일 수 있거나, 하나 이상의 서열 차이, 예를 들어, 치환, 추가, 또는 결실을 포함할 수 있음이 이해된다. 일부 예에서, 유전 요소 서열의 제2 카피 또는 이의 일부는 유전 요소 서열의 제1 카피에 대해 5'에 위치한다. 일부 예에서, 유전 요소 서열의 제2 카피 또는 이의 일부는 유전 요소 서열의 제1 카피에 대해 3'에 위치한다. 일부 예에서, 유전 요소 서열의 제2 카피 또는 이의 일부 및 유전 요소 서열의 제1 카피는 탠덤 작제물에서 서로 인접해 있다. 일부 예에서, 유전 요소 서열의 제2 카피 또는 이의 일부 및 유전 요소 서열의 제1 카피는, 예를 들어, 스페이서 서열에 의해 분리된다.

[0390] 일부 구현예에서, 본 명세서에 기재된 탠덤 작제물은 치료용 이펙터를 인코딩하는 (예를 들어, ORF1 분자가 유

래된 아넬로바이러스에 대한) 이종 서열 및 캡시드에 결합하는 단백질 결합 서열을 포함하는 유전 요소를 캡슐화하는 캡시드(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF, 예를 들어, ORF1 분자를 포함하는 캡시드)를 포함하는 벡터(예를 들어, 아넬로벡터), 비히클, 또는 입자(예를 들어, 바이러스 입자)의 유전 요소를 생성하는 데 사용될 수 있다. 구현예에서, 벡터는 유전 요소를 포유동물 세포, 예를 들어, 인간 세포 내로 운반할 수 있다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스 게놈 서열에 대하여 약 50% 미만(예를 들어, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5% 미만 또는 그 이하)의 동일성을 갖는다. 일부 구현예에서, 유전 요소는 야생형 아넬로바이러스 게놈 서열에 대하여 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 5.5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 또는 80% 이하의 동일성을 갖는다. 일부 구현예에서, 유전 요소는 약 2000, 3000, 4000, 4500, 또는 5000개 초과의, 비-아넬로바이러스 게놈 서열의 인접 뉴클레오티드를 갖는다. 일부 구현예에서, 유전 요소는 약 2000 내지 5000, 2500 내지 4500, 3000 내지 4500, 2500 내지 4500, 3500 또는 4000, 4500개(예를 들어, 약 3000 내지 4500개) 초과의, 비-아넬로바이러스 게놈 서열의 뉴클레오티드 뉴클레오티드를 갖는다.

[0391] 본 명세서의 시스템 및 방법의 일부 구현예에서, 벡터(예를 들어, 아넬로벡터)는, 예를 들어, 본 명세서에 기재된 바와 같은, 유전 요소 또는 유전 요소 작제물, 예를 들어, 탠덤 작제물인 제1 핵산 분자, 및 하나 이상의 추가 단백질(예를 들어, Rep 분자 및/또는 캡시드 단백질)을 인코딩하는 제2 핵산 분자를 세포 내로 도입함으로써 제조된다. 일부 구현예에서, 제1 핵산 분자 및 제2 핵산 분자는 (예를 들어, 본 명세서에 기재된 유전 요소 작제물에서, 예를 들어, 시스로) 서로 부착된다. 일부 구현예에서, 제1 핵산 분자 및 제2 핵산 분자는 (예를 들어, 트랜스로) 분리되어 있다. 일부 구현예에서, 제1 핵산 분자는 플라스미드, 코스미드, 박미드, 미니서클, 또는 인공 염색체이다. 일부 구현예에서, 제2 핵산 분자는 플라스미드, 코스미드, 박미드, 미니서클, 또는 인공 염색체이다. 일부 구현예에서, 제2 핵산 분자는 숙주 세포의 게놈에 통합된다.

[0392] 일부 구현예에서, 방법은 제1 핵산 분자 및/또는 제2 핵산 분자를 숙주 세포 내로 도입하는 단계를 추가로 포함한다. 일부 구현예에서, 제2 핵산 분자는 제1 핵산 분자 이전에, 이와 동시에, 또는 이후에 숙주 세포 내로 도입된다. 다른 구현예에서, 제2 핵산 분자는 숙주 세포의 게놈에 통합된다. 일부 구현예에서, 제2 핵산 분자는, 예를 들어, 본 명세서에 기재된 바와 같은, 헬퍼 작제물, 헬퍼 바이러스 또는 다른 헬퍼 벡터이거나 이를 포함하거나 이의 일부이다.

[0393] *시스/트랜스 작제물*

[0394] 일부 구현예에서, 본 명세서에 기재된 바와 같은 유전 요소 작제물은 하나 이상의 아넬로바이러스 ORF, 예를 들어, 단백질성 외부 구성성분(예를 들어, 본 명세서에 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 도입하는 데 적합할 수 있다. 다른 구현예에서, 본 명세서에 기재된 바와 같은 유전 요소 작제물은 하나 이상의 아넬로바이러스 ORF, 예를 들어, 단백질성 외부 구성성분(예를 들어, 본 명세서에 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, 하나 이상의 아넬로바이러스 ORF를 인코딩하는 제2 핵산 작제물의 도입을 통해, 또는 숙주 세포의 게놈에 통합된 아넬로바이러스 ORF 카세트를 통해) 트랜스로 제공되는 하나 이상의 아넬로바이러스 ORF와 함께, 유전 요소를 숙주 세포 내로 도입하는 데 적합할 수 있다. 일부 구현예에서, ORF1 분자는, 예를 들어, 본 명세서에 기재된 바와 같이, 트랜스로 제공된다. 일부 구현예에서, ORF2 분자는, 예를 들어, 본 명세서에 기재된 바와 같이, 트랜스로 제공된다. 일부 구현예에서, ORF1 분자 및 ORF1 분자는 둘 다, 예를 들어, 본 명세서에 기재된 바와 같이, 트랜스로 제공된다.

[0395] 일부 구현예에서, 유전 요소 작제물은 아넬로바이러스 ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 구현예에서, 유전 요소의 서열을 포함하지 않는 유전 요소의 부분은 (예를 들어, 프로모터 및 아넬로바이러스 ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편을 인코딩하는 서열을 포함하는 카세트에) 아넬로바이러스 ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편을 인코딩하는 서열을 포함한다. 추가 구현예에서, 유전 요소 서열을 포함하는 작제물의 일부는 아넬로바이러스 ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함한다. 구현예에서, 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같음)에 이러한 유전 요소의 봉입은

복제-구성성분 아넬로벡터(예를 들어, 세포 감염 시, 예를 들어, 본 명세서에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF를 인코딩하는, 추가 핵산 작제물을 세포에 도입하지 않고 세포가 아넬로벡터의 추가 카피를 생성할 수 있게 하는 아넬로벡터)를 생성한다.

[0396] 다른 구현예에서, 유전 요소는 아넬로바이러스 ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 젤리-롤 영역)을 인코딩하는 서열을 포함하지 않는다. 구현예에서, 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같음)에 이러한 유전 요소의 봉입은 복제-부적격 아넬로벡터(예를 들어, 세포 감염 시, 예를 들어, 본 명세서에 기재된 바와 같은 하나 이상의 아넬로바이러스 ORF를 인코딩하는, 하나 이상의 추가 작제물의 부재 시, 감염된 세포가 추가 아넬로벡터를 생성할 수 없게 하는 아넬로벡터)를 생성한다.

[0397] *발현 카세트*

[0398] 일부 구현예에서, 유전 요소 작제물은 폴리펩티드 또는 비-코딩 RNA(예를 들어, miRNA 또는 siRNA)의 발현을 위한 하나 이상의 카세트를 포함한다. 일부 구현예에서, 유전 요소 작제물은 이펙터(예를 들어, 외인성 또는 내인성 이펙터), 예를 들어, 본 명세서에 기재된 바와 같은 폴리펩티드 또는 비-코딩 RNA의 발현을 위한 카세트를 포함한다. 일부 구현예에서, 유전 요소 작제물은 아넬로바이러스 단백질(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2, 또는 이의 기능적 단편)의 발현을 위한 카세트를 포함한다. 발현 카세트는 일부 구현예에서 유전 요소 서열 내에 위치할 수 있다. 구현예에서, 이펙터에 대한 발현 카세트는 유전 요소 서열 내에 위치한다. 구현예에서, 아넬로바이러스 단백질에 대한 발현 카세트는 유전 요소 서열 내에 위치한다. 다른 구현예에서, 발현 카세트는 유전 요소의 서열 외부에 있는 유전 요소 작제물 내의 위치(예를 들어, 백본 내)에 위치한다. 구현예에서, 아넬로바이러스 단백질에 대한 발현 카세트는 유전 요소의 서열 외부에 있는 유전 요소 작제물 내의 위치(예를 들어, 백본)에 위치한다.

[0399] 폴리펩티드 발현 카세트는 일반적으로 프로모터 및 폴리펩티드, 예를 들어, 이펙터(예를 들어, 본 명세서에 기재된 바와 같은 외인성 또는 내인성 이펙터) 또는 아넬로바이러스 단백질을 인코딩하는 코딩 서열(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2, 또는 이의 기능적 단편을 인코딩하는 서열)을 포함한다. (예를 들어, 폴리펩티드의 발현을 유도하기 위해) 폴리펩티드 발현 카세트에 포함될 수 있는 예시적인 프로모터는 제한 없이, 예를 들어, 본 명세서에 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 a 쇄 프로모터, 뉴런-특이적 에놀라제(NSE) 프로모터, 또는 신경필라멘트 경쇄 프로모터), 및 유도성 프로모터(예를 들어, 아연-유도성 양 메탈로티오닌(MT) 프로모터; 덱사메타손(Dex)-유도성 마우스 유방 종양 바이러스(MMTV) 프로모터; T7 중합효소 프로모터 시스템, 테트라사이클린-억제성 시스템, 테트라사이클린-유도성 시스템, RU486-유도성 시스템, 라파마이신-유도성 시스템)를 포함한다. 일부 구현예에서, 발현 카세트는, 예를 들어, 본 명세서에 기재된 바와 같은, 인핸서를 추가로 포함한다.

[0400] *유전 요소 작제물의 설계 및 생성*

[0401] 유전 요소 작제물을 합성하기 위해 다양한 방법이 이용 가능하다. 예를 들어, 유전 요소 작제물 서열은 합성하기에 더 쉬운 더 작은 중첩하는 조각(예를 들어, 약 100 bp 내지 약 10 kb 범위의 세그먼트 또는 개별 ORF)으로 나뉠 수 있다. 이들 DNA 세그먼트는 중첩하는 단일-가닥 올리고뉴클레오티드의 세트로부터 합성된다. 그 다음 생성되는 중첩 신톤(synthon)은 더 큰 조각의 DNA, 예를 들어, 유전 요소 작제물로 조립된다. 세그먼트 또는 ORF는 유전 요소 작제물 내로, 예를 들어, 5' 및 3' 말단에서 시험관 내 재조합 또는 특유한 제한 부위에 의해 조립되어, 결찰을 가능하게 할 수 있다.

[0402] 유전 요소 작제물은 작제물 서열을 올리고-길이 단편으로 분석하는 설계 알고리즘으로 합성하여, 서열 공간의 복잡성을 고려하는 합성을 위한 적합한 설계 조건을 생성할 수 있다. 그 다음 올리고는 반도체-기반의 고밀도 칩 상에서 화학적으로 합성되며, 칩마다 200,000개 초과의 개별 올리고가 합성된다. 올리고를 조립 기법, 예컨대 BioFab®으로 조립하여, 더 작은 올리고로부터 더 긴 DNA 세그먼트를 구축한다. 이는 병렬 방식으로 행해지며, 따라서 수백 내지 수천개의 합성 DNA 세그먼트가 한번에 구축된다.

[0403] 각각의 유전 요소 작제물 또는 유전 요소 작제물의 세그먼트는 검증된 서열일 수 있다. 일부 구현예에서, RNA 또는 DNA의 고-처리량 서열결정은 AnyDot.chips(Genovoxx, 독일)를 사용하여 일어날 수 있으며, 이는 생물학적

과정(예를 들어, miRNA 발현 또는 대립형질 가변성(SNP 검출))의 모니터링을 가능하게 한다. 다른 고-처리량 서열결정 시스템은 문헌[Venter, J., et al. Science 16 Feb. 2001]; 문헌[Adams, M. et al, Science 24 Mar. 2000]; 및 문헌[M. J, Levene, et al. Science 299:682-686, January 2003]; 및 미국 공개 출원 20030044781 및 2006/0078937에 개시된 것을 포함한다. 전체적으로 이러한 시스템은 핵산 분자에서 측정되는 중합 반응을 통해 염기를 일시적으로 첨가함으로써 복수의 염기를 갖는 표적 핵산 분자를 서열결정하는 것을 수반하며, 즉 서열결정될 주형 핵산 분자 상의 핵산 중합 효소의 활성이 실시간으로 추적된다. 일부 구현예에서, 샷건(shotgun) 서열결정이 수행된다.

[0404] 유전 요소 작제물은 복제 또는 패키징을 위한 인자가 유전 요소에 비하여 시스 또는 트랜스로 공급될 수 있도록 설계될 수 있다. 예를 들어, 시스로 공급되는 경우, 유전 요소는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 또는 ORF2t/3을 인코딩하는 하나 이상의 유전자를 포함할 수 있다. 일부 구현예에서, 복제 및/또는 패키징 신호는, 예를 들어, 증폭 및/또는 캡슐화를 유도하기 위하여 유전 요소 내로 혼입될 수 있다. 일부 구현예에서, 이펙터는 게놈 내의 특정 부위 내로 삽입된다. 일부 구현예에서, 하나 이상의 바이러스 ORF는 이펙터로 대체된다.

[0405] 또 다른 예에서, 복제 또는 패키징 인자가 트랜스로 공급되는 경우, 유전 요소는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 또는 ORF2t/3 중 하나 이상을 인코딩하는 유전자가 결여될 수 있으며; 이 단백질 또는 단백질들은, 예를 들어, 또 다른 핵산, 예를 들어, 헬퍼 핵산에 의해 공급될 수 있다. 일부 구현예에서, 최소 시스 신호(예를 들어, 5' UTR 및/또는 GC-풍부 영역)가 유전 요소 내에 존재한다. 일부 구현예에서, 유전 요소는 복제 또는 패키징 인자(예를 들어, 복제효소 및/또는 캡시드 단백질)를 인코딩하지 않는다. 이러한 인자는 일부 구현예에서, 하나 이상의 헬퍼 핵산(예를 들어, 숙주 세포 게놈 내로 통합되는 헬퍼 바이러스 핵산, 헬퍼 플라스미드 또는 헬퍼 핵산)에 의해 공급될 수 있다. 일부 구현예에서, 헬퍼 핵산은 증폭 및/또는 패키징을 유도하기에 충분한 단백질 및/또는 RNA를 발현하지만, 이들 자체의 패키징 신호가 결여될 수 있다. 일부 구현예에서, 유전 요소 및 헬퍼 핵산은 (예를 들어, 동시에 또는 개별적으로) 숙주 세포 내로 도입되어, 유전 요소의 증폭 및/또는 패키징을 초래하지만, 헬퍼 핵산의 증폭 및/또는 패키징을 초래하지 않는다.

[0406] 일부 구현예에서, 유전 요소 작제물은 컴퓨터-지원 설계 도구를 사용하여 설계될 수 있다.

[0407] 작제물을 제조하는 일반적인 방법은, 예를 들어, 문헌[Khudyakov & Fields, *Artificial DNA: Methods and Applications*, CRC Press (2002)]; 문헌[Zhao, *Synthetic Biology: Tools and Applications*, (First Edition), Academic Press (2013)]; 및 문헌[Egli & Herdewijn, *Chemistry and Biology of Artificial Nucleic Acids*, (First Edition), Wiley-VCH (2012)]에 기재되어 있다.

[0408] ***이펙터***

[0409] 본 명세서에 기재된 조성물 및 방법은, 예를 들어, 본 명세서에 기재된 바와 같은, 이펙터(예를 들어, 외인성 이펙터 또는 내인성 이펙터)를 인코딩하는 서열을 포함하는 아넬로벡터의 유전 요소를 생성하는 데 사용될 수 있다. 이펙터는 일부 예에서, 내인성 이펙터 또는 외인성 이펙터일 수 있다. 일부 구현예에서, 이펙터는 치료용 이펙터이다. 일부 구현예에서, 이펙터는 폴리펩티드(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 치료용 폴리펩티드 또는 펩티드)를 포함한다. 일부 구현예에서, 이펙터는 비-코딩 RNA(예를 들어, miRNA, siRNA, shRNA, mRNA, lncRNA, RNA, DNA, 안티센스 RNA, 또는 gRNA)를 포함한다. 일부 구현예에서, 이펙터는, 예를 들어, 본 명세서에 기재된 바와 같은, 조절 핵산을 포함한다.

[0410] 일부 구현예에서, 이펙터-인코딩 서열은, TATA 박스 상류의 5' 비-코딩 영역, 5' UTR, 폴리-A 신호의 하류, 또는 GC-풍부 영역 상류의 3' 비-코딩 영역에서, 예를 들어, 비-코딩 영역, 예를 들어, 오픈-리딩 프레임의 3' 및 유전 요소의 GC-풍부 영역의 5'에 배치된 비-코딩 영역에서, 유전 요소에 삽입될 수 있다. 일부 구현예에서, 이펙터-인코딩 서열은, 예를 들어, 코딩 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 및/또는 ORF2t/3을 인코딩하는 서열)에서, 유전 요소에 삽입될 수 있다. 일부 구현예에서, 이펙터-인코딩 서열은 오픈 리딩 프레임의 전부 또는 일부를 대체한다. 일부 구현예에서, 유전 요소는 이펙터-인코딩 서열에 작동 가능하게 연결된 조절 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 프로모터 또는 인핸서)을 포함한다.

[0411] ***숙주 세포***

[0412] 본 명세서에 기재된 아넬로벡터는, 예를 들어, 숙주 세포에서 생성될 수 있다. 일반적으로, 아넬로벡터 유전 요

소 및 아넬로벡터 단백질성 외부(예를 들어, 아넬로바이러스 ORF1 핵산 또는 아넬로바이러스 ORF1 분자에 의해 인코딩된 폴리펩티드)의 구성성분을 포함하는 숙주 세포가 제공된다. 그 다음 숙주 세포는 단백질성 외부 내에 유전 요소를 봉입하기에 적합한 조건(예를 들어, 본 명세서에 기재된 바와 같은 배양 조건) 하에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 아넬로벡터를 숙주 세포로부터, 예를 들어, 주변 상청액으로 방출하기에 적합한 조건 하에서 추가로 인큐베이션된다. 일부 구현예에서, 숙주 세포는 세포 용해물로부터 아넬로벡터의 수확을 위해 용해된다. 일부 구현예에서, 아넬로벡터는 높은 세포 밀도로 성장한 숙주 세포주에 도입될 수 있다. 일부 구현예에서, 숙주 세포는 Expi-293 세포이다.

[0413] *숙주 세포 내로의 유전 요소의 도입*

[0414] 유전 요소, 또는 유전 요소의 서열을 포함하는 핵산 작제물은 숙주 세포 내로 도입될 수 있다. 일부 구현예에서, 유전 요소 자체가 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소(예를 들어, 본 명세서에 기재된 바와 같음)의 서열을 포함하는 유전 요소 작제물이 숙주 세포 내로 도입된다. 유전 요소 또는 유전 요소 작제물은, 예를 들어, 당업계에 알려진 방법을 사용하여 숙주 세포 내로 도입될 수 있다. 예를 들어, 유전 요소 또는 유전 요소 작제물은 형질감염(예를 들어, 안정한 형질감염 또는 일시적 형질감염)에 의해 숙주 세포 내로 도입될 수 있다. 구현예에서, 유전 요소 또는 유전 요소 작제물은 lipofectamine 형질감염에 의해 숙주 세포 내로 도입된다. 구현예에서, 유전 요소 또는 유전 요소 작제물은 칼슘 포스페이트 형질감염에 의해 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소 또는 유전 요소 작제물은 전기천공에 의해 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소 또는 유전 요소 작제물은 유전자 총을 사용하여 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소 또는 유전 요소 작제물은 뉴클레오펙션에 의해 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소 또는 유전 요소 작제물은 PEI 형질감염에 의해 숙주 세포 내로 도입된다. 일부 구현예에서, 유전 요소는 숙주 세포를 유전 요소를 포함하는 아넬로벡터와 접촉시킴으로써 숙주 세포 내로 도입된다.

[0415] 구현예에서, 유전 요소 작제물은 일단 숙주 세포 내로 도입되면 복제할 수 있다. 구현예에서, 유전 요소는 일단 숙주 세포 내로 도입되면 유전 요소 작제물로부터 생성될 수 있다. 일부 구현예에서, 유전 요소는, 예를 들어, 주형으로서 유전 요소 작제물을 사용하여, 중합효소에 의해 숙주 세포에서 생성된다.

[0416] 일부 구현예에서, 유전 요소 또는 유전 요소를 포함하는 벡터는 아넬로벡터의 발현을 달성하기 위해 바이러스 중합효소 단백질을 발현하는 세포주 내로 도입(예를 들어, 형질감염)된다. 이를 위해, 아넬로벡터 중합효소 단백질을 발현하는 세포주가 적절한 숙주 세포로 이용될 수 있다. 숙주 세포는 다른 바이러스 기능 또는 추가 기능을 제공하도록 유사하게 조작될 수 있다.

[0417] 본 명세서에 개시된 아넬로벡터를 제조하기 위해, 유전 요소 작제물이 복제 및 생성에 필요한 아넬로벡터 단백질 및 기능을 제공하는 세포를 형질감염시키는 데 사용될 수 있다. 대안적으로, 세포는 본 명세서에 개시된 유전 요소 또는 유전 요소를 포함하는 벡터에 의한 형질감염 전, 도중 또는 후에 아넬로벡터 단백질 및 기능을 제공하는 제2 작제물(예를 들어, 바이러스)로 형질감염될 수 있다. 일부 구현예에서, 제2 작제물은 불완전한 바이러스 입자의 생성을 보완하는 데 유용할 수 있다. 제2 작제물(예를 들어, 바이러스)은 숙주 범위 제한 또는 온도 민감성과 같은 조건부 성장 결함을 가질 수 있으며, 이는 예를 들어, 형질감염 바이러스의 후속 선택을 가능하게 한다. 일부 구현예에서, 제2 작제물은 아넬로벡터의 발현을 달성하기 위해 숙주 세포에 의해 이용되는 하나 이상의 복제 단백질을 제공할 수 있다. 일부 구현예에서, 숙주 세포는 하나 이상의 복제 단백질과 같은 바이러스 단백질을 인코딩하는 벡터로 형질감염될 수 있다. 일부 구현예에서, 제2 작제물은 항바이러스 민감성을 포함한다.

[0418] 본 명세서에 개시된 유전 요소 또는 유전 요소를 포함하는 벡터는 일부 예에서 당업계에 알려진 기법을 사용하여 아넬로벡터로 복제 및 생성될 수 있다. 예를 들어, 다양한 바이러스 배양 방법이, 예를 들어, 미국 특허 4,650,764; 미국 특허 5,166,057; 미국 특허 5,854,037; 유럽 특허 공개 EP 0702085A1; 미국 특허 출원 일련번호 09/152,845; 국제 특허 공개 PCT WO97/12032; WO96/34625; 유럽 특허 공개 EP-A780475; WO 99/02657; WO 98/53078; WO 98/02530; WO 99/15672; WO 98/13501; WO 97/06270; 및 EPO 780 47SA1에 기재되어 있으며, 이들 각각은 본 명세서에 전문이 참조로 포함된다.

[0419] *시스 또는 트랜스로 단백질(들)을 제공하는 방법*

[0420] 일부 구현예(예를 들어, 본 명세서에 기재된 시스 구현예)에서, 유전 요소 작제물은 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2, 또는 이의 기능적 단편)에 대한 코딩 서열을 포함하는 하나 이상의 발현 카세트를 추가로 포함한다. 구현예에서, 유전 요소 작제물은 아넬로바이

러스 ORF1, 또는 이의 스플라이스 변이체 또는 기능적 단편에 대한 코딩 서열을 포함하는 발현 카세트를 포함한다. 이펙터뿐만 아니라 하나 이상의 아넬로바이러스 ORF에 대한 발현 카세트를 포함하는 이러한 유전 요소 작제물은, 숙주 세포 내로 도입될 수 있다. 이러한 유전 요소 작제물을 포함하는 숙주 세포는, 일부 예에서, 추가의 핵산 작제물 또는 발현 카세트의 숙주 세포 게놈 내로의 통합을 요구하지 않으면서, 단백질성 외부에 대한, 그리고 단백질성 외부 내의 유전 요소의 봉입을 위한 유전 요소 및 구성성분을 생성할 수 있다. 다시 말하면, 이러한 유전 요소 작제물은, 예를 들어, 본 명세서에 기재된 바와 같이, 숙주 세포에서 시스 아넬로벡터 생성 방법에 사용될 수 있다.

[0421] 일부 구현예(예를 들어, 본 명세서에 기재된 트랜스 구현예)에서, 유전 요소는 하나 이상의 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 또는 ORF1/2, 또는 이의 기능적 단편)에 코딩 서열을 포함하는 발현 카세트를 포함하지 않는다. 구현예에서, 유전 요소 작제물은 아넬로바이러스 ORF1, 또는 이의 스플라이스 변이체 또는 기능적 단편에 대한 코딩 서열을 포함하는 발현 카세트를 포함하지 않는다. 이펙터에 대한 발현 카세트를 포함하지만 하나 이상의 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1 또는 이의 스플라이스 변이체 또는 기능적 단편)에 대한 발현 카세트가 결여된 이러한 유전 요소 작제물은 숙주 세포 내로 도입될 수 있다. 이러한 유전 요소 작제물을 포함하는 숙주 세포는, 일부 예에서, 아넬로벡터의 하나 이상의 구성성분(예를 들어, 단백질성 외부 단백질)의 생성을 위해 추가의 핵산 작제물 또는 발현 카세트의 숙주 세포 게놈으로의 통합을 필요로 할 수 있다. 일부 구현예에서, 이러한 유전 요소 작제물을 포함하는 숙주 세포는 아넬로바이러스 ORF1 분자를 인코딩하는 추가의 핵 작제물의 부재시 단백질성 외부 내에 유전 요소를 봉입할 수 없다. 다시 말하면, 이러한 유전 요소 작제물은, 예를 들어, 본 명세서에 기재된 바와 같이, 숙주 세포에서 트랜스 아넬로벡터 생성 방법에 사용될 수 있다.

[0422] 일부 구현예(예를 들어, 본 명세서에 기재된 시스 구현예)에서, 유전 요소 작제물은 하나 이상의 비-아넬로바이러스 ORF(예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질, 예를 들어, AAV Rep2 단백질)에 대한 코딩 서열을 포함하는 하나 이상의 발현 카세트를 추가로 포함한다. 이펙터뿐만 아니라 하나 이상의 비-아넬로바이러스 ORF에 대한 발현 카세트를 포함하는 이러한 유전 요소 작제물은, 숙주 세포 내로 도입될 수 있다. 이러한 유전 요소 작제물을 포함하는 숙주 세포는, 일부 예에서, 추가의 핵산 작제물 또는 발현 카세트의 숙주 세포 게놈 내로의 통합을 요구하지 않으면서, 단백질성 외부에 대한, 그리고 단백질성 외부 내의 유전 요소의 봉입을 위한 유전 요소 및 구성성분을 생성할 수 있다. 다시 말하면, 이러한 유전 요소 작제물은, 예를 들어, 본 명세서에 기재된 바와 같이, 숙주 세포에서 시스 아넬로벡터 생성 방법에 사용될 수 있다.

[0423] 일부 구현예(예를 들어, 본 명세서에 기재된 트랜스 구현예)에서, 유전 요소는 하나 이상의 비-아넬로바이러스 ORF(예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질, 예를 들어, AAV Rep2 단백질)에 대한 코딩 서열을 포함하는 발현 카세트를 포함하지 않는다. 이펙터에 대한 발현 카세트를 포함하지만 하나 이상의 비-아넬로바이러스 ORF(예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질, 예를 들어, AAV Rep2 단백질)에 대한 발현 카세트가 결여된 이러한 유전 요소 작제물은 숙주 세포 내로 도입될 수 있다. 이러한 유전 요소 작제물을 포함하는 숙주 세포는, 일부 예에서, (예를 들어, 유전 요소의 복제를 위한) 아넬로벡터의 하나 이상의 구성성분의 생성을 위해 추가의 핵산 작제물 또는 발현 카세트의 숙주 세포 게놈으로의 통합을 필요로 할 수 있다. 일부 구현예에서, 이러한 유전 요소 작제물을 포함하는 숙주 세포는, 예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질, 예를 들어, AAV Rep2 단백질을 인코딩하는, 추가 핵 작제물의 부재시 유전 요소를 복제할 수 없다. 다시 말하면, 이러한 유전 요소 작제물은, 예를 들어, 본 명세서에 기재된 바와 같이, 숙주 세포에서 트랜스 아넬로벡터 생성 방법에 사용될 수 있다.

[0424] *헬퍼 및 비-아넬로바이러스 분자*

[0425] 일부 구현예에서, 비-아넬로바이러스 바이러스 유래의 분자(예를 들어, 핵산 분자 또는 폴리펩티드), 또는 이에 기반한 분자는 숙주 세포에 존재한다. 비-아넬로바이러스 바이러스 유래의 분자, 또는 이에 기반한 분자는, 일부 구현예에서 본 명세서에 기재된 바와 같은 아넬로벡터의 생성에 기여할 수 있다. 예를 들어, 비-아넬로바이러스 바이러스 유래의 분자, 또는 이에 기반한 분자는 동족 복제 원점(예를 들어, AAV 복제 원점)을 포함하는 아넬로벡터 유전 요소의 복제를 촉진하는 비-아넬로바이러스 Rep 분자(예를 들어, AAV Rep 분자)를 포함할 수 있다.

[0426] 일부 구현예에서, AAV Rep 단백질은 하기 표 60에 열거된 바와 같은 아미노산 서열, 또는 이에 대하여 적어도

50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, AAV Rep 단백질은 서열번호 1030 내지 1042 중 임의의 것의 아미노산 서열, 또는 이에 대하여 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다.

[0427] **[표 60] 예시적인 AAV Rep 단백질 서열**

명칭	서열	서열번호
AAV2 Rep 서열		
Rep 코딩 영역	ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGC GACCTTGACGAGCATCTGCCCGGCATTTCTGACAGCTTT GTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCA GATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCC CTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGACG GAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTC TTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATG CACGTGCTCGTGGAAACCACCGGGGTGAAATCCATGGTT TTGGGACGTTTCCTGAGTCAGATTCGCGAAAAACTGATT CAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAAAC TGGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGC GGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTAC TTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGG ACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTC ACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCAC GTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAAT CCCAATTCTGATGCGCCGGTGATCAGATCAAAAACTTCA GCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAG GGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAG GCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGG TCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATT ATGAGCCTGACTAAAACCGCCCCCGACTACCTGGTGGGC CAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTAT AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCG GCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGC AAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACC GGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTG CCCTTCTACGGGTGCGTAAACTGGACCAATGAGAACTTT CCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGG GAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCC AAAGCCATTCTCGGAGGAAGCAAGGTGCGCGTGGACCAG AAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTG ATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGAC GGGAACTCAACGACCTTCGAACACCAGCAGCCGTTGCAA GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGAT CATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGAC TTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAG CATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGA CCCGCCCCCAGTGACGCAGATATAAGTGAGCCCAAACGG GTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCG GAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAA TGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCC TGCAGACAATGCGAGAGAATGAATCAGAATTCAAATATC TGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTT CCCGTGTCAGAATCTCAACCCGTTTCTGTCGTCAAAAAG GCGTATCAGAAACTGTGCTACATTCATCATATCATGGGA AAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAAT GTGGATTTGGATGACTGCATCTTTGAACAATAAATGATT TAAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTG GCTCGAGGACACTCTCTCTGA	1030
Rep78 AA	MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPP DSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALF FVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLI QRIYRGIEPTLPNWFAVTKTRNGAGGGNKVVDECYIPNY LLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDK GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKI MSLTKTAPDYLVGQQPVEDISSNRIYKILELNGYDPQYA ASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKD FFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISEPKR VRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMLFP CRQCERMNQNSNICFTHGQKDCLECFPVSESQPVSVVKK AYQKLCYIHHIMGKVPDACTACDLVNVDLDDCIFEQ	1031

[0428]

[0429]

Rep68 AA	MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPP DSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALF FVQFEKGESYFHMHVLVETTGVKSMVLGRFLSQIREKLI QRIYRGIEPTLPNWFAVTKTRNGAGGGNKVVDECYIPNY LLPKTQPELQWAWTNMEQYLSACLNLTERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDK GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNAGKI MSLTKTAPDYLVGQQPVEDISSNRIYKILELNGYDPQYA ASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIDPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFKFELTRRLDHDFGKVTKQEVKD FFRWAKDHVVEVEHEFYVKKGGAKKRPAPSDADISEPKR VRESVAQPSTSDAEASINYADRLARGHSL	1032
Rep52 AA	MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNAGKIMSLTKTAPDYLVGQQPVEDISSNRIYKIL ELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDF GKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAP SDADISEPKRVRESVAQPSTSDAEASINYADRYQNKCSR HVGMNLMLFPCRQCERMNQNSNICFTHGQKDCLECFPVS ESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLVNVDL DDCIFEQ	1033
Rep40 AA	MELVGWLVDKGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNAGKIMSLTKTAPDYLVGQQPVEDISSNRIYKIL ELNGYDPQYAASVFLGWATKKFGKRNTIWLFGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIDPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDF GKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPAP SDADISEPKRVRESVAQPSTSDAEASINYADRLARGHSL	1034

[0430]

AAV3 Rep 서열		
Rep 코딩 영역	ATGCCGGGGTTCTACGAGATTGTCCTGAAGGTCCCGAGT GACCTGGACGAGCGCCTGCCGGGCATTTCTAACTCGTTT GTTAACTGGGTGGCCGAGAAGGAATGGGACGTGCCGCCG GATTCTGACATGGATCCGAATCTGATTGAGCAGGCACCC CTGACCGTGGCCGAAAAGCTTCAGCGCGAGTTCCTGGTG GAGTGGCGCCGCGTGAGTAAGGCCCCGGAGGCCCTCTTT TTTGTCCAGTTCGAAAAGGGGGAGACCTACTTCCACCTG CACGTGCTGATTGAGACCATCGGGGTCAAATCCATGGTG GTCGGCCGCTACGTGAGCCAGATTAAAGAGAAGCTGGTG ACCCGCATCTACCGCGGGGTCGAGCCGCAGCTTCCGAAC TGGTTCGCGGTGACCAAAACGCGAAATGGCGCCGGGGGC GGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTAC CTGCTCCCCAAGACCCAGCCCGAGCTCCAGTGGGCGTGG ACTAACATGGACCAGTATTTAAGCGCCTGTTTGAATCTC GCGGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGCAC GTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAAC CCCAATTCTGACGCGCCGGTCATCAGGTCAAAAACCTCA GCCAGGTACATGGAGCTGGTCGGGTGGCTGGTGGACCGC GGGATCACGTCAGAAAAGCAATGGATTCAGGAGGACCAG GCCTCGTACATCTCCTTCAACGCCGCCTCCAACTCGCGG TCCCAGATCAAGGCCGCGCTGGACAATGCCTCCAAGATC ATGAGCCTGACAAAGACGGCTCCGGACTACCTGGTGGG- CAGCAACCCGCCGGAGGACATTACCAAAAATCGGATCTA CCAAATCCTGGAGCTGAACGGGTACGATCCGCAGTACGC GGCCTCCGTCTTCCTGGGCTGGGCGCAAAAGAAGTTCGG GAAGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGAC GGGTAAAACCAACATCGCGGAAGCCATCGCCCACGCCGT GCCCTTCTACGGCTGCGTAAACTGGACCAATGAGAACTT TCCCTTCAACGATTGCGTCGACAAGATGGTGATCTGGTG GGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGAGCGC CAAGGCCATTCTGGGCGGAAGCAAGGTGCGCGTGGACCA AAAGTGCAAGTCATCGGCCCAGATCGAACCCACTCCCGT GATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGA CGGGAACAGCACCACCTTCGAGCATCAGCAGCCGCTGCA GGACCGGATGTTTGAATTTGAACTTACCCGCCGTTTGGA CCATGACTTTGGGAAGGTCACCAAACAGGAAGTAAAGGA CTTTTTCCGGTGGGCTTCCGATCACGTGACTGACGTGGC TCATGAGTTCTACGTCAGAAAGGGTGGAGCTAAGAAACG CCCCGCCTCCAATGACGCGGATGTAAGCGAGCCAAAACG GGAGTGCACGTCACTTGCGCAGCCGACAACGTCAGACGC GGAAGCACCGGCGGACTACGCGGACAGGTACCAAAACAA ATGTTCTCGTCACGTGGGCATGAATCTGATGCTTTTTCC CTGTAAAACATGCGAGAGAATGAATCAAATTTCCAATGT CTGTTTTACGCATGGTCAAAGAGACTGTGGGGAATGCTT CCCTGGAATGTCAGAATCTCAACCCGTTTCTGTCGTCAA AAAGAAGACTTATCAGAAACTGTGTCCAATTCATCATAT CCTGGGAAGGGCACCCGAGATTGCCTGTTCGGCCTGCGA TTTGGCCAATGTGGACTTGGATGACTGTGTTTCTGAGCA ATAAATGACTTAAACCAGGTATGGCTGCTGACGGTTATC TTCCAGATTGGCTCGAGGACAACCTTTCTGA	1035
Rep78 AA	MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWDVPP DSDMDPNLIEQAPLTVAEKLQREFLVEWRRVSKAPEALF FVQFEKGETYFHLHVLIETIGVKSMVVGRYVSQIKEKLV TRIYRGVEPQLPNWFAVTKTRNGAGGNKVVDDCYIPNY LLPKTQPELQWAWTNMDQYLSACLNLAERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDR GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKI MSLTKTAPDYLVGSNPPEDITKNRIYQILELNGYDPQYA ASVFLGWAQKKFGKRNTIWLFGPATTGKTNIAEAIAHAV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIEPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFEFELTRRLDHDFGKVTKQEVKD FFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKR ECTSLAQPTTSDAEAPADYADRYQNKCSRHVGMNLMLFP CKTCERMNQISNVCFTHGQRDCGECFPGMSESQPVSVVK KKTYQKLCPIHHILGRAPEIACSACDLANVDLDDCVSEQ	1036

Rep68 AA	MPGFYEIVLKVPSDLDERLPGISNSFVNWVAEKEWDVPP DSDMDPNLIEQAPLTVAEKLQREFLVEWRRVSKAPEALF FVQFEKGETYFHLHVLIETIGVKSMVVGRYVSQIKEKLV TRIYRGVEPQLPNWFAVTKTRNGAGGGNKVVDDCYIPNY LLPKTQPELQWAWTNMDQYLSACLNLAERKRLVAQHLTH VSQTQEQNKENQNPNSDAPVIRSKTSARYMELVGWLVDR GITSEKQWIQEDQASYISFNAASNSRSQIKAALDNASKI MSLTKTAPDYLVGSNPPEDITKNRIYQILELNGYDPQYA ASVFLGWAQKKFGKRNTIWLFGPATTGKTNIAEAIAHAV PFYGCVNWTNENFPFNDCVDKMVIWWEEGKMTAKVVESA KAILGGSKVRVDQKCKSSAQIEPTPVIVTSNTNMCAVID GNSTTFEHQQPLQDRMFEFELTRRLDHDFGKVTKQEVKD FFRWASDHVTDVAHEFYVRKGGAKKRPASNDADVSEPKR ECTSLAQPTTSDAEAPADYADRLARGQPF	1037
Rep52 AA	MELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNASKIMSLTKTAPDYLVGSNPPEDITKNRIYQIL ELNGYDPQYAASVFLGWAQKKFGKRNTIWLFGPATTGKT NIAEAIAHAVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIEPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFEFELTRRLDHDF GKVTKQEVKDFFRWASDHVTDVAHEFYVRKGGAKKRPAS NDADVSEPKRECTSLAQPTTSDAEAPADYADRYQNKCSR HVGMNLMLFPCKTCERMNQISNVCFTHGQRDCGECFPGM SESQPVSVVKKKTYQKLCPIHHILGRAPEIACSACDLAN VDLDDCVSEQ	1038
Rep40 AA	MELVGWLVDRGITSEKQWIQEDQASYISFNAASNSRSQI KAALDNASKIMSLTKTAPDYLVGSNPPEDITKNRIYQIL ELNGYDPQYAASVFLGWAQKKFGKRNTIWLFGPATTGKT NIAEAIAHAVPFYGCVNWTNENFPFNDCVDKMVIWWEEG KMTAKVVESAKAILGGSKVRVDQKCKSSAQIEPTPVIVT SNTNMCAVIDGNSTTFEHQQPLQDRMFEFELTRRLDHDF GKVTKQEVKDFFRWASDHVTDVAHEFYVRKGGAKKRPAS NDADVSEPKRECTSLAQPTTSDAEAPADYADRLARGQPF	1039
AAV5 Rep 서열		
Rep78 AA	MATFYEVIVRVPFDVEEHLPGISDSFVDWVTGQIWELPP ESDLNLTLVEQPQLTVADRIRRVFLYEWNKFSKQESKFF VQFEKGSEYFHLHTLVETSGISSMVLGRYVSQIRAQLVK VVFQGIEPQINDWVAITKVKKGGANKVVDSGYIPAYLLP KVQPELQWAWTNLDEYKLAALNLEERKRLVAQFLAESSQ RSQEAASQREFSADPVIKSKTSQKYMALVNWLVEHGITS EKQWIQENQESYLSFNSTGNSRSQIKAALDNATKIMSLT KSAVDYLVGSSVPEDISKNRIWQIFEMNGYDPAYAGSIL YGWCQRSFNKRNTVWLYGPATTGKTNIAEAIAHTVPFYG CVNWTNENFPFNDCVDKMLIWWEEGKMTNKVVESAKAIL GGSKVRVDQKCKSSVQIDSTPVIVTSNTNMCVVVDGNST TFEHQQPLEDRMFKFELTKRLPPDFGKITKQEVKDFFAW AKVNQVPVTHEFKVPRELAGTKGAEKSLKRPLGDVTNTS YKSLEKRARLSFVPETPRSSDVTVDPAPLRPLNWNSRYD CKCDYHAQFDNISNKCDECEYLNRGKNGCICHNVTHCQI CHGIPPWEKENLSDFGDFDDANKEQ	1040

[0431]

Rep52 AA	MALVNWLVEHGITSEKQWIQENQESYLSFNSTGNSRSQI KAALDNATKIMSLTKSAVDYLVGSSVPEDISKNRIWQIF EMNGYDPAYAGSILYGWCQRSFNKRNTVWLYGPATTGKT NIAEAIAHTVPFYGCVNWTNENFPFNDCVDKMLIWWEEG KMTNKVVESAKAILGGSKVRVDQKCKSSVQIDSTPVIVT SNTNMCVVVDGNSTTFEHQQPLEDRMFKFELTKRLPPDF GKITKQEVKDFFAWAKVNQVPVTHEFKVPRELAGTKGAE KSLKRPLGDVTNTSYKSLEKRARLSFVPETPRSSDVTVD PAPLRPLNWNSRYDCKCDYHAQFDNISNKCDECEYLNRG KNGCICHNVTHCQICHGIPPWEKENLSDFGDFDDANKEQ	1041
Rep40 AA	MSLTKSAVDYLVGSSVPEDISKNRIWQIFEMNGYDPAYA GSILYGWCQRSFNKRNTVWLYGPATTGKTNIAEAIAHTV PFYGCVNWTNENFPFNDCVDKMLIWWEEGKMTNKVVESA KAILGGSKVRVDQKCKSSVQIDSTPVIVTSNTNMCVVVD GNSTTFEHQQPLEDRMFKFELTKRLPPDFGKITKQEVKD FFAWAKVNQVPVTHEFKVPRELAGTKGAEKSLKRPLGDV TNTSYKSLEKRARLSFVPETPRSSDVTVDPAPLRPLNWN SRYDCKCDYHAQFDNISNKCDECEYLNRGKNGCICHNVT HCQICHGIPPWEKENLSDFGDFDDANKEQ	1042

[0432]

[0433] 일부 구현예에서, 비-아넬로바이러스 바이러스 유래의 분자, 또는 이에 기반한 분자는 헬퍼 작제물을 통해 숙주 세포 내로 도입된다. 일부 구현예에서, 본 명세서에 기재된 방법은 헬퍼 작제물을 숙주 세포(예를 들어, 본 명세서에 기재된 바와 같은 유전 요소 작제물 또는 유전 요소를 포함하는 숙주 세포) 내로 도입하는 단계를 포함한다. 일부 구현예에서, 헬퍼 작제물은 유전 요소 작제물의 도입 이전에 숙주 세포 내로 도입된다. 일부 구현예에서, 헬퍼 작제물은 유전 요소 작제물의 도입과 동시에 숙주 세포 내로 도입된다. 일부 구현예에서, 헬퍼 작제물은 유전 요소 작제물의 도입 후에 숙주 세포 내로 도입된다.

[0434] 일부 구현예에서, 헬퍼 작제물은 비-아넬로바이러스 ORF를 인코딩하는 서열을 포함한다. 일부 구현예에서, 헬퍼 작제물은 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질을 인코딩하는 서열을 포함한다. 일부 구현예에서, 헬퍼 작제물은 AAV REP2 분자를 인코딩하는 서열을 포함한다. 일부 구현예에

서, 하나 이상의 헬퍼 작제물은 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자 중 하나 이상(예를 들어, 1, 2, 또는 3개 전부)을 인코딩하는 서열을 포함한다. 구현예에서, AAV Rep 분자, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자는 동일한 작제물 상에서 인코딩된다. 구현예에서, AAV Rep 분자, 아데노바이러스 E2A 분자, 아데노바이러스 E4 분자, 및 아데노바이러스 VARNA 분자는 상이한 작제물(예를 들어, 2, 3, 또는 4개의 별개 작제물) 상에서 인코딩된다.

[0435] 일부 구현예에서, 헬퍼 작제물은 아넬로바이러스 ORF(예를 들어, 아넬로바이러스 ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, 및/또는 ORF1/2 중 하나 이상)를 인코딩하는 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다.

[0436] *예시적인 세포 유형*

[0437] 아넬로벡터의 생성에 적합한 예시적인 숙주 세포는 제한 없이, 포유동물 세포, 예를 들어, 인간 세포 및 곤충 세포를 포함한다. 일부 구현예에서, 숙주 세포는 인간 세포 또는 세포주이다. 일부 구현예에서, 세포는 면역 세포 또는 세포주, 예를 들어, T 세포 또는 세포주, 암 세포주, 간 세포 또는 세포주, 뉴런, 신경아교세포, 피부 세포, 상피 세포, 중간엽 세포, 혈액 세포, 내피 세포, 눈 세포, 위장 세포, 전구 세포, 전구체 세포, 줄기 세포, 폐 세포, 심장 세포, 또는 근육 세포이다. 일부 구현예에서, 숙주 세포는 동물 세포(예를 들어, 마우스 세포, 래트 세포, 토끼 세포, 또는 햄스터 세포, 또는 곤충 세포)이다.

[0438] 일부 구현예에서, 숙주 세포는 림프계 세포이다. 일부 구현예에서, 숙주 세포는 T 세포 또는 무한증식 T 세포이다. 구현예에서, 숙주 세포는 Jurkat 세포이다. 구현예에서, 숙주 세포는 MOLT 세포(예를 들어, MOLT-4 또는 MOLT-3 세포)이다. 구현예에서, 숙주 세포는 MOLT-4 세포이다. 구현예에서, 숙주 세포는 MOLT-3 세포이다. 일부 구현예에서, 숙주 세포는 급성 림프구성 백혈병(ALL) 세포, 예를 들어, MOLT 세포, 예를 들어, MOLT-4 또는 MOLT-3 세포이다. 일부 구현예에서, 숙주 세포는 B 세포 또는 무한증식 B 세포이다. 일부 구현예에서, 숙주 세포는 유전 요소 작제물(예를 들어, 본 명세서에 기재된 바와 같음)을 포함한다.

[0439] 일부 구현예에서, 숙주 세포는 MOLT 세포(예를 들어, MOLT-4 또는 MOLT-3 세포)이다.

[0440] 일부 구현예에서, 숙주 세포는 급성 림프구성 백혈병(ALL) 세포, 예를 들어, MOLT 세포, 예를 들어, MOLT-4 또는 MOLT-3 세포이다.

[0441] 일부 구현예에서, 숙주 세포는 Expi-293 세포이다. 일부 구현예에서, 숙주 세포는 Expi-293F 세포이다.

[0442] 일 양태에서, 본 개시내용은 단백질성 외부에 봉입되는 유전 요소를 포함하는 아넬로벡터의 제조 방법을 제공하며, 방법은 아넬로벡터 유전 요소를 포함하는 MOLT-4 세포를 제공하는 단계, 및 아넬로벡터 유전 요소가 MOLT-4 세포의 단백질성 외부에 봉입될 수 있도록 하는 조건 하에서 MOLT-4 세포를 인큐베이션하는 단계를 포함한다. 일부 구현예에서, MOLT-4 세포는 단백질성 외부의 일부 또는 전부를 형성하는 하나 이상의 아넬로바이러스 단백질(예를 들어, 아넬로바이러스 ORF1 분자)을 추가로 포함한다. 일부 구현예에서, 아넬로벡터 유전 요소는 MOLT-4 세포에서, 예를 들어, 유전 요소 작제물(예를 들어, 본 명세서에 기재된 바와 같음)로부터 생성된다. 일부 구현예에서, 방법은 MOLT-4 세포 내로 아넬로벡터 유전 요소 작제물을 도입하는 단계를 추가로 포함한다.

[0443] 일 양태에서, 본 개시내용은 단백질성 외부에 봉입되는 유전 요소를 포함하는 아넬로벡터의 제조 방법을 제공하며, 방법은 아넬로벡터 유전 요소를 포함하는 MOLT-3 세포를 제공하는 단계, 및 아넬로벡터 유전 요소가 MOLT-3 세포의 단백질성 외부에 봉입될 수 있도록 하는 조건 하에서 MOLT-3 세포를 인큐베이션하는 단계를 포함한다. 일부 구현예에서, MOLT-3 세포는 단백질성 외부의 일부 또는 전부를 형성하는 하나 이상의 아넬로바이러스 단백질(예를 들어, 아넬로바이러스 ORF1 분자)을 추가로 포함한다. 일부 구현예에서, 아넬로벡터 유전 요소는 MOLT-3 세포에서, 예를 들어, 유전 요소 작제물(예를 들어, 본 명세서에 기재된 바와 같음)로부터 생성된다. 일부 구현예에서, 방법은 MOLT-3 세포 내로 아넬로벡터 유전 요소 작제물을 도입하는 단계를 추가로 포함한다.

[0444] 일부 구현예에서, 숙주 세포는 인간 세포이다. 구현예에서, 숙주 세포는 HEK293T 세포, HEK293F 세포, A549 세포, Jurkat 세포, Raji 세포, Chang 세포, HeLa 세포 Phoenix 세포, MRC-5 세포, NCI-H292 세포 또는 Wi38 세포이다. 일부 구현예에서, 숙주 세포는 비-인간 영장류 세포(예를 들어, Vero 세포, CV-1 세포, 또는 LLCMK2 세포)이다. 일부 구현예에서, 숙주 세포는 뮤린 세포(예를 들어, McCoy 세포)이다. 일부 구현예에서, 숙주 세포는 햄스터 세포(예를 들어, CHO 세포 또는 BHK 21 세포)이다. 일부 구현예에서, 숙주 세포는 MARC-145, MDBK, RK-13, 또는 EEL 세포이다. 일부 구현예에서, 숙주 세포는 상피 세포(예를 들어, 상피 계통의 세포주)이다.

[0445] 일부 구현예에서, 아넬로벡터는 연속 동물 세포주(예를 들어, 연속 증식될 수 있는 무한증식 세포주)에서 배양

된다. 본 발명의 일 구현예에 따르면, 세포주는 돼지 세포주를 포함할 수 있다. 본 발명의 맥락에서 고려되는 세포주는 돼지 신장 상피 세포주 PK-15 및 SK, 단골수성 세포주 3D4/31 및 고환 세포주 ST와 같은 무한증식 돼지 세포주를 포함하지만, 이에 제한되지 않는다.

[0446] ***배양 조건***

[0447] 단백질성 외부의 유전 요소 및 구성성분을 포함하는 숙주 세포는 단백질성 외부 내에 유전 요소를 봉입하기에 적합한 조건 하에서 인큐베이션되어, 아넬로벡터를 생성할 수 있다. 적합한 배양 조건은, 예를 들어, 실시예 4, 5, 7, 8, 9, 10, 11, 또는 15 중 임의의 것에 기재된 것을 포함한다. 일부 구현예에서, 숙주 세포는 액체 배지(예를 들어, Grace's Supplemented(TNM-FH), IPL-41, TC-100, Schneider's Drosophila, SF-900 II SFM, 또는 및 EXPRESS-FIVE™ SFM)에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 부착 배양에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 현탁 배양에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 튜브, 병, 마이크로캐리어 또는 플라스크에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 접시 또는 웰(예를 들어, 플레이트 상의 웰)에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 숙주 세포의 증식에 적합한 조건 하에서 인큐베이션된다. 일부 구현예에서, 숙주 세포는 숙주 세포가 내부에서 생성된 아넬로벡터를 주변 상청액으로 방출하기에 적합한 조건 하에서 인큐베이션된다.

[0448] 본 발명에 따른 아넬로벡터-함유 세포 배양물의 생성은 상이한 규모(예를 들어, 플라스크, 롤러 보틀(roller bottle) 또는 생물반응기)에서 수행될 수 있다. 감염될 세포의 배양을 위해 사용되는 배지는 세포 생존에 필요한 표준 영양물질을 포함하지만, 또한 세포 유형에 따라 추가의 영양물질을 포함할 수 있다. 선택적으로, 배지는 단백질-무함유 및/또는 혈청-무함유일 수 있다. 세포 유형에 따라, 세포는 현탁액 중에서 또는 기질 상에서 배양될 수 있다. 일부 구현예에서, 숙주 세포의 성장을 위하여, 그리고 아넬로벡터의 생성을 위하여 상이한 배지가 사용된다.

[0449] ***수확***

[0450] 숙주 세포에 의해 생성된 아넬로벡터는, 예를 들어, 당업계에 알려진 방법에 따라 수확될 수 있다. 예를 들어, 배양 중인 숙주 세포에 의해 주변 상청액으로 방출된 아넬로벡터는 상청액으로부터 수확될 수 있다(예를 들어, 실시예 4에 기재된 바와 같음). 일부 구현예에서, 상청액은 아넬로벡터를 얻기 위해 숙주 세포로부터 분리된다. 일부 구현예에서, 숙주 세포는 수확 전 또는 수확 동안 용해된다. 일부 구현예에서, 아넬로벡터는 숙주 세포 용해물로부터 수확된다(예를 들어, 실시예 10에 기재된 바와 같음). 일부 구현예에서, 아넬로벡터는 숙주 세포 용해물 및 상청액 둘 다로부터 수확된다. 일부 구현예에서, 아넬로벡터의 정제 및 단리는, 예를 들어, 문헌 [Rinaldi, et al., DNA Vaccines: Methods and Protocols (Methods in Molecular Biology), 3rd ed. 2014, Humana Press](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같이, 바이러스 생성에서 알려진 방법에 따라 수행된다. 일부 구현예에서, 아넬로벡터는 약제학적 부형제를 이용한 제형화 이전에, 생물물리학적 특성에 기반한 용질의 분리, 예를 들어, 이온 교환 크로마토그래피 또는 접선 유동 여과에 의해 수확되고/수확되거나 정제될 수 있다.

[0451] ***시험관 내 조립 방법***

[0452] 아넬로벡터는, 예를 들어, 시험관 내 조립에 의해, 예를 들어, 무세포 현탁액 또는 상청액에서 생성될 수 있다. 일부 구현예에서, 유전 요소는, 예를 들어, 조립을 가능하게 하는 조건 하에서, 시험관 내에서 ORF1 분자에 접촉된다.

[0453] 일부 구현예에서, 바큘로바이러스 작제물은 아넬로바이러스 단백질을 생성하는 데 사용된다. 그 다음 이들 단백질은, 예를 들어, 유전 요소, 예를 들어, RNA를 포함하는 유전 요소를 캡시드화하기 위해 시험관 내 조립에 사용될 수 있다. 일부 구현예에서, 하나 이상의 아넬로바이러스 단백질을 인코딩하는 폴리뉴클레오티드는 숙주 세포에서, 예를 들어, 곤충 또는 동물 세포에서 발현을 위한 프로모터에 융합된다. 일부 구현예에서, 폴리뉴클레오티드는 바큘로바이러스 발현 시스템으로 클로닝된다. 일부 구현예에서, 숙주 세포, 예를 들어, 곤충 세포는 바큘로바이러스 발현 시스템으로 감염되고 일정 기간 동안 인큐베이션된다. 일부 구현예에서, 감염된 세포는 약 1, 2, 3, 4, 5, 10, 15, 또는 20일 동안 인큐베이션된다. 일부 구현예에서, 감염된 세포는 아넬로바이러스 단백질을 회수하기 위해 용해된다.

[0454] 일부 구현예에서, 단리된 아넬로바이러스 단백질은 정제된다. 일부 구현예에서, 아넬로바이러스 단백질은 킬레이트화 정제, 헤파린 정제, 구배 침강 정제, 및/또는 SEC 정제를 포함하지만 이에 제한되지 않는 정제 기법을 사용하여 정제된다. 일부 구현예에서, 정제된 아넬로바이러스 단백질은 유전 요소와 혼합되어 유전 요소, 예를

들어, RNA를 포함하는 유전 요소를 캡시드화한다. 일부 구현예에서, 유전 요소는 ORF1 단백질, ORF2 단백질, 또는 이의 변형된 버전을 사용하여 캡시드화된다. 일부 구현예에서 2개의 핵산은 캡시드화된다. 예를 들어, 제1 핵산은 mRNA, 예를 들어, 화학적으로 변형된 mRNA일 수 있고, 제2 핵산은 DNA일 수 있다.

[0455] 일부 구현예에서, 아넬로바이러스(AV) ORF1(예를 들어, 야생형 ORF1 단백질, 예를 들어, 조립 효율, 수율 또는 안정성을 개선시키기 위한, 돌연변이를 보유하는 ORF1 단백질, 키메라 ORF1 단백질, 또는 이의 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 구현예에서, AV ORF1을 인코딩하는 DNA는 태그되지 않을 수 있다. 일부 구현예에서, AV ORF1을 인코딩하는 DNA는 N-말단 및/또는 C-말단에 융합된 태그를 함유할 수 있다. 일부 구현예에서, AV ORF1을 인코딩하는 DNA는 ORF1 단백질 내에 돌연변이, 삽입 또는 결실을 보유하여, 예를 들어, 면역염색 검정(ELISA 또는 웨스턴 블롯을 포함하지만 이에 제한되지 않음)을 통해, 예를 들어, 정제 및/또는 동일성 결정을 돕기 위해, 태그를 도입할 수 있다. 일부 구현예에서, AV ORF1을 인코딩하는 DNA는 단독으로 또는 임의의 수의 헬퍼 단백질과 조합하여 발현될 수 있다. 일부 구현예에서, AV ORF1을 인코딩하는 DNA는 AV ORF2 및/또는 ORF3 단백질과 조합하여 발현된다.

[0456] 일부 구현예에서, 조립 효율을 개선시키기 위한 돌연변이를 보유하는 ORF1 단백질은 입자 조립(서열번호 3 내지 5)을 촉발시키는 pH 민감성 핵산 결합을 허용하는 ARG 아암(arm)의 pI를 변경하기 위해 N-말단 아르기닌 아암(ARG 아암) 내로 도입되는 돌연변이를 보유하는 ORF1 단백질을 포함할 수 있지만 이에 제한되지 않는다. 일부 구현예에서, 안정성을 개선시키는 돌연변이를 보유하는 ORF1 단백질은 표준 젤리롤 베타-배럴의 베타 가닥 F 및 G와 접촉하는 인터프로토머(interprotomer)에 돌연변이를 포함하여 프로토머 표면의 소수성 상태를 변경하고 캡시드 형성의 열역학적 우호성을 개선시킬 수 있다.

[0457] 일부 구현예에서, 키메라 ORF1 단백질은 또 다른 캡시드 단백질, 예를 들어, 부리 깃털 질환 바이러스(BFDV) 캡시드 단백질, 또는 E형 간염 캡시드 단백질 유래의 비슷한 부분으로 대체된 이들 서열의 일부 또는 일부들, 예를 들어, BFDV 캡시드 단백질 유래의 비슷한 구성성분으로 대체된 Ring 9 ORF1의 ARG 아암 또는 F 및 G 베타 가닥을 갖는 ORF1 단백질을 포함할 수 있지만 이에 제한되지 않는다. 일부 구현예에서, 키메라 ORF1 단백질은 또한 또 다른 AV ORF1 단백질의 비슷한 부분으로 대체된 이들 서열의 일부 또는 일부들(예를 들어, Ring 9 ORF1의 비슷한 부분으로 대체된 Ring 2 ORF1의 젤리롤 단편 또는 C-말단 부분)을 갖는 ORF1 단백질을 포함할 수 있다.

[0458] 일부 구현예에서, 본 개시내용은 아넬로벡터를 제조하는 방법을 기재하며, 방법은 (a) (i) RNA를 포함하는 유전 요소, 및 (ii) ORF1 분자를 포함하는 혼합물을 제공하는 단계; 및 (b) ORF1 분자를 포함하는 단백질성 외부 내부에 유전 요소를 봉입하기에 적합한 조건 하에서 혼합물을 인큐베이션하는 단계를 포함하고 이를 통해 아넬로벡터를 제조하며; 선택적으로 혼합물은 세포에 포함되지 않는다. 일부 구현예에서, 방법은, (a)를 제공하기 이전에, 예를 들어, 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)에서 ORF1 분자를 발현하는 단계를 추가로 포함한다. 일부 구현예에서, 발현하는 단계는 ORF1 분자를 생성하기에 적합한 조건 하에서 ORF1 분자를 인코딩하는 핵산 분자(예를 들어, 바큘로바이러스 발현 벡터)를 포함하는 숙주 세포(예를 들어, 곤충 세포 또는 포유동물 세포)를 인큐베이션하는 것을 포함한다. 일부 구현예에서, 방법은 (a)를 제공하기 이전에, 숙주 세포에 의해 발현된 ORF1 분자를 정제하는 단계를 추가로 포함한다. 일부 구현예에서, 방법은 무세포 시스템에서 수행된다. 일부 구현예에서, 본 개시내용은 아넬로벡터 조성물을 제조하는 방법을 기재하며, (a) 선행하는 구현예 중 어느 하나에 따른 복수의 아넬로벡터 또는 조성물을 제공하는 단계; (b) 선택적으로, 본 명세서에 기재된 오염물, 광학 밀도 측정(예를 들어, OD 260), 입자 수(예를 들어, HPLC에 의함), 감염성(예를 들어, 형광 및/또는 ELISA에 의해 결정된 바와 같은, 예를 들어, 입자:감염 단위 비율) 중 하나 이상에 대하여 복수의 아넬로벡터 또는 조성물을 평가하는 단계; 및 (c) 예를 들어, (b)의 하나 이상의 매개변수가 명시된 임계값을 충족하는 경우, 예를 들어, 대상체에게 투여하기에 적합한 약제학적 조성물로서, 복수의 아넬로벡터를 제형화하는 단계를 포함한다.

[0459] ***농축 및 정제***

[0460] 수확된 아넬로벡터는, 예를 들어, 아넬로벡터 제제를 생성하기 위해, 정제되고/되거나 농축될 수 있다. 일부 구현예에서, 수확된 아넬로벡터는, 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 또는 오염물로부터 단리된다. 일부 구현예에서, 정제 단계는 제제로부터 혈청, 숙주 세포 DNA, 숙주 세포 단백질, 유전 요소가 결여된 입자, 및/또는 페놀 레드 중 하나 이상을 제거하는 것을 포함한다. 일부 구현예에서, 수확된 아넬로벡터는, 예를 들어, 바이러스 입자를 농축시키기 위해 당업계에 알려진 방법을 사용하여, 수확 용액에 존

재하는 다른 구성요소 또는 오염물에 비해 농축된다.

[0461] 일부 구현예에서, 생성된 제제 또는 제제를 포함하는 약제학적 조성물은 허용 가능한 기간 및 온도에 걸쳐 안정할 것이고/이거나, 요망되는 투여 경로 및/또는 투여 경로가 필요로 할 임의의 디바이스, 예를 들어, 니들 또는 주사기와 병용 가능할 것이다.

[0462] **II. 아넬로벡터**

[0463] 일부 양태에서, 본 명세서에 기재된 발명은 아넬로벡터, 아넬로벡터 제제, 및 치료용 조성물을 사용하고 제조하기 위한 조성물 및 방법을 포함한다. 일부 구현예에서, 아넬로벡터는 본 명세서에 기재된 바와 같은 조성물 및 방법을 사용하여 제조된다. 일부 구현예에서, 아넬로벡터는 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스), 또는 이의 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 구현예에서, 아넬로바이로스-기반 아넬로벡터는 해당 아넬로바이러스에 대해 외인성인 적어도 하나의 요소, 예를 들어, 외인성 이펙터 또는 아넬로벡터의 유전 요소 내부에 배치된 외인성 이펙터를 인코딩하는 핵산 서열 및/또는 아넬로바이러스(예를 들어, 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV)) 이외의 바이러스 유래의 외인성 핵산 서열을 포함한다. 일부 구현예에서, 아넬로바이러스-기반 아넬로벡터는 해당 아넬로바이러스 유래의 또 다른 요소에 대하여 이종인 적어도 하나의 요소, 예를 들어, 프로모터 요소와 같은 또 다른 연결된 핵산 서열에 대하여 이종인 이펙터-인코딩 핵산 서열을 포함한다. 일부 구현예에서, 아넬로벡터는 유전 요소의 나머지 및/또는 단백질성 외부에 비하여 이종인 적어도 하나의 요소(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 이펙터를 인코딩하는 외인성 요소)를 포함하는 유전 요소(예를 들어, 환형 DNA, 예를 들어, 단일 가닥 DNA)를 포함한다. 아넬로벡터는 숙주, 예를 들어, 인간 내로의 페이로드를 위한 운반 비히클(예를 들어, 실질적으로 비-병원성 운반 비히클)일 수 있다. 일부 구현예에서, 아넬로벡터는 진핵 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포에서 복제할 수 있다. 일부 구현예에서, 아넬로벡터는 포유동물(예를 들어, 인간) 세포에서 실질적으로 비-병원성 및/또는 실질적으로 비-통합성이다. 일부 구현예에서, 아넬로벡터는 포유동물, 예를 들어, 인간에서 실질적으로 비-면역원성이다. 일부 구현예에서, 아넬로벡터는 복제-결핍이다. 일부 구현예에서, 아넬로벡터는 복제-적격이다.

[0464] 일부 구현예에서, 아넬로벡터는, 예를 들어, PCT 출원 PCT/US2018/037379(이는 본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같은, 큐론 또는 이의 구성성분(예를 들어, 이펙터를 인코딩하는 서열, 및/또는 단백질성 외부를 포함하는, 예를 들어, 유전 요소)을 포함한다. 일부 구현예에서, 아넬로벡터는, 예를 들어, PCT 출원 PCT/US19/65995(이는 본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같은, 아넬로벡터, 또는 이의 구성성분(예를 들어, 이펙터를 인코딩하는 서열, 및/또는 단백질성 외부를 포함하는, 예를 들어, 유전 요소)을 포함한다.

[0465] 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, 하기의 특성 중 하나 또는 둘 다를 갖는 유전 요소: 환형이고/이거나 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 아넬로벡터를 포함하며; 유전 요소는 단백질성 외부 내에 봉입되고; 아넬로벡터는 유전 요소를 진핵 세포 내로 운반할 수 있다.

[0466] 본 명세서에 기재된 아넬로벡터의 일부 구현예에서, 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 구현예에서, 대상체에 투여되는 복수의 아넬로벡터로부터 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 구현예에서, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터의 집단의 유전 요소는 비슷한 AAV 바이러스의 집단의 것보다 더 낮은 빈도로, 예를 들어, 비슷한 AAV 바이러스의 집단보다 약 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 또는 그 이상 더 낮은 빈도로 숙주 세포의 게놈 내로 통합된다.

[0467] 일 양태에서, 본 발명은 (i) 프로모터 요소 및 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터, 예를 들

어, 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, 본 명세서에 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 아넬로벡터를 포함하며; 유전 요소는 단백질성 외부 내에 봉입되며; 아넬로벡터는 유전 요소를 진핵 세포 내로 운반할 수 있다.

[0468] 일 양태에서, 본 발명은 다음을 포함하는 아넬로벡터를 포함한다:

[0469] 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

[0470] b) 유전 요소와 회합된, 예를 들어, 이를 둘러싸거나 봉입하는 단백질성 외부.

[0471] 일부 구현예에서, 아넬로벡터는 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 게놈을 갖는 단일 가닥 DNA(ssDNA) 바이러스의 하위군을 지칭한다. 따라서, 동물 환형 ssDNA 바이러스는 원핵생물을 감염시키는 ssDNA 바이러스(즉, 마이크로바리대(Microviridae) 및 이노비리대(Inoviridae)) 및 식물을 감염시키는 ssDNA 바이러스(즉, 제미니비리대(Geminiviridae) 및 나노비리대(Nanoviridae))와 구별 가능하다. 동물 환형 ssDNA 바이러스는 또한 비-식물 진핵생물을 감염시키는 선형 ssDNA 바이러스(즉, 파보비리대(Parvoviridiae))와 구별 가능하다.

[0472] 일부 구현예에서, 아넬로벡터는 숙주 세포 기능을, 예를 들어, 일시적으로 또는 장기간 조절한다. 특정 구현예에서, 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, 또는 더 긴 시간 동안 또는 이들 사이의 임의의 시간 동안 지속된다. 특정 구현예에서, 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.

[0473] 일부 구현예에서, 유전 요소는 프로모터 요소를 포함한다. 구현예에서, 프로모터 요소는 RNA 중합효소 II-의존성 프로모터, RNA 중합효소 III-의존성 프로모터, PGK 프로모터, CMV 프로모터, EF-1α 프로모터, SV40 프로모터, CAGG 프로모터, 또는 UBC 프로모터, TTV 바이러스 프로모터, 조직 특이적, U6(pollIII), 활성화 단백질(TetR-VP16, Gal4-VP16, dCas9-VP16 등)에 대한 상류 DNA 결합 부위가 있는 최소 CMV 프로모터로부터 선택된다. 구현예에서, 프로모터 요소는 TATA 박스를 포함한다. 구현예에서, 프로모터 요소는, 예를 들어, 본 명세서에 기재된 바와 같은 야생형 아넬로바이러스에 대하여 내인성이다.

[0474] 일부 구현예에서, 유전 요소는 하기의 특성 중 하나 이상을 포함한다: 단일-가닥, 환형, 음성 가닥, 및/또는 DNA. 구현예에서, 유전 요소는 에피솜을 포함한다. 일부 구현예에서, 이펙터를 배제한 유전 요소의 일부는 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.

[0475] 본 명세서에 기재된 바와 같은 아넬로벡터, 아넬로벡터를 포함하는 조성물, 이러한 아넬로벡터의 사용 방법 등은 일부 예에서 부분적으로, 상이한 이펙터, 예를 들어, (예를 들어, IFN 또는 miR-625에 대한) miRNA, shRNA 등 및 단백질 결합 서열, 예를 들어, 캡시드 단백질, 예컨대 Q99153에 결합하는 DNA 서열을 단백질성 외부, 예를 들어, 문헌[Arch Virol (2007) 152: 1961-1975]에 개시된 캡시드와 조합하여, 이어서 이펙터를 세포(예를 들어, 동물 세포, 예를 들어, 인간 세포 또는 비-인간 동물 세포, 예컨대 돼지 또는 마우스 세포)로 운반하는 데 사용될 수 있는 아넬로벡터를 생성하는 방법을 예시하는 실시예에 기초한다. 구현예에서, 이펙터는 인터페론과 같은 인자의 발현을 침묵화시킬 수 있다. 실시예에는 이펙터를, 예를 들어, 아넬로바이러스로부터 유래된 서열 내로 삽입함으로써 아넬로벡터가 제조될 수 있는 방법이 추가로 기재된다. 이들 실시예에 기초하여 이하의

설명은 실시예에서 고려되는 특정 발견 및 조합의 다양한 변형을 고려한다. 예를 들어, 당업자는 특정 miRNA가 단지 이펙터의 예로서 사용되며, 다른 이펙터가, 예를 들어, 다른 조절 핵산 또는 치료용 펩티드일 수 있음을 실시예로부터 이해할 것이다. 유사하게, 실시예에 사용되는 특정 캡시드는 하기 기재되는 실질적으로 비-병원성인 단백질에 의해 대체될 수 있다. 실시예에 기재된 특정 아넬로바이러스 서열은 또한 하기 기재되는 아넬로바이러스 서열에 의해 대체될 수 있다. 이들 고려사항은 단백질 결합 서열, 조절 서열, 예컨대 프로모터 등에 유사하게 적용된다. 이와 독립적으로, 당업자는 특히 실시예와 밀접하게 관련이 있는 구현예를 고려할 것이다.

[0476] 일부 구현예에서, 아넬로벡터, 또는 아넬로벡터에 포함되는 유전 요소는 세포(예를 들어, 인간 세포) 내로 도입된다. 일부 구현예에서, 예를 들어, 일단 아넬로벡터 또는 유전 요소가 세포 내로 도입되면, 예를 들어, 아넬로벡터의 유전 요소에 의해 인코딩된 이펙터(예를 들어, RNA, 예를 들어, miRNA)는 세포(예를 들어, 인간 세포)에서 발현된다. 구현예에서, 아넬로벡터 또는 그 내부에 포함된 유전 요소의 세포 내로의 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). 구현예에서, 아넬로벡터 또는 그 내부에 포함된 유전 요소의 도입은 세포에 의해 생성되는 인터페론의 수준을 감소시킨다. 구현예에서, 아넬로벡터 또는 그 내부에 포함된 유전 요소의 세포 내로의 도입은 세포의 기능을 조절한다(예를 들어, 증가시키거나 감소시킨다). 구현예에서, 아넬로벡터 또는 그 내부에 포함된 유전 요소의 세포 내로의 도입은 세포의 생존력을 조절한다(예를 들어, 증가시키거나 감소시킨다). 구현예에서, 아넬로벡터 또는 그 내부에 포함된 유전 요소의 세포 내로의 도입은 세포(예를 들어, 암 세포)의 생존력을 감소시킨다.

[0477] 일부 구현예에서, 본 명세서에 기재된 아넬로벡터(예를 들어, 합성 아넬로벡터)는 70% 미만의 항체 출현율(예를 들어, 약 60%, 50%, 40%, 30%, 20% 또는 10% 미만의 항체 출현율)을 유도한다. 구현예에서, 항체 출현율은 당업계에 알려진 방법에 따라 결정된다. 구현예에서, 항체 출현율은, 예를 들어, 문헌[Tsuda et al., 1999; *J. Virol. Methods 77*: 199-206](본 명세서에 참조로 포함됨)에 기재된 항-TTV 항체 검출 방법 및/또는 문헌[Kakkola et al., 2008; *Virology* 382: 182-189](본 명세서에 참조로 포함됨)에 기재된 항-TTV IgG 혈청학적 출현율의 결정 방법에 따라 생물학적 샘플에서 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 또는 그에 기반한 아넬로벡터에 대한 항체를 검출함으로써 결정된다. 아넬로바이러스 또는 이에 기초한 아넬로벡터에 대한 항체는 또한 항-바이러스 항체를 검출하기 위한 당업계의 방법, 예를 들어, 문헌[Calcedo et al., 2013; *Front. Immunol.* 4(341): 1-7](본 명세서에 참조로 포함됨)에 기재된 바와 같은 예를 들어, 항-AAV 항체의 검출 방법에 의해 검출될 수 있다.

[0478] 일부 구현예에서, 복제 결핍, 복제 결함, 또는 복제 비적격 유전 요소는 유전 요소의 복제에 필요한 필수 기구 또는 구성성분의 전부를 인코딩하지 않는다. 일부 구현예에서, 복제 결함 유전 요소는 복제 인자를 인코딩하지 않는다. 일부 구현예에서, 복제 결함 유전 요소는 하나 이상의 ORF(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 및/또는 ORF2t/3)를 인코딩하지 않는다. 일부 구현예에서, 유전 요소에 의해 인코딩되지 않는 기구 또는 구성성분은 (예를 들어, 헬퍼, 예를 들어, 헬퍼 바이러스 또는 헬퍼 플라스미드를 사용하여, 또는 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.

[0479] 일부 구현예에서, 패키징 결핍, 패키징 결함, 또는 패키징 부적격 유전 요소는 단백질성 외부(예를 들어, 단백질성 외부는, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함하는, 캡시드 또는 이의 일부를 포함함)에 패키징될 수 없다. 일부 구현예에서, 패키징 결핍 유전 요소는 야생형 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음)와 비교하여 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 외부에 패키징된다. 일부 구현예에서, 패키징 결함 유전 요소는 심지어 (예를 들어, 본 명세서에 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 외부에 패키징될 수 없다. 일부 구현예에서, 패키징 결핍 유전 요소는 심지어 (예를 들어, 본 명세서에 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, 본 명세서에 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 외부에 패키징된다.

[0480] 일부 구현예에서, 패키징 적격 유전 요소는 단백질성 외부(예를 들어, 단백질성 외부는, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1 핵산에 의해 인코딩된 폴리펩티드를 포함하는, 캡시드 또는 이의 일부를

포함함)에 패키징될 수 있다. 일부 구현예에서, 패키징 적격 유전 요소는 야생형 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음)와 비교하여 적어도 20%(예를 들어, 적어도 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%, 또는 그 이상)의 효율로 단백질성 외부에 패키징된다. 일부 구현예에서, 패키징 적격 유전 요소는 (예를 들어, 본 명세서에 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 외부에 패키징될 수 있다. 일부 구현예에서, 패키징 적격 유전 요소는 (예를 들어, 본 명세서에 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, 본 명세서에 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 외부에 패키징된다.

[0481] *아넬로바이러스*

[0482] 일부 구현예에서, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터는 아넬로바이러스로부터 유래된 서열 또는 발현 생성물을 포함한다. 일부 구현예에서, 아넬로벡터는 아넬로바이러스에 비하여 외인성인 하나 이상의 서열 또는 발현 생성물을 포함한다. 일부 구현예에서, 아넬로벡터는 아넬로바이러스에 비하여 내인성인 하나 이상의 서열 또는 발현 생성물을 포함한다. 일부 구현예에서, 아넬로벡터는 아넬로벡터 내의 하나 이상의 다른 서열 또는 발현 생성물에 비하여 이종인 하나 이상의 서열 또는 발현 생성물을 포함한다. 아넬로바이러스는 일반적으로 음의 극성을 갖는 단일-가닥 환형 DNA 게놈을 갖는다. 아넬로바이러스는 일반적으로 임의의 인간 질환과 연관된 바 없다. 그러나, 아넬로바이러스 감염을 인간 질환과 연관시키려는 시도는 대조군 코호트 집단(들)에서의 무증상 아넬로바이러스 바이러스혈증의 높은 유병률, 아넬로바이러스 과 내의 현저한 게놈 다양성, 종래에 시험관 내에서 작용제를 전파할 수 없음, 아넬로바이러스 질환의 동물 모델(들)의 결여에 의해 혼동되었다(문헌[Yzebe et al., Panminerva Med. (2002) 44:167-177]; 문헌[Biagini, P., Vet. Microbiol. (2004) 98:95-101]).

[0483] 아넬로바이러스는 일반적으로 구비강 또는 배변-구강 감염, 모에서 유아로의 전염 및/또는 자궁 내의 전염에 의해 전염된다(문헌[Gerner et al., Ped. Infect. Dis. J. (2000) 19:1074-1077]). 감염된 사람은 일부 예에서, 장기의(수개월 내지 수년) 아넬로바이러스 바이러스혈증이라는 특징이 나타날 수 있다. 인간은 한가지 초과의 유전자군 또는 바이러스주로 동시-감염될 수 있다(문헌[Saback, et al., Scad. J. Infect. Dis. (2001) 33:121-125]). 이들 유전자군이 감염된 인간 내에서 재조합될 수 있다는 제안이 존재한다(문헌[Rey et al., Infect. (2003) 31:226-233]). 이중 가닥 아이소폼(복제성) 중간체는 몇몇의 조직, 예컨대 간, 말초 혈액 단핵 세포 및 골수에서 관찰된 바 있다(문헌[Kikuchi et al., J. Med. Virol. (2000) 61:165-170]; 문헌[Okamoto et al., Biochem. Biophys. Res. Commun. (2002) 270:657-662]; 문헌[Rodriguez-Inigo et al., Am. J. Pathol. (2000) 156:1227-1234]).

[0484] 일부 구현예에서, 유전 요소는 아미노산 서열 또는 이의 기능적 단편, 또는 본 명세서에 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.

[0485] 일부 구현예에서, 본 명세서에 기재된 바와 같은 아넬로벡터는 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 서열 또는 이의 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, 본 명세서에 기재된 바와 같은 유전 요소)를 포함한다.

[0486] 일부 구현예에서, 본 명세서에 기재된 바와 같은 아넬로벡터는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스의 TATA 박스, cap 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 이의 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, 본 명세서에 기재된 바와 같은 유전 요소)를 포함한다. 일부 구현예에서, 핵산 분자는 캡시드 단백질을 인코딩하는 서열, 예를 들어, 본 명세서에 기재된 아넬로바이러스 중 임의의 것의 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 서열을 포함한다. 구현예에서, 핵산 분자는 아넬로바이러스 ORF1 단백질(또는 이의 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 인코딩하

는 서열을 포함한다.

[0487] 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 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. 구현예에서, 핵산 분자는 표 A1의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 A1의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 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. 구현예에서, 핵산 분자는 표 A1의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 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. 구현예에서, 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.

[0488] 구현예에서, 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B1의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B1의 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. 구현예에서, 핵산 분자는 표 B1의 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.

[0489] 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B3의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 B3의 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. 구현예에서, 핵산 분자는 표 B3의 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.

[0490] 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 아넬로바이러스 TAIP 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 C1의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%,

80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 C1의 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. 구현예에서, 핵산 분자는 표 C1의 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.

[0491] 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 E1의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 E1의 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. 구현예에서, 핵산 분자는 표 E1의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 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. 구현예에서, 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.

[0492] 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F1의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%,

75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F1의 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. 구현예에서, 핵산 분자는 표 F1의 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.

[0493] 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F3의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F3의 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. 구현예에서, 핵산 분자는 표 F3의 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.

[0494] 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 아넬로바이러스 개시인자 요소 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F5의 아넬로바이러스 전사 시작 부위 뉴클레오티드 서열에 대하여 적어도 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 아넬로바이러스 3개의 오픈-리딩 프레임 영역 뉴클레오티드 서열에 대하여 적어도 약 70%,

75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 핵산 서열을 포함한다. 구현예에서, 핵산 분자는 표 F5의 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. 구현예에서, 핵산 분자는 표 F5의 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.

[0495] 일부 구현예에서, 유전 요소는 아미노산 서열 또는 이의 기능적 단편, 또는 본 명세서에 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.

[0496] 일부 구현예에서, 본 명세서에 기재된 바와 같은 아넬로벡터는 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 서열 또는 이의 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, 본 명세서에 기재된 바와 같은 유전 요소)를 포함한다. 구현예에서, 아넬로벡터는 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. 구현예에서, 아넬로벡터는 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.

[0497] 일부 구현예에서, 본 명세서에 기재된 바와 같은 아넬로벡터는 본 명세서에 기재된 아넬로바이러스(예를 들어, 표 A 내지 M 중 어느 하나에서 거론되거나, 그에 열거된 서열에 의해 인코딩된 아넬로바이러스 서열) 중 어느 하나의 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 이의 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, 본 명세서에 기재된 바와 같은 유전 요소)를 포함한다. 일부 구현예에서, 핵산 분자는 캡시드 단백질을 인코딩하는 서열, 예를 들어, 본 명세서에 기재된 아넬로바이러스(예를 들어, 표 A 내지 M 중 어느 하나에서 거론되거나, 그에 열거된 서열에 의해 인코딩된 아넬로바이러스 서열) 중 어느 하나의 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 서열을 포함한다. 구현예에서, 핵산 분자는 아넬로바이러스 ORF1 또는 ORF2 단백질(예를 들어, 표 A2 내지 M2 중 어느 하나에 나타낸 바와 같은 ORF1 또는 ORF2 아미노산 서열, 또는 표 A1 내지 M1 중 어느 하나에 나타낸 바와 같은 핵산 서열에 의해 인코딩된 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. 구현예에서, 핵산 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 표 A2 내지 M2 중 어느 하나에 나타낸 바와 같은 ORF1 아미노산 서열 또는 표 A1 내지 M1 중 어느 하나에 나타낸 바와 같은 핵산 서열에 의해 인코딩된 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.

[0498] 일부 구현예에서, 본 명세서에 기재된 바와 같은 아넬로벡터는 키메라 아넬로벡터이다. 일부 구현예에서, 키메라 아넬로벡터는 아넬로바이러스 이외의 바이러스 유래의 하나 이상의 요소, 폴리펩티드, 또는 핵산을 추가로 포함한다.

[0499] 구현예에서, 키메라 아넬로벡터는 복수의 상이한 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음)로부터의 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 아넬로벡터는 하나의 아넬로바이러스 유래의 ORF1 분자(예를 들어, Ring1 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또한 99% 아미노산 서열 동일성을 갖는 ORF1 분자) 및 상이한 아넬로바이러스 유래의 ORF2 분자(예를 들어, Ring2 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 아넬로벡터는 아넬로바이러스 유래의 제1 ORF1 분자(예를 들어, Ring1 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 아미노산 서열 동일성을 갖는 ORF1 분자) 및 상이한 아넬로바이러스 유래의 제2 ORF1 분자(예를 들어, Ring2 ORF1 분자, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 아미노산을 갖는 ORF1 분자)를 포함할 수 있다.

[0500] 일부 구현예에서, 아넬로벡터는, 예를 들어, 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의

적어도 하나의 부분 및 상이한 바이러스 유래의 적어도 하나의 부분(예를 들어, 본 명세서에 기재된 바와 같음)을 포함하는, 키메라 폴리펩티드(예를 들어, 아넬로바이러스 ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, 및/또는 ORF2t/3)를 포함한다.

[0501] 일부 구현예에서, 아넬로벡터는, 예를 들어, 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 적어도 하나의 부분 및 상이한 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 키메라 ORF1 분자는 하나의 아넬로바이러스 유래의 ORF1 젤리-롤 도메인, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스 유래의 ORF1 아미노산 서열(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 키메라 ORF1 분자는 하나의 아넬로바이러스 유래의 ORF1 아르기닌-풍부 영역, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스 유래의 ORF1 아미노산 서열(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 키메라 ORF1 분자는 하나의 아넬로바이러스 유래의 ORF1 초가변 도메인, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스 유래의 ORF1 아미노산 서열(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 키메라 ORF1 분자는 하나의 아넬로바이러스 유래의 ORF1 N22 도메인, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스 유래의 ORF1 아미노산 서열(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 키메라 ORF1 분자는 하나의 아넬로바이러스 유래의 ORF1 C-말단 도메인, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열, 및 상이한 아넬로바이러스 유래의 ORF1 아미노산 서열(예를 들어, 본 명세서에 기재된 바와 같음), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다.

[0502] 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는

하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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. 구현예에서, 아넬로벡터는 하나의 아넬로바이러스(예를 들어, 본 명세서에 기재된 바와 같음) 유래의 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, 본 명세서에 기재된 바와 같음) 유래의 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.

[0503] 본 명세서에 포함된 서열 또는 하위서열이 (예를 들어, 본 명세서에 기재된 바와 같이, 예를 들어, 아넬로벡터의 유전 요소를 형성하기 위해) 본 명세서에 기재된 조성물 및 방법에서 이용될 수 있는 추가의 예시적인 아넬로바이러스 게놈은, 예를 들어, PCT 출원 PCT/US2018/037379 및 PCT/US19/65995(본 명세서에 전문이 참조로 포함됨)에 기재되어 있다. 일부 구현예에서, 예시적인 아넬로바이러스 서열은 PCT/US19/65995(본 명세서에 참조로 포함됨)의 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 구현예에서, 예시적인 아넬로바이러스 서열은 PCT/US19/65995(본 명세서에 참조로 포함됨)의 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 구현예에서, 예시적인 아넬로바이러스 서열은, 예를 들어, PCT/US19/65995(본 명세서에 참조로 포함됨)의 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 이를 인코딩하는 핵산 서열을 포함한다.

[0504] [표 A1] 예시적인 아넬로바이러스 핵산 서열(알파토르크바이러스, 클레이드 3)

명칭	Ring1
속/클레이드	*알파토르크바이러스*, 클레이드 3
수탁번호	AJ620231.1

전체 서열 : 3753 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TGCTACGTCACTAACCCACGTGTCCTCTACAGGCCAATCGCAGTCTATGT
CGTGCACTTCCTGGGCATGGTCTACATAATTATATAAATGCTTGCACTTC
CGAATGGCTGAGTTTTTGCTGCCCGTCCGCGGAGAGGAGCCACGGCAGGG
GATCCGAACGTCCTGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGAAG
TCAAGGGGCAATTCGGGCTCAGGACTGGCCGGGCTTTGGGCAAGGCTCTT
AAAAATGCACTTTTCTCGAATAAGCAGAAAGAAAAGGAAAGTGCTACTGC
TTTGCGTGCCAGCAGCTAAGAAAAAACCAACTGCTATGAGCTTCTGGAAA
CCTCCGGTACACAATGTCACGGGGATCCAACGCATGTGGTATGAGTCCTT
TCACCGTGGCCACGCTTCTTTTTGTGGTTGTGGGAATCCTATACTTCACA
TTACTGCACTTGCTGAAACATATGGCCATCCAACAGGCCCGAGACCTTCT
GGGCCACCGGGAGTAGACCCCAACCCCCACATCCGTAGAGCCAGGCCTGC
CCCGGCCGCTCCGGAGCCCTCACAGGTTGATTCGAGACCAGCCCTGACAT
GGCATGGGGATGGTGGAAGCGACGGAGGCGCTGGTGGTTCCGGAAGCGGT
GGACCCGTGGCAGACTTCGCAGACGATGGCCTCGATCAGCTCGTCGCCGC
CCTAGACGACGAAGAGTAAGGAGGCGCAGACGGTGGAGGAGGGGGAGACG
AAAAACAAGGACTTACAGACGCAGGAGACGCTTTAGACGCAGGGGACGAA
AAGCAAAACTTATAATAAAACTGTGGCAACCTGCAGTAATTAAAAGATGC
AGAATAAAGGGATACATACCACTGATTATAAGTGGGAACGGTACCTTTGC
CACAAACTTTACCAGTCACATAAATGACAGAATAATGAAAGGCCCCTTCG
GGGGAGGACACAGCACTATGAGGTTCAGCCTCTACATTTTGTTTGAGGAG
CACCTCAGACACATGAACTTCTGGACCAGAAGCAACGATAACCTAGAGCT
AACCAGATACTTGGGGGCTTCAGTAAAAATATACAGGCACCCAGACCAAG
ACTTTATAGTAATATACAACAGAAGAACCCCTCTAGGAGGCAACATCTAC
ACAGCACCCTCTCTACACCCAGGCAATGCCATTTTAGCAAAACACAAAAT
```

[0505]

ATTAGTACCAAGTTTACAGACAAGACCAAAGGGTAGAAAAGCAATTAGAC
TAAGAATAGCACCCCCCACACTCTTTACAGACAAGTGGTACTTTCAAAAG
GACATAGCCGACCTCACCCTTTTCAACATCATGGCAGTTGAGGCTGACTT
GCGGTTTCCGTTCTGCTCACCACAAACTGACAACACTTGCATCAGCTTCC
AGGTCCTTAGTTCCGTTTACAACAACTACCTCAGTATTAATACCTTTAAT
AATGACAACTCAGACTCAAAGTTAAAAGAATTTTTAAATAAAGCATTTCC
AACAACAGGCACAAAAGGAACAAGTTTAAATGCACTAAATACATTTAGAA
CAGAAGGATGCATAAGTCACCCACAACTAAAAAAACCAAACCCACAAATA
AACAAACCATTAGAGTCACAATACTTTGCACCTTTAGATGCCCTCTGGGG
AGACCCCATATACTATAATGATCTAAATGAAAACAAAAGTTTGAACGATA
TCATTGAGAAAATACTAATAAAAAACATGATTACATACCATGCAAAACTA
AGAGAATTTCCAAATTCATACCAAGGAAACAAGGCCTTTTGCCACCTAAC
AGGCATATACAGCCCACCATACCTAAACCAAGGCAGAATATCTCCAGAAA
TATTTGGACTGTACACAGAAATAATTTACAACCCTTACACAGACAAAGGA
ACTGGAAACAAAGTATGGATGGACCCACTAACTAAAGAGAACAACATATA
TAAAGAAGGACAGAGCAAATGCCTACTGACTGACATGCCCCTATGGACTT
TACTTTTTGGATATACAGACTGGTGTAAAAAGGACACTAATAACTGGGAC
TTACCACTAAACTACAGACTAGTACTAATATGCCCTTATACCTTTCCAAA
ATTGTACAATGAAAAAGTAAAAGACTATGGGTACATCCCGTACTCCTACA
AATTCGGAGCGGGTCAGATGCCAGACGGCAGCAACTACATACCCTTTCAG
TTTAGAGCAAAGTGGTACCCCACAGTACTACACCAGCAACAGGTAATGGA
GGACATAAGCAGGAGCGGGCCCTTTGCACCTAAGGTAGAAAAACCAAGCA
CTCAGCTGGTAATGAAGTACTGTTTTAACTTTAACTGGGGCGGTAACCCT
ATCATTGAACAGATTGTTAAAGACCCCAGCTTCCAGCCCACCTATGAAAT
ACCCGGTACCGGTAACATCCCTAGAAGAATACAAGTCATCGACCCGCGGG
TCCTGGGACCGCACTACTCGTTCCGGTCATGGGACATGCGCAGACACACA
TTTAGCAGAGCAAGTATTAAGAGAGTGTCAGAACAACAAGAAACTTCTGA
CCTTGTATTCTCAGGCCCAAAAAAGCCTCGGGTCGACATCCCAAAACAAG
AAACCCAAGAAGAAAGCTCACATTCACTCCAAAGAGAATCGAGACCGTGG
GAGACCGAGGAAGAAAGCGAGACAGAAGCCCTCTCGCAAGAGAGCCAAGA
GGTCCCCTTCCAACAGCAGTTGCAGCAGCAGTACCAAGAGCAGCTCAAGC
TCAGACAGGGAATCAAAGTCCTCTTCGAGCAGCTCATAAGGACCCAACAA
GGGGTCCATGTAAACCCATGCCTACGGTAGGTCCCAGGCAGTGGCTGTTT
CCAGAGAGAAAGCCAGCCCCAGCTCCTAGCAGTGGAGACTGGGCCATGGA
GTTTCTCGCAGCAAAAATATTTGATAGGCCAGTTAGAAGCAACCTTAAAG
ATACCCCTTACTACCCATATGTTAAAAACCAATACAATGTCTACTTTGAC
CTTAAATTTGAATAAACAGCAGCTTCAAACTTGCAAGGCCGTGGGAGTTT

[0506] CACTGGTCGGTGTCTACCTCTAAAGGTCACTAAGCACTCCGAGCGTAAGC

GAGGAGTGCGACCCTCCCCCCTGGAACAACTTCTTCGGAGTCCGGCGCTA
CGCCTTCGGCTGCGCCGGACACCTCAGACCCCCCCTCCACCCGAAACGCT
TGCGCGTTTCGGACCTTCGGCGTCGGGGGGGTCGGGAGCTTTATTAAACG
GACTCCGAAGTGCTCTTGGACACTGAGGGGGTGAACAGCAACGAAAGTGA
GTGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGT
GTCCGGGGTCGCCATAGGCTTCGGGCTCGTTTTTAGGCCTTCCGGACTAC
AAAAATCGCCATTTTGGTGACGTCACGGCCGCCATCTTAAGTAGTTGAGG
CGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAA
TGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACA
CGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACT
TCCTTCCTCTTTTTCAAAAAAAAGCGGAAGTGCCGCCGCGGCGGCGGGGG
GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCGCCCCCCCCC
GCGCATGCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCCCCCCGGCCCCCC
CCG (서열번호 16)

주석:

추정의 도메인	*염기 범위*
TATA 박스	83 – 88
Cap 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
ORF2t/3	336 – 388 ; 2565 – 3015
ORF1	599 – 2830
ORF1/1	599 – 715 ; 2363 – 2830
ORF1/2	599 – 715 ; 2565 – 2789
3개의 오픈-리딩 프레임 영역	2551 – 2786
폴리(A) 신호	3011 – 3016
GC-풍부 영역	3632 – 3753

[0507]

[0508] [표 A2] 예시적인 아넬로바이러스 아미노산 서열(알파토르크바이러스, 클레이드 3)

Ring1 (*알파토르크바이러스* 클레이드 3)	
ORF2	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEE (서열번호 17)
ORF2/2	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEELLKTPASSPPMKYPVPVTSLEEYKSSTRGSWDRTTRSGHGT CADTHLAEQVLRECQNNKKLLTLYSQAQKSLGSTSQNKKPKKKAHIHSKENRDRG RPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSSSSS (서열번호 18)
ORF2/3	MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEEPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDRSPLA REPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQWLFP ERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLKFE (서열번호 19)
ORF2t/3	MSFWKPPVHNVTGIQRMWPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDR SPLAREPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQ WLFPERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLK FE (서열번호 20)
ORF1	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGRRK TRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDR IMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQ DFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFT DKWYFQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFN NDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQ YFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFC HLTGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQS KCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDY GYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSGPFAPKVEKP STQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSF RSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQRESR PWETEEESETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHV NPCLR (서열번호 21)
ORF1/1	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRIVKDPSFQPTYEIPG TGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPR VDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQEVPFQQQLQQQYQEQLKL RQGIKVLFEQLIRTQQGVHVNPCLR (서열번호 22)
ORF1/2	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRAQKSLGSTSQNKK PKKKAHIHSKENRDRGRPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSS SSS (서열번호 23)

[0509]

[0510] [표 B1] 예시적인 아넬로바이러스 핵산 서열(베타토르크바이러스)

명칭	Ring2
속/클레이드	*베타토르크바이러스*
수탁번호	JX134045.1

전체 서열 : 2797 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCCCCTTTTTGCAACAACTTCTGCTTTTTTCCAACTGC
CGGAAAACCACATAATTTGCATGGCTAACCACAAACTGATATGCTAATTA
ACTTCCACAAAACAACTTCCCCTTTTAAAACCACACCTACAAATTAATTA
TTAAACACAGTCACATCCTGGGAGGTACTACCACACTATAATACCAAGTG
CACTTCCGAATGGCTGAGTTTATGCCGCTAGACGGAGAACGCATCAGTTA
CTGACTGCGGACTGAACTTGGGCGGGTGCCGAAGGTGAGTGAAACCACCG
AAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAGAAA
CTTAAAATTATTTTATTTTTCAGATGAGCGACTGCTTTAAACCAACATGC
TACAACAACAAAACAAAGCAAACTCACTGGATTAATAACCTGCATTTAAC
CCACGACCTGATCTGCTTCTGCCCAACACCAACTAGACACTTATTACTAG
CTTTAGCAGAACAACAAGAAACAATTGAAGTGTCTAAACAAGAAAAAGAA
AAAATAACAAGATGCCTTATTACTACAGAAGAAGACGGTACAACTACAGA
CGTCCTAGATGGTATGGACGAGGTTGGATTAGACGCCCTTTTCGCAGAAG
ATTTCGAAGAAAAAGAAGGGTAAGACCTACTTATACTACTATTCCTCTAA
AGCAATGGCAACCGCCATATAAAAGAACATGCTATATAAAAGGACAAGAC
TGTTTAATATACTATAGCAACTTAAGACTGGGAATGAATAGTACAATGTA
TGAAAAAAGTATTGTACCTGTACATTGGCCGGGAGGGGGTTCTTTTTCTG
TAAGCATGTTAACTTTAGATGCCTTGTATGATATACATAAACTTTGTAGA
AACTGGTGGACATCCACAAACCAAGACTTACCACTAGTAAGATATAAAGG
```

[0511]

ATGCAAAATAACATTTTATCAAAGCACATTTACAGACTACATAGTAAGAA
TACATACAGAACTACCAGCTAACAGTAACAAACTAACATACCCAAACACA
CATCCACTAATGATGATGATGTCTAAGTACAAACACATTATACCTAGTAG
ACAAACAAGAAGAAAAAAGAAACCATACACAAAAATATTTGTAAAACCAC
CTCCGCAATTTGAAAACAAATGGTACTTTGCTACAGACCTCTACAAAATT
CCATTACTACAAATACACTGCACAGCATGCAACTTACAAAACCCATTTGT
AAAACCAGACAAATTATCAAACAATGTTACATTATGGTCACTAAACACCA
TAAGCATACAAAATAGAAACATGTCAGTGGATCAAGGACAATCATGGCCA
TTTAAAATACTAGGAACACAAAGCTTTTATTTTTACTTTTACACCGGAGC
AAACCTACCAGGTGACACAACACAAATACCAGTAGCAGACCTATTACCAC
TAACAAACCCAAGAATAAACAGACCAGGACAATCACTAAATGAGGCAAAA
ATTACAGACCATATTACTTTCACAGAATACAAAAACAAATTTACAAATTA
TTGGGGTAACCCATTTAATAAACACATTCAAGAACACCTAGATATGATAC
TATACTCACTAAAAAGTCCAGAAGCAATAAAAAACGAATGGACAACAGAA
AACATGAAATGGAACCAATTAAACAATGCAGGAACAATGGCATTAACACC
ATTTAACGAGCCAATATTCACACAAATACAATATAACCCAGATAGAGACA
CAGGAGAAGACACTCAATTATACCTACTCTCTAACGCTACAGGAACAGGA
TGGGACCCACCAGGAATTCCAGAATTAATACTAGAAGGATTTCCACTATG
GTTAATATATTGGGGATTTGCAGACTTTCAAAAAAACCTAAAAAAAGTAA
CAAACATAGACACAAATTACATGTTAGTAGCAAAAACAAAATTTACACAA
AAACCTGGCACATTCTACTTAGTAATACTAAATGACACCTTTGTAGAAGG
CAATAGCCCATATGAAAAACAACCTTTACCTGAAGACAACATTAAATGGT
ACCCACAAGTACAATACCAATTAGAAGCACAAAACAAACTACTACAAACT
GGGCCATTTACACCAAACATACAAGGACAACTATCAGACAATATATCAAT
GTTTTATAAATTTTACTTTAAATGGGGAGGAAGCCCACCAAAAGCAATTA
ATGTTGAAAATCCTGCCCACCAGATTCAATATCCCATACCCCGTAACGAG
CATGAAACAACTTCGTTACAGAGTCCAGGGGAAGCCCCAGAATCCATCTT
ATACTCCTTCGACTATAGACACGGGAACTACACAACAACAGCTTTGTCAC
GAATTAGCCAAGACTGGGCACTTAAAGACACTGTTTCTAAAATTACAGAG
CCAGATCGACAGCAACTGCTCAAACAAGCCCTCGAATGCCTGCAAATCTC
GGAAGAAACGCAGGAGAAAAAAGAAAAAGAAGTACAGCAGCTCATCAGCA
ACCTCAGACAGCAGCAGCAGCTGTACAGAGAGCGAATAATATCATTATTA
AAGGACCAATAACTTTTAACTGTGTAAAAAAGGTGAAATTGTTTGATGAT
AAACCAAAAAAACCGTAGATTTACACCTGAGGAATTTGAAACTGAGTTACA
AATAGCAAAATGGTTAAAGAGACCCCCAAGATCCTTTGTAAATGATCCTC
CCTTTTACCCATGGTTACCACCTGAACCTGTTGTAAACTTTAAGCTTAAT
TTTACTGAATAAAGGCCAGCATTAATTCACTTAAGGAGTCTGTTTATTTA

[0512] AGTTAAACCTTAATAAACGGTCACCGCCTCCCTAATACGCAGGCGCAGAA

AGGGGGCTCCGCCCCCTTTAACCCCCAGGGGGCTCCGCCCCCTGAAACCC
CCAAGGGGGCTACGCCCCCTTACACCCCC (서열번호 54)

주석:

추정의 도메인	*염기 범위*
TATA 박스	237– 243
Cap 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
3개의 오픈-리딩 프레임 영역	2441 – 2586
폴리(A) 신호	2808 – 2813
GC-풍부 영역	2868 – 2929

[0513]

[0514] [표 B2] 예시적인 아넬로바이러스 아미노산 서열(베타토르크바이러스)

Ring2 (*베타토르크바이러스*)	
ORF2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEG (서열번호 55)
ORF2/2	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGFNIPYPVTSMKQLRY RVQGKPQNPSYTPSTIDTGTTQQQLCHELAKTGHLKTLFLKLQSQIDSNCSNKPSNA CKSRKKRRRKKKKKYSSSSATSDSSSSCTESE (서열번호 56)
ORF2/3	MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGARSTATAQTSPRMP ANLGRNAGEKRKRSTAAHQQPQTAAAAVQRANNIIIKGPITFNCVKKVKLFDDKPK NRRFTPEEFETELQIAKWLKRPPRSFVNDPPFYPWLPPEPVVNFKLNFTE (서열번호 NO: 57)
ORF1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQPPYKR TCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKL CRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACN LQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNK HIQEHLDMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNP DRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEA QNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNE HETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLK QALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (서열번호 58)
ORF1/1	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRIQYPIPRNEHETTSLQSPGE APESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEE TQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (서열번호 59)
ORF1/2	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRSQIDSNCSNKPSNACKSRK KRRRKKKKKYSSSSATSDSSSSCTESE (서열번호 60)

[0515]

[0516] [표 B3] 예시적인 아넬로바이러스 핵산 서열(감마토르크바이러스)

명칭	Ring3.1
속/클레이드	*감마토르크바이러스*
수탁번호	

전체 서열 : 3264 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAAAATGGCGGCAACCAATCATTTTATACTTTCACTTTCCAATTACAAGC
CGCCACGTCACAGAACAGGGGTGGAGACTTTAAAACTATATAACCAAGTG
ATGTGACGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCGTGGGCGGGAGCCCGAGGTGA
GTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCTAGCGGA
ACGGGCAAGAAACTTAAAATATGTTTTGTTTCAGATGCAGACACCTGCTT
CACAGATAAGCTCAGACGACTTCTTTGTACACACTCCATTTAATGCAGTA
ACTAAACAGCAAATATGGATGTCTCAAATTGCTGATGGACATGACAACAT
TTGTCACTGCCACCGTCCTTTTGCTCACCTGCTTGCTAATATTTTTCCTC
CTGGTCATAAAGACAGGGATCTTACCATTAATCAAATACTTGCTAGAGAT
CTTACAGAAACATGCCATTCTGGTGGAGACGAAGGAACAAGCGGTGGTGG
GGTCGCCGCTTCCGCTACCGCCGCTACAACAAATATAAAACCAGAAGGAG
ACGCAGAATACCCAGAAGACGAAATAGAAGATTTACTAAGACACGCAGGA
GAAGAAAAAGAAAGAAGGTAAGAAGAAAACTTAAAAAAATTACTATTAAA
CAATGGCAGCCAGATTCAGTGAAAAAATGTAAAATTAAAGGATATAGTAC
TTTAGTTATGGGTGCACAAGGAAAACAATACAACTGTTACACAAACCAAG
```

[0517]

CAAGTGACTATGTTCAGCCTAAAGCACCACAAGGTGGGGGCTTTGGCTGT
GAAGTATTTAATTTAAAATGGCTATACCAAGAATATACTGCACACAGAAA
TATTTGGACAAAAACAAATGAATATACAGACCTTTGTAGATACACTGGAG
CTCAAATAATTTTATACAGGCACCCAGATGTTGATTTTATAGTCAGCTGG
GACAATCAGCCACCTTTTTTACTTAACAAATATACATATCCAGAACTGCA
ACCACAAAACCTTTTACTAGCTAGAAGGAAAAGAATTATTCTTAGTCAAA
AATCAAACCCCAAAGGAAAACTAAGAATTAAACTAAGAATACCACCACCA
AAACAAATGATAACAAAATGGTTTTTTCAAAGAGACTTTTGTGATGTGAA
TCTGTTTAAACTATGTGCTTCTGCTGCTTCTTTCCGCTACCCAGGTATCA
GTCATGGAGCTCAAAGTACTATTTTTTCTGCATATGCTTTAAACACTGAC
TTTTATCAATGCAGTGACTGGTGCCAAACTAACACAGAAACTGGCTACCT
AAACATTAAAACACAACAAATGCCACTATGGTTTCATTACAGAGAGGGTG
GCAAAGAGAAATGGTATAAATACACCAACAAAGAACACAGACCATATACA
AATACATATCTTAAAAGTATTAGCTATAATGATGGATTGTTTTCTCCTAA
AGCCATGTTTGCATTTGAAGTAAAAGCGGGGGGTGAAGGAACAACAGAAC
CACCACAAGGCGCCCAATTAATTGCTAACCTTCCACTCATTGCACTAAGA
TATAATCCACATGAAGACACAGGCCATGGCAATGAAATTTACCTTACATC
AACTTTTAAAGGTACATATGACAAACCTAAAGTTACTGATGCTCTATACT
TTAACAATGTACCCCTGTGGATGGGATTTTATGGCTACTGGGACTTTATA
TTACAAGAAACAAAAAACAAAGGTGTCTTTGATCAACATATGTTTGTTGT
TAAATGTCCTGCCTTAAGGCCCATATCACAAGTCACAAAACAAGTATACT
ACCCACTTGTAGACATGGACTTTTGTTCAGGGAGACTGCCATTTGATGAA
TATTTATCCAAAGACATTAAAAGTCATTGGTATCCCACTGCAGAAAGACA
AACAGTTACAATAAATAATTTTGTTACAGCAGGTCCATACATGCCTAAAT
TTGAACCCACAGACAAAGACAGTACATGGCAATTAAACTATCACTATAAA
TTTTTTTTTAAGTGGGGTGGTCCACAAGTCACAGACCCAACTGTTGAAGA
CCCATGCAGCAGAAACAAATATCCTGTCCCCGATACAATGCAACAAACAA
TACAAATTAAAAACCCTGAAAAGCTGCACCCAGCAACCCTCTTCCATGAC
TGGGACCTTAGAAGGGGCTTCATTACACAAGCAGCTATTAAAAGAATGTC
AGAAAACCTCCAAATTGATTCATCTTTCGAATCTGATGGCACAGAATCAC
CCAAAAAAAAGAAAAGATGCACCAAAGAAATCCCAACACAAAACCAAAAG
CAAGAAGAGATCCAAGAATGTCTCCTCTCACTCTGCGAAGAGCCTACATG
CCAAGAAGAAACAGAGGACCTCCAGCTCTTCATCCAGCAGCAGCAGCAGC
AGCAGTACAAGCTCAGAAAAAACCTCTTCAAACTCCTCACTCACCTGAAA
AAAGGACAGAGAATAAGTCAACTACAAACGGGACTTTTAGAGTAATACCA
TTTAAACCAGGTTTTGAACAAGAAACAGAAAAAGAACTTGCCATAGCTTT
CTGCAGACCACCTAGAAAATATAAAAATGATCCCCCTTTTTATCCCTGGT
TACCATGGACACCCCTTGTACACTTTAACCTTAATTACAAAGGCTAGGCC
AACACTGTTCACTTAGTGGTGTATGTTTAATAAAGTTTCACCCCCAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCG
GCGCTCGCGCGCGCTGCGCGCGCGCGAGCGCCGTCACGCGCCGGCGCTCG
CGCGCCGCGCGTATGTGCTAACACACCACGCACCTAGATTGGGGTGCGCG
CGCTAGCGCGCGCACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGC
GCGGGTCGGACCACTTCGGGCTCGGGGGGGCGCGCCTGCGGCGCTTTTTT
ACTAAACAGACTCCGAGCCGCCATTTGGCCCCCCCTAAGCTCCGCCCCCC
TCATGAATATTCATAAAGGAAACCACATAATTAGAATTGCCGACCACAAA
CTGCCATATGCTAATTAGTTCCCCTTTTACACAGTAAAAAGGGGAAGTGG
GGGGGCATAGCCCCCCCACACCCCCCGCGGGGGGGGCAGAGCCCCCCCCC
GCACCCCCCCCCTACGTCACAATCCACGCCCCCGCCGCCATCTTGGGTGC

GGCAGGGCGGGGGC (서열번호 878)

주석:

추정의 도메인	*염기 범위*
TATA 박스	87– 93

[0518]

Cap 부위	110 – 117
전사 시작 부위	117
5′ UTR 보존된 도메인	185 – 255
ORF2	285 – 671
ORF2/2	285 – 667 ; 2063 – 2498
ORF2/3	285 – 667 ; 2295 – 2697
TAIP	385 - 585
ORF1	512 – 2545
ORF1/1	512 – 667 ; 2063 – 2545
ORF1/2	512 – 667 ; 2295 – 2498
3개의 오픈-리딩 프레임 영역	2295 – 2495
폴리(A) 신호	2729 – 2734
GC-풍부 영역	3141 – 3264

[0519]

[0520] **[표 B4] 예시적인 아넬로바이러스 아미노산 서열(감마토르크바이러스)**

Ring 3.1 (*감마토르크바이러스*)	
ORF2	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERR (서열번호 879)
ORF2/2	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERSGVVHKSQTQLLKTHAAETNILSPIQCNKQYKLKTLKSCTQQPS SMTGTLEGASLHKQLLKECQKTSKLIHLSNLMAQNHPKKRKDAPKKSQHK TKSKKRSKNVSSHSAKSLHAKKKQRTSSSSSSSSSSSTSSEKTSSNSSLT (서열번호 880)
ORF2/3	MQTPASQISSDDFFVHTPFNAVTKQQIWMSQIADGHDNICHCHRPFAHLLAN IFPPGHKDRDLTINQ ILARDLTETCHSGGDEGTSGGGVAASATAATTNIKPEGDAEYPEDEIEDLLR HAGEEKERRITQKKEKMHQRNPNTKPKARRDPRMSPLTLRRAYMPRRNRG PPALHPAAAAAAAVQAQKKPLQTPHSPEKRTENKSTTNGTFRVIPFKPGFEQE TEKELAIAFCRPPRKYKNDPPFYPWLPWTPLVHFNLNYKG (서열번호 881)

[0521]

TAIP	MDMTTFVTATVLLLTCLLIFFLLVIKTGILPLIKYLLEILQKHAILVETKEQAV VGSPLPLPPLQQI (서열번호 882)
ORF1	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKVRRKLKKITIKQWQP DSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDYVQPKAPQGGGFGCEVF NLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQIILYRHPDVDFIVSWDNQP PFLLNKYTYPELQPQNLLLARRKRIILSQKSNPKGKLRIKLRIPPPKQMITKWF FQRDFCDVNLFKLCASAASFRYPGISHGAQSTIFSAYALNTDFYQCSDWCQT NTETGYLNIKTQQMPLWFHYREGGKEKWYKYTNKEHRPYTNTYLKSISYN DGLFSPKAMFAFEVKAGGEGTTEPPQGAQLIANLPLIALRYNPHEDTGHGNE IYLTSTFKGTYDKPKVTDALYFNNVPLWMGFYGYWDFILQETKNKGVFDQ HMFVVKCPALRPISQVTKQVYYPLVDMDFCSGRLPFDEYLSKDIKSHWYPT AERQTVTINNFVTAGPYMPKFEPTDKDSTWQLNYHYKFFFKWGGPQVTDP TVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWDLRRGFITQAAIKR MSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQEEIQECLLSLCEEPTCQ EETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKKGQRISQLQTGLLE (서열번호 883)
ORF1/1	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKWGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWD LRRGFITQAAIKRMSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQEEIQE CLLSLCEEPTCQEETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKKGQRISQL QTGLLE (서열번호 884)
ORF1/2	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRK RKKNHPKKRKDAPKKSQHKTKSKKRSKNVSSHSAKSLHAKKKQRTSSSSSS SSSSSSTSSEKTSSNSSLT (서열번호 885)

[0522]

[0523] [표 C1] 예시적인 아넬로바이러스 핵산 서열(감마토르크바이러스)

명칭	Ring4
속/클레이드	*감마토르크바이러스*
수탁번호	

전체 서열 : 3176 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAAAATGGCGGGAGCCAATCATTTTATACTTTCACTTTCCAATTAAAAAT
GGCCACGTCACAAACAAGGGGTGGAGCCATTTAAACTATATAACTAAGTG
GGGTGGCGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCATGGGCGGGAGCCGAGGTGAG
TGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCTAGCGGAA
CGGGCAAGAAACTTAAAACAATATTTGTTTTACAGATGGTTAGTATATCC
TCAAGTGATTTTTTTAAGAAAACGAAATTTAATGAGGAGACGCAGAACCA
AGTATGGATGTCTCAAATTGCTGACTCTCATGATAATATCTGCAGTTGCT
GGCATCCATTTGCTCACCTTCTTGCTTCCATATTTCCTCCTGGCCACAAA
GATCGTGATCTTACTATTAACCAAATTCTTCTAAGAGATTATAAAGAAAA
ATGCCATTCTGGTGGAGAAGAAGGAGAAAATTCTGGACCAACAACAGGTT
TAATTACACCAAAAGAAGAAGATATAGAAAAAGATGGCCCAGAAGGCGCC
GCAGAAGAAGACCATACAGACGCCCTGTTCGCCGCCGCCGTAGAAAACTT
CGAAAGGTAAAGAGAAAAAAAAAATCTTTAATTGTTAGACAATGGCAACC
AGACAGTATAAGAACTTGTAAAATTATAGGACAGTCAGCTATAGTTGTTG
GGGCTGAAGGAAAGCAAATGTACTGTTATACTGTCAATAAGTTAATTAAT
GTGCCCCCAAAAACACCATATGGGGGAGGCTTTGGAGTAGACCAATACAC
ACTGAAATACTTATATGAAGAATACAGATTTGCACAAAACATTTGGACAC
AATCTAATGTACTGAAAGACTTATGCAGATACATAAATGTTAAGCTAATA
TTCTACAGAGACAACAAAACAGACTTTGTCCTTTCCTATGACAGAAACCC
ACCTTTTCAACTAACAAAATTTACATACCCAGGAGCACACCCACAACAAA
TCATGCTTCAAAAACACCACAAATTCATACTATCACAAATGACAAAGCCT
AATGGAAGACTAACAAAAAAACTCAAAATTAAACCTCCTAAACAAATGCT
TTCTAAATGGTTCTTTTCAAAACAATTCTGTAAATACCCTTTACTATCTC
TTAAAGCTTCTGCACTAGACCTTAGGCACTCTTACCTAGGCTGCTGTAAT
GAAAATCCACAGGTATTTTTTTATTATTTAAACCATGGATACTACACAAT
AACAAACTGGGGAGCACAATCCTCAACAGCATACAGACCTAACTCCAAGG
TGACAGACACAACATACTACAGATACAAAAATGACAGAAAAAATATTAAC
ATTAAAAGCCATGAATACGAAAAAAGTATATCATATGAAAACGGTTATTT
TCAATCTAGTTTCTTACAAACACAGTGCATATATACCAGTGAGCGTGGTG
AAGCCTGTATAGCAGAAAAACCACTAGGAATAGCTATTTACAATCCAGTA
AAAGACAATGGAGATGGTAATATGATATACCTTGTAAGCACTCTAGCAAA
CACTTGGGACCAGCCTCCAAAAGACAGTGCTATTTTAATACAAGGAGTAC
CCATATGGCTAGGCTTATTTGGATATTTAGACTACTGTAGACAAATTAAA
GCTGACAAAACATGGCTAGACAGTCATGTACTAGTAATTCAAAGTCCTGC
TATTTTTACTTACCCAAATCCAGGAGCAGGCAAATGGTATTGTCCACTAT
CACAAAGTTTTATAAATGGCAATGGTCCGTTTAATCAACCACCTACACTG
CTACAAAAAGCAAAGTGGTTTCCACAAATACAATACCAACAAGAAATTAT
TAATAGCTTTGTAGAATCAGGACCATTTGTTCCCAAATATGCAAATCAAA
CTGAAAGCAACTGGGAACTAAAATATAAATATGTTTTTACATTTAAGTGG
GGTGGACCACAATTCCATGAACCAGAAATTGCTGACCCTAGCAAACAAGA
GCAGTATGATGTCCCCGATACTTTCTACCAAACAATACAAATTGAAGATC
CAGAAGGACAAGACCCCAGATCTCTCATCCATGATTGGGACTACAGACGA
GGCTTTATTAAAGAAAGATCTCTTAAAAGAATGTCAACTTACTTCTCAAC
TCATACAGATCAGCAAGCAACTTCAGAGGAAGACATTCCCAAAAAGAAAA
```

[0524]

AGAGAATTGGACCCCAACTCACAGTCCCACAACAAAAAGAAGAGGAGACA
CTGTCATGTCTCCTCTCTCTCTGCAAAAAAGATACCTTCCAAGAAACAGA
GACACAAGAAGACCTCCAGCAGCTCATCAAGCAGCAGCAGGAGCAGCAGC
TCCTCCTCAAGAGAAACATCCTCCAGCTCATCCACAAACTAAAAGAGAAT
CAACAAATGCTTCAGCTTCACACAGGCATGTTACCTTAACCAGATTTAAA
CCTGGATTTGAAGAGCAAACAGAGAGAGAATTAGCAATTATATTTCATAG
GCCCCCTAGAACCTACAAAGAGGACCTTCCATTCTATCCCTGGCTACCAC
CTGCACCCCTTGTACAATTTAACCTTAACTTCAAAGGCTAGGCCAACAAT
GTACACTTAGTAAAGCATGTTTATTAAAGCACAACCCCCAAAATAAATGT
AAAAATAAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCGGCGCT
CGCGCGCATGTGCGCCTCTGGCGCAAATCACGCAACGCTCGCGCGCCCGC
GTATGTCTCTTTACCACGCACCTAGATTGGGGTGCGCGCGCTAGCGCGCG
CACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGCGCGGGTCGGACC
ACTTCGGGCTCGGGGGGGCGCGCCTGCGGCGCTTTTTTACTAAACAGACT
CCGAGCCGCCATTTGGCCCCCTAAGCTCCGCCCCCCTCATGAATATTCAT
AAAGGAAACCACATAATTAGAATTGCCGACCACAAACTGCCATATGCTAA
TTAGTTCCCCTTTTACAAAGTAAAAGGGGAAGTGAACATAGCCCCACACC
CGCAGGGGCAAGGCCCCGCACCCCTACGTCACTAACCACGCCCCCGCCGC
CATCTTGGGTGCGGCAGGGCGGGGGC (서열번호 886)

주석:

추정의 도메인	*염기 범위*
TATA 박스	87– 93
Cap 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
3개의 오픈-리딩 프레임 영역	2209 – 2439
폴리(A) 신호	2672 – 2678
GC-풍부 영역	3076 – 3176

[0525]

[0526] [표 C2] 예시적인 아넬로바이러스 아미노산 서열(감마토르크바이러스)

Ring4 (*감마토르크바이러스*)	
ORF2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLR DYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALFAAAV ENFER (서열번호 887)
ORF2/2	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGA AEEDHTDALFAAAVENFESGVDHNSMNQKLLTLANKSSMMSPILSTKQYKL KIQKDKTPDLSSMIGTTDEALLKKDLLKECQLTSQLIQISKQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSSR ETSSSSSTN (서열번호 888)
ORF2/3	MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPP GHKDRDLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGA AEEDHTDALFAAAVENFERSASNFRGRHSQKEKENWTPTHSPTTKRRGDTV MSPLSLQKRYLPRNRDTRRPPAAHQAAAGAAAPPQEKHPPAHPQTKRESTN ASASHRHVTLTRFKPGFEEQTERELAIIFHRPPRTYKEDLPFYPWLPPAPLVQ FNLNFKG (서열번호 889)
TAIP	MRRRRTKYGCLKLLTLMIISAVAGIHLLTFLLPYFLLATKIVILLLTKFF (서열번호 890)

[0527]

ORF1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINV PPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFY RDNKTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGR LTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNENPQ VFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHE YEKSISYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPVKDNGDGNM IYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCRQIKADKTWLDSHV LVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTLLQKAKWFPQIQYQ QEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKWGGPQFHEPEIADPSK QEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSLKRMSTYFST HTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETLSCLLSLCKKDTFQETETQE DLQQLIKQQQEQQLLLKRNILQLIHKLKENQQMLQLHTGMLP (서열번호 891)
ORF1/1	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKWGGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDY RRGFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETL SCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKENQQM LQLHTGMLP (서열번호 892)
ORF1/2	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRK LRKISKQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSSR ETSSSSTN (서열번호 893)

[0528]

[0529] **[표 E1] 예시적인 아넬로바이러스 핵산 서열(알파토르크바이러스) - 클레이드 1**

명칭 Ring5.2

속/클레이드 *알파토르크바이러스* 클레이드 1

수탁번호

전체 서열 : 3696 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
ATTTTGTTCAGCCCGCCAATTTCTCTTTCAAACAGGCCAATCAGCTACTA
CTTCGTGCACTTCCTGGGGCGTGTCCTGCCGCTCTATATAAGCAGAGGCG
GTGACGAATGGTAGAGTTTTTCTTGGCCCGTCCGCGGCGAGAGCGCGAGC
GAAGCGAGCGATCGAGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACA
CACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTATGGGCA
AGATTCTTAAAAAATTCCCCCGATCCCTTTGCCGCCAGGACATAAAAACA
TGCCGTGGAGACCGCCGGTCCATAGTGTCCAGGGGCGAGAGGATCAGTGG
TTCGCAAGCTTTTTTCACGGCCACGATTCGTTTTGCGGCTGCGGTGACCC
TCTTGGCCATATTAATAGCATTGCTCATCGCTTTCCTCGCGCCGGTCCAC
CAAGGCCCCCTCCGGGGGCTAGATCAGCCTAACCCCCGGGAGCAGGGCCCG
GCCGGACCCGGAGGGCCGCCCGCCATCTTGGCCCTGCCGGCTCCGCCCGC
GGAGCCTGACGACCCGCAGCCACGGCGTGGTGGTGGGGACGGTGGCGCCG
CCGCTGGCGCCGCAGACGACCATACACAACGAGACTACGACGAAGAAGAG
CTAGACGAGCTTTTCCGCGCCGCCGCCGAAGACGATTTGTAAGTAGGAGA
TGGCGCCGGCCTTACAGGCGCAGGAGGAGACGCGGGCGACGCAGACGCAG
ACGCAGACGCAGACATAAGCCCACCCTAATACTCAGACAGTGGCAACCTG
ACTGTATCAGACACTGTAAAATAACAGGATGGATGCCCCTCATTATCTGT
GGAAAGGGGTCCACCCAGTTCAACTACATCACCCACGCGGACGATATCAC
CCCCAGGGGAGCCTCCTACGGAGGCAATTTCACAAACATGACTTTCTCCC
```

[0530]

TGGAGGCCATATATGAACAGTTCCTATACCACAGAAACAGGTGGTCGGCC
TCTAACCACGACCTAGAACTGTGCAGATACAAGGGGACCACCTTAAAACT
CTACAGACACCCAGAAGTAGACTACATAGTTACCTACAGCAGAACAGGAC
CCTTTGAAATCAGCCACATGACCTACCTCAGCACTCACCCCATGCTAATG
CTGCTAAACAAGCACCACATTGTGGTGCCCAGCTTAAAGACTAAGCCCAG
AGGCAGAAAGGCCATAAAAGTCAGGATAAGGCCCCCAAAACTCATGAACA
ACAAGTGGTACTTCACCAGAGACTTCTGTAACATAGGCCTCTTCCAGCTC
TGGGCCACAGGCTTAGAACTCAGAAACCCCTGGCTCAGAATGAGCACCCT
GAGCCCCTGCATAGGCTTTAATGTCCTCAAAAACAGCATTTACACAAACC
TCAGCAACCTGCCACAATACAAAAACGAAAGACTAAACATCATTAACAAC
ATACTTCACCCACAAGAAATTACAGGTACAAACAACAAAAAGTGGCAGTA
CACATACACCAAACTCATGGCCCCTATTTACTATTCAGCAAACAGGGCCA
GCACCTATGACTGGGAAAATTACAGCAAAGAAACAAACTACAATAATACA
TATGTTAAATTTACCCAGAAAAGACAGGAAAAACTAACTAAAATTAGAAA
AGAGTGGCAGATGCTTTATCCACAACAACCCACAGCACTGCCAGACTCCT
ATGACCTCCTACAAGAGTATGGCCTCTACAGTCCATACTACCTAAACCCC
ACAAGAATAAACCTAGACTGGATGACCCCATACACACACGTCAGATACAA
TCCCCTAGTAGACAAGGGCTTTGGAAACAGAATATACATCCAGTGGTGCT
CAGAAGCAGATGTTAGCTACAACAGGACAAAATCCAAGTGTCTGCTACAA
GACATGCCCCTGTTTTTCATGTGCTATGGCTACATAGACTGGGCAATAAA
AAACACTGGAGTGTCATCTCTAGTGAAGGACGCCAGAATCTGCATCAGGT
GTCCCTACACAGAGCCACAACTAGTTGGCTCCACAGAAGACATAGGCTTT
GTACCCATCTCAGAAACCTTCATGAGGGGCGACATGCCGGTACTTGCACC
ATACATACCGTTAAGCTGGTTTTGCAAGTGGTATCCCAACATAGCTCACC
AAAAGGAAGTCCTTGAGTCAATCATTTCCTGCAGCCCCTTCATGCCCCGT
GACCAAGACATGAACGGTTGGGATATCACAATCGGTTACAAAATGGACTT
CTTATGGGGCGGTTCCCCTCTCCCCTCACAGCCAATCGACGACCCCTGCC
AGCAGGGAACCCACCCGATTCCCGACCCCGATAAACACCCTCGCCTCCTA
CAAGTCTCGAACCCGAAACTACTCGGACCGAGGACAGTGTTCCACAAGTG
GGACATCAGACGTGGGCAGTTTAGCAAAAGAAGTATTAAGAGAGTGTCAG
AATACTCAAGCGATGATGAATCTCTTGCGCCAGGTCTCCCATCAAAGCGA
AACAAGCTCGACTCGGCGTTCCGAGGAGAAAATCGAGAGCAAAAAGAATG
CTATTCTCTCCTCAAAGCGCTCGAGGAAGAAGAGACCCCAGAAGAAGAAG
AACCAGCACCCCAAGAAAAAGCCCAGAAAGAGGAGCTACTCCACCAGCTC
CAGCTCCAGAGACGCCACCAGCGAGTCCTCAGACGAGGGCTCAAGCTCGT
CTTTACAGACATCCTCCGACTCCGCCAGGGAGTCCACTGGAACCCGGAGC
TCACATAGCGCCCCCACCTTACATACCAGACCTGCTTTTTCCCAATACTG
GTAAAAAAAAAAAATTCTCTCCCTTCGATTGGGAGACAGAGGCGCAAATA
GCGGGGTGGATGCGGCGGCCCATGCGCTTCTATCCCTCAGACACCCCTCA
CTACCCGTGGCTACCCCCCGAGCGAGATATCCCGAAAATATGTAACATAA
ACTTCAAAATAAAGCTTCAAGAGTGAGTGATTCGAGGCCCTCCTCTGTTC
ACTTAGCGGTGTCTACCTCTTAAGGTCACTAAGCACTCCGAGCGTAAGCG
AGGAGTGCGACCCTCTACCAAGGGGCAACTTCCTCGGGGTCCGGCGCTAC
GCGCTTCGCGCTGCGCCGGACATCTCGGACCCCTCGACCCGAATCGCTTG
CGCGATTCGGACCTGCGGCCTCGGGGGGGTCGGGGGCTTTACTAAACAGA
CTCCGAGGTGCCATTGGACACTGTAGGGGGTGAACAGCAACGAAAGTGAG
TGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTGGATAGTG
ACTTCCGGGTCCGCCTGGGGGCCGCCATTTTAGCTTCGGCCGCCATTTTA
GGCCCTCGCGGGCCTCCGTAGGCGCGCTTTAGTGACGTCACGGCAGCCAT
TTTGTCGTGACGTTTGAGACACGTGATGGGGGCGTGCCTAAACCCGGAAG
CATCCCTGGTCACGTGACTCTGACGTCACGGCGGCCATCTTGTGCTGTCC
GCCATCTTGTAACTTCCTTCCGCTTTTTCAAAAAAAAAGAGGAAGTGTGA
CGTAGCGGCGGGGGGGCGGCGCGCTTCGCGCGCCGCCCACCAGGGGGCGC
TGCGCGCCCCCCGCGCATGCGCAGGGGCCTCTCGAGGGGCTCCGCCCCCC

[0531] CCCCGTGCTAAATTTACCGCGCATGCGCGACCACGCCCCCGCCGCC (서열번호 894)

주석:

추정의 도메인	*염기 범위*
TATA 박스	85– 91
Cap 부위	108 – 115
전사 시작 부위	115
5′ UTR 보존된 도메인	178 – 248
ORF2	300 – 692
ORF2/2	300 – 688 ; 2282 – 2804
ORF2/3	300 – 688 ; 2484 – 2976
ORF2t/3	300 – 349 : 2484 - 2976
TAIP	322 - 471
ORF1	572 – 2758
ORF1/1	572 – 688 ; 2282 – 2758
ORF1/2	572 – 688 ; 2484 – 2804
3개의 오픈-리딩 프레임 영역	2484 – 2755
폴리(A) 신호	3018 –3023
GC-풍부 영역	3555 – 3696

[0532]

[0533] **[표 D2] 예시적인 아넬로바이러스 아미노산 서열(알파토르크바이러스) 클레이드 1**

Ring 5.2 (*알파토르크바이러스*) 클레이드 1	
ORF2	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG AADDHTQRDYDEEELDELFRAAAEDDL (서열번호 895)
ORF2/2	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG AADDHTQRDYDEEELDELFRAAAEDDFQSTTPASREPTRFPTPINTLASYKS RTRNYSDRGQCSTSGTSDVGSLAKEVLRECQNTQAMMNLLRQVSHQSETSS TRRSEEKIESKKNAILSSKRSRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDAT SESSDEGSSSSLQTSSDSARESTGTRSSHSAPTLHTRPAFSQYW (서열번호 896)
ORF2/3	MPWRPPVHSVQGREDQWFASFFHGHDSFCGCGDPLGHINSIAHRFPRAGPP RPPPGLDQPNPREQGPAGPGGPPAILALPAPPAEPDDPQPRRGGGDGGAAAG AADDHTQRDYDEEELDELFRAAAEDDLSPIKAKQARLGVPRRKSRAKRMLF SPQSARGRRDPRRRRTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHP PTPPGSPLEPGAHIAPPPYIPDLLFPNTGKKKKFSPFDWETEAQIAGWMRRPM RFYPSDTPHYPWLPPERDIPKICNINFKIKLQ (서열번호 897)

[0534]

ORF2t/3	MPWRPPVHSVQGREDQWSPIKAKQARLGVPRRKSRAKRMLFSPQSARGRR DPRRRRTSTPRKSPERGATPPAPAPETPPASPQTRAQARLYRHPPTPPGSPLEP GAHIAPPPYIPDLLFPNTGKKKKFSPFDWETEAQIAGWMRRPMRFYPSDTPH YPWLPPERDIPKICNINFKIKLQE (서열번호 898)
TAIP	IVSRGERISGSQAFFTATIRFAAAVTLLAILIALLIAFLAPVHQGPLRG (서열번호 899)
ORF1	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRRPYR RRRRRGRRRRRRRRRHKPTLILRQWQPDCIRHCKITGWMPLIICGKGSTQFN YITHADDITPRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRY KGTTLKLYRHPEVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHIVVPSL KTKPRGRKAIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLELRNPWLR MSTLSPCIGFNVLKNSIYTNLSNLPQYKNERLNIINNILHPQEITGTNNKKWQ YTYTKLMAPIYYSANRASTYDWENYSKETNYNNTYVKFTQKRQEKLTKIR KEWQMLYPQQPTALPDSYDLLQEYGLYSPYYLNPTRINLDWMTPYTHVRY NPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLLQDMPLFFMCYGYIDWAIK NTGVSSLVKDARICIRCPYTEPQLVGSTEDIGFVPISETFMRGDMPVLAPYIPL SWFCKWYPNIAHQKEVLESIISCSPFMPRDQDMNGWDITIGYKMDFLWGGS PLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKWDIRRGQFS KRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENREQKECYSLLKALEEE ETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGLKLVFTDILRLRQGVH WNPELT (서열번호 900)
ORF1/1	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFPIDDPCQQGT HPIPDPDKHPRLLQVSNPKLLGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDE SLAPGLPSKRNKLDSAFRGENREQKECYSLLKALEEEETPEEEEPAPQEKAQ KEELLHQLQLQRRHQRVLRRGLKLVFTDILRLRQGVHWNPELT (서열번호 901)
ORF1/2	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSHQSETSSTR RSEEKIESKKNAILSSK RSRKKRPQKKKNQHPKKKPRKRSYSTSSSSRDATSESSDEGSSSSLQTSSDSA RESTGTRSSHSAPTLHTRPAFSQYW (서열번호 902)

[0535]

[0536] [표 F1] 예시적인 아넬로바이러스 핵산 서열(베타토르크바이러스)

명칭	Ring9
속/클레이드	*베타토르크바이러스*
수탁번호	MH649263.1

전체 서열 : 2845 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TTATTAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTAACTTCCTTATTTAACATTACTTCCCTTTTAACCAATGAATA
TTCATACAACACATCACACTTCCTGGGAGGAGACATAAAACTATATAACT
AACTACACAGACGAATGGCTGAGTTTATGCCGCTAGACGGAGGACGCACA
GCTACTGCTGCGACCTGAACTTGGGCGGGTGCCGAAGGTGAGTGTAACCA
CCGTAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAG
ATTATTAATACAAACTTATTTTTACAGATGAGCAAACAACTAAAACCAAC
TTTATACAAAGACAAATCATTGGAATTACAATGGCTAAACAACATTTTTA
GCTCTCACGACCTGTGCTGCGGCTGCAACGATCCAGTTTTACATTTACTG
ATTTTAATTAACAAAACCGGAGAAGCACCTAAACCAGAAGAAGACATTAA
AAATATAAAATGCCTCCTTACTGGCGCCAAAAATACTACCGAAGAAGATA
TAGACCTTTCTCCTGGAGAACTAGAAGAATTATTCAAAGAAGAAAAAGAT
GGAGATACCGCAAACCAAGAAAAACATACTGGAGAAGAAAACTGCGGGTA
AGAAAACGTTTTTATAAAAGAAAGTTAAAAAAAATTGTACTTAAACAGTT
TCAACCAAAAATTATTAGAAGATGTACAATATTTGGAACAATCTGCCTAT
TTCAAGGCTCTCCAGAAAGAGCCAACAATAATTATATTCAAACAATCTAC
TCCTACGTACCAGATAAAGAACCAGGAGGAGGGGGATGGACTTTAATAAC
TGAAAGCTTAAGTAGTTTATGGGAAGACTGGGAACATTTAAAAAATGTAT
GGACTCAAAGTAACGCTGGTTTACCACTTGTAAGATACGGGGGAGTAACA
TTATACTTTTATCAATCTGCCTATACTGACTATATTGCTCAAGTTTTCAA
CTGTTATCCTATGACAGACACAAAATACACACATGCAGACTCAGCACCAA
ACAGAATGTTATTAAAAAAACATGTAATAAGAGTACCTAGCAGAGAAACA
CGCAAAAAAAAGAAAGCCATACAAAAGAGTTAGAGTAGGACCTCCTTCTCA
AATGCAAAACAAATGGTACTTTCAAAGAGACATATGTGAAATACCATTAA
TAATGATTGCAGCCACAGCCGTTGACTTTAGATATCCCTTTTGTGCAAGC
GACTGTGCTAGTAACAACTTAACTCTAACATGTTTAAACCCACTATTGTT
TCAAAACCAAGACTTTGACCACCCATCCGATACACAAGGCTACTTTCCAA
AACCTGGAGTATATCTATACTCAACACAAAGAAGTAACAAGCCAAGTTCT
TCAGACTGTATATACTTAGGAAACACAAAAGACAATCAAGAAGGTAAATC
TGCAAGTAGTCTAATGACTCTAAAAACACAAAAAATAACAGATTGGGGAA
ATCCATTTTGGCATTATTATATAGACGGTTCTAAAAAAATATTTTCTTAC
TTTAAACCCCCATCACAATTAGACAGCAGCGACTTTGAACACATGACAGA
ATTAGCAGAACCAATGTTTATACAAGTTAGATACAACCCAGAAAGAGACA
CAGGACAAGGAAACTTAATATACGTAACAGAAAACTTTAGAGGACAACAC
```

[0537]

TGGGACCCTCCATCTAGTGACAACCTAAAATTAGATGGATTTCCCTTATA
TGACATGTGCTGGGGTTTCATAGACTGGATAGAAAAAGTTCATGAAACAG
AAAACTTACTTACCAACTACTGCTTCTGTATTAGAAGCAGCGCTTTCAAT
GAAAAAAAAACAGTTTTTATACCTGTAGATCATTCATTTTTAACAGGTTT
TAGCCCATATGAAACTCCAGTTAAATCATCAGACCAAGCTCACTGGCACC
CACAAATAAGATTTCAAACAAAATCAATAAATGACATTTGTTTAACAGGC
CCCGGTTGTGCTAGGTCCCCATATGGCAATTACATGCAGGCAAAAATGAG
TTATAAATTTCATGTAAAATGGGGAGGATGTCCAAAAACTTATGAAAAAC
CATATGATCCTTGTTCACAGCCCAATTGGACTATTCCCCATAACCTCAAT
GAAACAATACAAATCCAGAATCCAAACACATGCCCACAAACAGAACTCCA
AGAATGGGACTGGCGACGTGATATTGTTACAAAAAAAGCTATCGAAAGAA
TTAGACAACACACGGAACCTCATGAAACTTTGCAAATCTCTACAGGTTCC
AAACACAACCCACCAGTACACAGACAAACATCACCGTGGACGGACTCAGA
AACGGACTCGGAAGAGGAAAAAGACCAAACACAAGAGATCCAGATCCAGC
TCAACAAGCTCAGAAAGCATCAACAGCATCTCAAGCAGCAGCTCAAGCAG
TACCTGAAACCCCAAAATATAGAATAGTTGCAAGCAACATAAAAGTTGAA
CTTTTTCCTACTAAAAAACCTTTTAAAAACAGACGCTTTACTCCTTCTGA
AAGAGAAACAGAAAGACAATGTGCTAAAGCTTTTTGTAGACCAGAAAGAC
ATTTCTTTTATGATCCTCCTTTTTACCCTTACTGTGTACCTGAACCTATT
GTAAACTTTGCTTTGGGATATAAAATTTAAGGCCAACAAATTTCACTTAG
TGGTGTCTGTTTATTAAAGTTTAACCTTAATAAGCATACTCCGCCTCCCT
ACATTAAGGCGCCAAAAGGGGGCTCCGCCCCCTTAAACCCCAAGGGGGCT
CCGCCCCCTTAAACCCCCAAGGGGGCTCCGCCCCCTTACACCCCC （서열번호 1001）

주석:

추정의 도메인	*염기 범위*
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
3개의 오픈-리딩 프레임 영역	2296 – 2454
GC-풍부 영역	2734 – 2845

[0538]

[0539] [표 F2] 예시적인 아넬로바이러스 아미노산 서열(베타토르크바이러스)

Ring9 (*베타토르크바이러스*)	
ORF2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CG (서열번호 1002)
ORF2/2	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGPIGLFPITSMKQYKSRIQTHAHKQNSKNGTGDVILLQKKLSKELDNTRNL MKLCKSLQVPNTTHQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESIN SISSSSSSST (서열번호 1003)
ORF2/3	MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGFQTQPTSTQTNITVDGLRNGLGRGKRPNTRDPDPAQQAQKASTASQAAA QAVPETPKYRIVASNIKVELFPTKKPFKNRRFTPSERETERQCAKAFCRPERH FFYDPPFYPYCVPEPIVNFALGYKI (서열번호 1004)
ORF1	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVRKR FYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKE PGGGGWTLITESLSSLWEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSA YTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPSRETRKKRKPYK RVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLT CLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDN QEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKPPSQLDSSDFEH MTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFP LYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFS PYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNYMQAKMSYK FHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWD WRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEE KDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE (서열번호 1005)
ORF1/1	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRPNW TIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIERIRQHTEPHETLQIS TGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQL KQYLKPQNIE (서열번호 1006)
ORF1/2	MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVPNT THQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESINSISSSSSSST (서열번호 1007)

[0540]

[0541] [표 F3] 예시적인 아넬로바이러스 핵산 서열(베타토르크바이러스)

명칭	Ring10
속/클레이드	*베타토르크바이러스*
수탁번호	JX134044.1

전체 서열 : 2912 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCCTCTTTTTCCACAACTTCCTCTTTTACTAATGAATA
TTCATGTAATTAATTAATAATCACCGTAATTCCGGGGAGGAGCCTTTAAA
CTATAAAACTAACTACACATTCGAATGGCTGAGTTTATGCCGCCAGACGG
AGACGGGATCACTTCAGTGACTCCAGGCTGATCAAGGGCGGGTGCCGAAG
GTGAGTGAAACCACCGTAGTCAAGGGGCAATTCGGGCTAGATCAGTCTGG
CGGAACGGGCAAGAAACTTAAAATGTACTTTATTTTACAGAAATGTTCAA
ATCTCCAACATACTTAACAACTAAAGGCAAAAACAATGCCTTAATCAACT
GCTTCGTTGGAGACCACGATCTTCTGTGCAGCTGTAACAATCCTGCCTAC
CATTGCCTCCAAATACTTGCAACTACCTTAGCACCTCAACTAAAACAAGA
AGAAAAACAACAAATAATACAATGCCTTGGTGGTACAGACGCCGTAGCTA
CAACCCGTGGAGACGAAGAAATTGGTTTAGAAGACCTAGAAAAACTATTT
ACAGAAGATACAGAAGAAGACGCCGCTGGGTAAGAAGAAAACCTTTTTAC
AAACGTAAAATTAAGAGACTAAATATAGTAGAATGGCAACCTAAATCAAT
TAGAAAATGTAGAATAAAAGGAATGCTATGCTTGTTTCAAACGACAGAAG
ACAGACTGTCATATAACTTTGATATGTATGAAGAGTCTATTATACCAGAA
AAACTGCCGGGAGGGGGGGGATTTAGCATTAAGAATATAAGCTTATATGC
CTTATACCAAGAACACATACATGCACACAACATATTTACACACACAAACA
CAGACAGACCACTAGCAAGATACACAGGCTGTTCTTTAAAATTCTACCAA
AGCAAAGACATAGACTACGTAGTAACATATTCTACATCACTCCCACTAAG
AAGCTCAATGGGAATGTACAACTCCATGCAACCATCCATACATCTAATGC
AACAAAACAAACTAATTGTACCAAGCAAACAAACACAAAAAAGAAGAAAA
CCATATATTAAAAAACATATATCACCACCAACACAAATGAAATCTCAATG
GTACTTTCAACATAACATTGCAAACATACCGCTACTAATGATAAGAACCA
CAGCATTAACATTAGATAATTACTATATAGGAAGCAGACAATTAAGTACA
AATGTCACTATACATACACTTAACACAACATACATCCAAAACAGAGACTG
GGGAGACAGAAATAAAACTTACTACTGCCAAACATTAGGAACACAAAGAT
ACTTCCTATATGGAACACATTCAACTGCACAAAATATTAATGACATAAAG
CTACAAGAACTAATACCTTTAACAAACACACAAGACTATGTACAAGGCTT
TGATTGGACAGAAAAAGACAAACATAACATAACAACCTACAAAGAATTCT
TAACTAAAGGAGCAGGAAATCCATTTCACGCAGAATGGATAACAGCACAA
AACCCAGTAATACACACAGCAAACAGTCCTACACAAATAGAACAAATATA
CACCGCTTCAACAACAACATTCCAAAACAAAAAACTAACAGACCTACCAA
CGCCAGGATATATATTTATAACTCCAACAGTAAGCTTAAGATACAACCCA
TACAAAGACCTAGCAGAAAGAAACAAATGCTACTTTGTAAGAAGCAAAAT
AAATGCACACGGGTGGGACCCAGAACAACACCAAGAATTAATAAACAGTG
```

[0542]

ACCTACCACAATGGTTACTATTATTTGGCTACCCAGACTACATAAAAAGA
ACACAAAACTTTGCATTAGTAGACACAAATTACATACTAGTAGACCACTG
CCCATACACAAATCCAGAAAAAACACCATTTATACCTTTAAGCACATCAT
TTATAGAAGGTAGAAGCCCATACAGTCCTTCAGACACACATGAACCAGAT
GAAGAAGACCAAAACAGGTGGTACCCATGCTACCAATATCAACAAGAATC
AATAAATTCAATATGTCTTAGCGGTCCAGGCACACCAAAAATACCAAAAG
GAATAACAGCAGAAGCAAAAGTAAAATATTCCTTTAATTTTAAGTGGGGT
GGTGACCTACCACCAATGTCTACAATTACAAACCCGACAGACCAGCCAAC
ATATGTTGTTCCCAATAACTTCAATGAAACAACTTCGTTACAGAATCCAA
CCACCAGACCAGAGCACTTCTTGTACTCCTTTGACGAAAGGAGGGGACAA
CTTACAGAAAAAGCTACAAAACGCTTGCTTAAAGACTGGGAAACTAAAGA
AACTTCTTTATTGTCTACAGAATACAGATTCGCGGAGCCAACACAAACAC
AAGCCCCACAAGAGGACCCGTCCTCGGAAGAAGAAGAAGAGAGCAACCTC
TTCGAGCGACTCCTCCGACAGCGAACCAAGCAGCTCCAGCTCAAGCGCAG
AATAATACAAACATTGAAAGACCTACAAAAATTAGAATAACTAACAGCAA
AAACACCGTTTACCTATTTCCACCTGAACAAAAGAACAGAAGACTAACAC
CATGGGAAATACAAGAAGACAAAGAAATAGCCAATTTATTTGGCAGACCA
CATAGATACTTTTTAAAAGACATTCCTTTCTATTGGGATATACCCCCAGA
GCCTAAAGTAAACTTTGATTTAAATTTTCAATAAAGAAATAAAGGGCAAG
GCCCCATTAACTCAAAGTCGGTGTCTACCTCTTTAAGTTTAACTTTACTA
AACGGACTCCGCCTCCCTAAATTTGGGCGCCAAAAGGGGGCTCCGCCCCC
TTAAACCCCAGGGGGCTCCGCCCCCTAAAACCCCCAAGGGGGCTACGCCC
CCTTACACCCCC (서열번호 1008)

주석:

추정의 도메인	*염기 범위*
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
3개의 오픈-리딩 프레임 영역	2276 – 2502
GC-풍부 영역	2803 – 2912

[0543]

[0544] [표 F4] 예시적인 아넬로바이러스 아미노산 서열(베타토르크바이러스)

Ring10 (*베타토르크바이러스*)	
ORF2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAG (서열번호 NO: 1009)
ORF2/2	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPI TSMKQLRYRIQPPDQSTSCTPLTKGGDNLQKKLQNACLKTGKLKKLLYCLQ NTDSRSQHKHKPHKRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (서열번호 1010)
ORF2/3	MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGIQIRGANT NTSPTRGPVLGRRRREQPLRATPPTANQAAPAQAQNNTNIERPTKIRITNSKN TVYLFPPEQKNRRLTPWEIQEDKEIANLFGRPHRYFLKDIPFYWDIPPEPKVN FDLNFQ (서열번호 1011)
ORF1	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWVRRKPFYKRKIKR LNIVEWQPKSIRKCRIKGMLCLFQTTEDRLSYNFDMYEESIIPEKLPGGGGFSI KNISLYALYQEHIHAHNIFTHTNTDRPLARYTGCSLKFYQSKDIDYVVTYSTS LPLRSSMGMYNSMQPSIHLMQQNKLIVPSKQTQKRRKPYIKKHISPPTQMKS QWYFQHNIANIPLLMIRTTALTLDNYYIGSRQLSTNVTIHTLNTTYIQNRDW GDRNKTYYCQTLGTQRYFLYGTHSTAQNINDIKLQELIPLTNTQDYVQGFD WTEKDKHNITTYKEFLTKGAGNPFHAEWITAQNPVIHTANSPTQIEQIYTAS TTTFQNKKLTDLPTPGYIFITPTVSLRYNPYKDLAERNKCYFVRSKINAHGW DPEQHQELINSDLPQWLLLFGYPDYIKRTQNFALVDTNYILVDHCPYTNPEK TPFIPLSTSFIEGRSPYSPSDTHEPDEEDQNRWYPCYQYQQESINSICLSGPGTP KIPKGITAEAKVKYSFNFKWGGDLPPMSTITNPTDQPTYVVPNNFNETTSLQ NPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEPTQT QAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE (서열번호 1012)
ORF1/1	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWPTYVVPNNFNETT SLQNPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEP TQTQAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE
ORF1/2	MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWNTDSRSQHKHKPH KRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (서열번호 1013)

[0545]

[0546] [표 F5] 예시적인 아넬로바이러스 핵산 서열(알파토르크바이러스, 클레이드 4)

명칭	Ring20
속/클레이드	*알파토르크바이러스* 클레이드 4
수탁번호	AF122914.3

전체 서열 : 3853 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
GGCTTAGTGCGTCACCACCCACGTGACCCGCCTCCGCCAATTAACAGGTA
CTTCGTACACTTCCTGGGCGGGCTTATAAGACTAATATAAGTAGCTGCAC
TTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCAGCGGTGAAGCCACGGA
GGGAGCTCAGCGCGTCCCGAGGGCGGGTGCCGGAGGTGAGTTTACACACC
GCAGTCAAGGGGCAATTCGGGCTCGGGACTGGCCGGGCTTTGGGCAAGGC
TCTTAAAAAAAGCTATGTTTATTGGCAGGCACTACCGAAAGAAAAGGGCGC
TGCTACTGCTATCTGTGCATTCTACAAAGACAAAAGGGAAACTTCTAATA
GCTATGTGGACTCCCCCACGCAATGATCAACAATACCTTAACTGGCAATG
GTACACTTCTGTACTTAGCTCCCACTCTGCTATGTGCGGGTGTTCCGACG
CTATCGCTCATCTTAATCATCTTGCTAATCTGCTTCGTGCCCCGCAAAAT
CCGCCCCCGCCTGATAATCCAAGACCCCTACCCGTGCGAGCACTGCCTGC
TCCCCCGGCTGCCCACGAGGCAGCCGGTGATCGAGCACCATGGCCTATGG
GTGGTGGAGGAGACGCCGGAGGCGCTGGCGCAGGTGGAGACGCCGACCAT
GGAGGCGCCGCTGGAGGACCCGCAGACGCAGACCTGCTAGACGCCGTGGC
CGCCGCAGAAACGTAAGGAGACGGCGCAGAGGGAGGTGGAGAAGGAGGTA
CAGGAGGTGGAAAAGAAAGGGCAGACGTAGAAGAAAAGCAAAAATAATAA
TAAGACAGTGGCAGCCAAACTACAGAAGAAGATGTAATATAGTGGGCTAC
CTCCCTATACTTATCTGTGGTGGAAATACTGTTTCTAGAAACTATGCCAC
ACACTCAGACGATACTAACTATCCAGGACCCTTTGGGGGAGGCATGACCA
CAGACAAATTCAGCCTTAGAATACTATATGATGAATACAAAAGATTTATG
AACTACTGGACAGCCTCAAATGAGGACCTAGATCTCTGTAGATATCTAGG
ATGCACTTTTTACTTCTTTAGACACCCTGAAGTAGACTTTATTATAAAAA
TAAACACCATGCCCCCATTCTTAGATACAACCATAACAGCACCTAGCATA
CACCCAGGCCTCATGGCCCTAGACAAAAGAGCCAGATGGATTCCTTCTCT
TAAAAATAGACCAGGTAAAAAACACTATATAAAAATTAGAGTAGGGGCTC
CTAAAATGTTCACAGATAAATGGTACCCTCAAACAGACCTCTGTGACATG
ACACTGCTAACTATCTATGCAACCGCAGCGGATATGCAATATCCGTTCGG
CTCACCACTAACTGACACTGTGGTTGTTAACTCCCAAGTTCTGCAATCCA
TGTATGATGAAACAATTAGCATATTACCTGATGAAAAAACTAAAAGAAAT
AGCCTTCTTACTTCTATAAGAAGCTACATACCTTTTTATAATACTACACA
AACAATAGCTCAATTAAAACCATTTGTAGATGCAGGAGGACACACAACAG
GCTCAACAACAACTACATGGGGACAACTATTAAACACAACTAAATTTACC
ACTACCACAACAACCACATACACATACCCTGGCACCACAAATACAGCAGT
AACATTTATAACAGCCAATGATACCTGGTACAGGGGAACAGCATATAAAG
ATAACATTAAAGATGTACCACAAAAAGCAGCACAATTATACTTTCAAACA
ACACAAAAACTACTAGGAAACACATTCCATGGCTCAGATGAAACACTTGA
```

[0547]

ATACCATGCAGGCCTATACAGCTCTATCTGGCTATCACCAGGTAGATCCT
ACTTTGAAACACCAGGTGCATACACAGACATTAAATATAACCCTTTTACA
GACAGAGGAGAAGGCAACATGCTGTGGATAGACTGGCTAAGTAAAAAAAA
CATGAAATATGACAAAGTGCAAAGTAAGTGCCTAGTAGCAGACCTACCAC
TGTGGGCAGCAGCATATGGTTATGTAGAATTCTGCTCTAAAAGCACAGGA
GACACAAACATACACATGAATGCCAGACTACTAATAAGAAGTCCTTTTAC
AGACCCCCAGCTAATAGTACACACAGACCCCACTAAAGGCTTTGTACCCT
ATTCTTTAAACTTTGGAAATGGTAAAATGCCAGGAGGTAGCAGCAATGTT
CCCATAAGAATGAGAGCTAAGTGGTACCCCACTTTATCCCACCAACAAGA
AGTTCTAGAGGCCTTAGCACAGTCAGGACCCTTTGCTTATCACTCAGACA
TTAAAAAAGTATCTCTAGGCATAAAATACCGTTTTAAGTGGATCTGGGGT
GGAAACCCCGTTCGCCAACAGGTTGTTAGAAATCCCTGCAAGGAACCCCA
CTCCTCGGGCAATAGAGTCCCTAGAAGCATACAAATCGTTGACCCGAGAT
ACAACTCACCGGAACTTACCATCCATGCCTGGGACTTCAGACGTGGCTTC
TTTGGCCCGAAAGCTATTCAAAGAATGCAACAACAACCAACTGCTACTGA
ATTTTTTTCAGCAGGCCGCAAGAGACCCAGAAGGGACACAGAAGTGTATC
AGTCCGACCAAGAAAAGGAGCAAAAAGAAAGCTCGCTTTTCCCCCCAGTC
AAGCTCCTCCGAAGAGTCCCCCCGTGGGAGGACTCGGAACAGGAGCAAAG
CGGGTCGCAAAGCTCAGAGGAAGAGACGGCGACCCTCTCCCAGCAGCTCA
AACAGCAGCTGCAGCAGCAGCGAGTCTTGGGAGTCAAACTCAGACTCCTG
TTCAACCAAGTCCAAAAAATCCAACAAAATCAAGATATCAACCCTACCTT
GTTACCAAGGGGGGGGGATCTAGTATCCTTCTTTCAGGCTGTACCATAAA
TATGTTTCCAGACCCTAAACCTTACTGCCCCTCCAGCAATGACTGGAAAG
AAGAGTATGAGGCCTGTAAATATTGGGATAGACCTCCCAGACACAACCTT
AGAGACCCCCCCTTTTACCCCTGGGCCCCTAAAAACAATCCTTGCAATGT
AAGCTTTAAACTTGGCTTCAAATAAACTAGGCCGTGGGAGTTTCACTTGT
CGGTGTCTACCTCTATAAGTCACTAAGCACTCCGAGCGCAGCGAGGAGTG
CGACCCTTCCCCCTGGTGCAACGCCCTCGGCGGCCGCGCGCTACGCCTTC
GGCTGCGCGCGGCACCTCGGACCCCCGCTCGTGCTGACACGCTTGCGCGT
GTCAGACCACTTCGGGCTCGCGGGGGTCGGGAAATTTGCTAAACAGACTC
CGAGTTGCCATTGGACACTGTAGCTATGAATCAGTAACGAAAGTGAGTGG
GGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGTATTG
GGGGTCGCCATAAACTTTGGGCTCCATTTTAGGCCTTCCGGACTACAAAA
ATCGCCATATTTGTGACGTCAGAGCCGCCATTTTAAGTCAGCTCTGGGGA
GGCGTGACTTCCAGTTCAAAGGTCATCCTCACCATAACTGGCACAAAATG
GCCGCCAACTTCTTCCGGGTCAAAGGTCACTGCTACGTCATAGGTGACGT
GGGGGGGGACCTACTTAAACACGGAAGTAGGCCCCGACACGTCACTGTCA
CGTGACAGTACGTCACAGCCGCCATTTTGTTTTACAAAATAGCCGACTTC
CTTCCTCTTTTTTAAAAAAAGGCGCCAAAAAACCGTCGGCGGGGGGGCCG
CGCGCTGCGCGCGCGGCCCCCGGGGGAGGCACAGCCTCCCCCCCCCGCGC
GCATGCGCGCGGGTCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCC
CCC （ 서열번호 1014）

주석:

추정의 도메인	*염기 범위*
TATA 박스	86 – 90
개시 요소	104 – 119
전사 시작 부위	114
5′ UTR 보존된 도메인	174 – 244
ORF2	354 – 716
ORF2/2	354 – 712; 2372 – 2873
ORF2/3	354 – 712; 2565 – 3075
ORF2t/3	354 – 400; 2565 – 3075
TAIP	373 – 690
ORF1	590 – 2899
ORF1/1	590 – 712; 2372 – 2899
ORF1/2	590 – 712; 2565 – 2873
3개의 오픈-리딩 프레임 영역	2551 – 2870
폴리(A)-신호	3071 – 3076
GC-풍부 영역	3733 – 3853

[0548]

[0549]

[0550] [표 F6] 예시적인 아넬로바이러스 아미노산 서열(알파토르크바이러스)

Ring20 (*알파토르크바이러스* 클레이드 4)	
ORF2	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAET (서열번호 1015)
ORF2/2	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAETLLEIPARNPTPRAIESLEAYKSLTRDTTHRN LPSMPGTSDVASLARKLFKECNNNQLLLNFFQQAARDPEGTQKCISPTKKRS KKKARFSPQSSSSEESPRGRTRNRSKAGRKAQRKRRRPSPSSSNSSCSSSESW ESNSDSCSTKSKKSNKIKISTLPCYQGGGI (서열번호 1016)
ORF2/3	MWTPPRNDQQYLNWQWYTSVLSSHSAMCGCSDAIAHLNHLANLLRAPQN PPPPDNPRPLPVRALPAPPAAHEAAGDRAPWPMGGGGDAGGAGAGGDADH GGAAGGPADADLLDAVAAAETPQETQKGHRSVSVRPRKGAKRKLAFPPSQ APPKSPPVGGLGTGAKRVAKLRGRDGDPLPAAQTAAAAAASLGSQTQTPV QPSPKNPTKSRYQPYLVTKGGGSSILLSGCTINMFPDPKPYCPSSNDWKEEY EACKYWDRPPRHNLRDPPFYPWAPKNNPCNVSFKLGFK (서열번호 1017)
ORF2t/3	MWTPPRNDQQYLNWQWPQETQKGHRSVSVRPRKGAKRKLAFPPSQAPPKS PPVGGLGTGAKRVAKLRGRDGDPLPAAQTAAAAAASLGSQTQTPVQPSPK NPTKSRYQPYLVTKGGGSSILLSGCTINMFPDPKPYCPSSNDWKEEYEACKY WDRPPRHNLRDPPFYPWAPKNNPCNVSFKLGFK (서열번호 1018)

[0551]

TAIP	MINNTLTGNGTLLYLAPTLLCAGVPTLSLILIILLICFVPRKIRPRLIIQDPYPCE HCLLPRLPTRQPVIEHHGLWVVEETPEALAQVETPTMEAPLEDPQTQTC (서열번호 1019)
ORF1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNVRRRRR GRWRRRYRRWKRKGRRRRKAKIIIRQWQPNYRRRCNIVGYLPILICGGNTV SRNYATHSDDTNYPGPFGGGMTTDKFSLRILYDEYKRFMNYWTASNEDLD LCRYLGCTFYFFRHPEVDFIIKINTMPPFLDTTITAPSIHPGLMALDKRARWIP SLKNRPGKKHYIKIRVGAPKMFTDKWYPQTDLCDMTLLTIYATAADMQYP FGSPLTDTVVVNSQVLQSMYDETISILPDEKTKRNSLLTSIRSYIPFYNTTQTI AQLKPFVDAGGHTTGSTTTTWGQLLNTTKFTTTTTTTYTYPGTTNTAVTFIT ANDTWYRGTAYKDNIKDVPQKAAQLYFQTTQKLLGNTFHGSDETLEYHAG LYSSIWLSPGRSYFETPGAYTDIKYNPFTDRGEGNMLWIDWLSKKNMKYDK VQSKCLVADLPLWAAAYGYVEFCSKSTGDTNIHMNARLLIRSPFTDPQLIVH TDPTKGFVPYSLNFGNGKMPGGSSNVPIRMRAKWYPTLSHQQEVLEALAQS GPFAYHSDIKKVSLGIKYRFKWIWGGNPVRQQVVRNPCKEPHSSGNRVPRSI QIVDPRYNSPELTIHAWDFRRGFFGPKAIQRMQQQPTATEFFSAGRKRPRRD TEVYQSDQEKEQKESSLFPPVKLLRRVPPWEDSEQEQSGSQSSEEETATLSQ QLKQQLQQQRVLGVKLRLLFNQVQKIQQNQDINPTLLPRGGDLVSFFQAVP (서열번호 1020)
ORF1/1	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNVVRNPC KEPHSSGNRVPRSIQIVDPRYNSPELTIHAWDFRRGFFGPKAIQRMQQQPTAT EFFSAGRKRPRRDTEVYQSDQEKEQKESSLFPPVKLLRRVPPWEDSEQEQSG SQSSEEETATLSQQLKQQLQQQRVLGVKLRLLFNQVQKIQQNQDINPTLLPR GGDLVSFFQAVP (서열번호 1021)
ORF1/2	MAYGWWRRRRRRWRRWRRRPWRRRWRTRRRRPARRRGRRRNAARDPEG TQKCISPTKKRSKKKARFSPQSSSSEESPRGRTRNRSKAGRKAQRKRRRPSPS SSNSSCSSSESWESNSDSCSTKSKKSNKIKISTLPCYQGGGI (서열번호 1022)

[0552]

[0553] 일부 구현예에서, 아넬로벡터는 본 명세서에 전문이 참조로 포함된 PCT 출원 PCT/US2018/037379에 열거된 서열을 포함하는 핵산을 포함한다. 일부 구현예에서, 아넬로벡터는 본 명세서에 전문이 참조로 포함된 PCT 출원 PCT/US2018/037379에 열거된 서열을 포함하는 폴리펩티드를 포함한다. 일부 구현예에서, 아넬로벡터는 본 명세서에 전문이 참조로 포함된 PCT 출원 PCT/US19/65995에 열거된 서열을 포함하는 핵산을 포함한다. 일부 구현예에서, 아넬로벡터는 본 명세서에 전문이 참조로 포함된 PCT 출원 PCT/US19/65995에 열거된 서열을 포함하는 폴리펩티드를 포함한다.

[0554] *ORF1 분자*

[0555] 일부 구현예에서, 아넬로벡터는 ORF1 분자 및/또는 ORF1 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)의 구조적 특징 및/또는 활성을 갖는 폴리펩티드를 포함한다. 일부 구현예에서, ORF1 분자는 아넬로바이러스 ORF1 단백질(예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질)에 비해 절단을 포함한다. ORF1 분자는 다른 ORF1 분자에 결합하여, 예를 들어, 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같음), 예를 들어, 캡시드를 형성할 수 있다. 일부 구현예에서, 단백질성 외부는 핵산 분자(예를 들어, 본 명세서에 기재된 바와 같은 유전 요소)를 봉입할 수 있다. 일부 구현예에서, 복수의 ORF1 분자는, 예를 들어, 단백질성 외부를 형성하기 위해, 다량체를 형성할 수 있다. 일부 구현예에서, 다량체는 동종다량체일 수 있다. 다른 구현예에

서, 다량체는 이종다량체일 수 있다.

[0556] ORF1 분자는 일부 구현예에서 하기 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.

[0557] *아르기닌-풍부 영역*

[0558] 아르기닌 풍부 영역은 본 명세서에 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 이의 조합)를 포함하는 적어도 약 40개 아미노산의 서열을 갖는다.

[0559] *젤리 롤 도메인*

[0560] 젤리-롤 도메인 또는 영역은 하기의 특성 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):

[0561] (i) 젤리-롤 도메인의 적어도 30%(예를 들어, 적어도 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, 또는 그 이상)의 아미노산이 하나 이상의 β-시트의 부분임;

[0562] (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

[0563] (iii) 젤리-롤 도메인의 3차 구조는 적어도 2개(예를 들어, 적어도 2, 3 또는 4개)의 β-시트를 포함함; 및/또는

[0564] (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 비를 포함함.

[0565] 특정 구현예에서, 젤리-롤 도메인은 2개의 β-시트를 포함한다.

[0566] 특정 구현예에서, 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 구현예에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 β-시트는 8개의 β-가닥을 포함한다. 특정 구현예에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 β-시트는 7개의 β-가닥을 포함한다. 특정 구현예에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 β-시트는 6개의 β-가닥을 포함한다. 특정 구현예에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 β-시트는 5개의 β-가닥을 포함한다. 특정 구현예에서, 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 β-시트는 4개의 β-가닥을 포함한다.

[0567] 일부 구현예에서, 젤리-롤 도메인은 제2 β-시트에 대한 역평행 배향의 제1 β-시트를 포함한다. 특정 구현예에서, 제1 β-시트는 약 4개(예를 들어, 3, 4, 5 또는 6개)의 β-가닥을 포함한다. 특정 구현예에서, 제2 β-시트는 약 4개(예를 들어, 3, 4, 5 또는 6개)의 β-가닥을 포함한다. 구현예에서, 제1 및 제2 β-시트는 총 약 8개(예를 들어, 6, 7, 8, 9, 10, 11 또는 12개)의 β-가닥을 포함한다.

[0568] 특정 구현예에서, 젤리-롤 도메인은 캡시드 단백질(예를 들어, 본 명세서에 기재된 바와 같은 ORF1 분자)의 구성성분이다. 특정 구현예에서, 젤리-롤 도메인은 자가-조립 활성을 갖는다. 일부 구현예에서, 젤리-롤 도메인을 포함하는 폴리펩티드는 젤리-롤 도메인을 포함하는 또 다른 카피의 폴리펩티드에 결합한다. 일부 구현예에서, 제1 폴리펩티드의 젤리-롤 도메인은 제2 카피의 폴리펩티드의 젤리-롤 도메인에 결합한다.

[0569] *N22 도메인*

[0570] ORF1 분자는 또한, 아넬로바이러스 N22 도메인(예를 들어, 본 명세서에 기재된 바와 같음, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 N22 도메인)의 구조 또는 활성을 포함하는 제3 영역, 및/또는 아넬로바이러스 C-말단 도메인(CTD)(예를 들어, 본 명세서에 기재된 바와 같음, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로바이러스 ORF1 단백질로부터의 CTD)의 구조 또는 활성을 포함하는 제4 영역을 포함할 수 있다. 일부 구현예에서, ORF1 분자는 N-말단에서 C-말단 순서로, 제1, 제2, 제3 및 제4 영역을 포함

한다.

[0571] *초가변 영역(HVR)*

[0572] ORF1 분자는 일부 구현예에서, 초가변 영역(HVR), 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 HVR을 추가로 포함할 수 있다. 일부 구현예에서, HVR은 제2 영역과 제3 영역 사이에 위치한다. 일부 구현예에서, 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.

[0573] *예시적인 ORF1 서열*

[0574] 예시적인 아넬로바이러스 ORF1 아미노산 서열, 및 예시적인 ORF1 도메인의 서열이 하기 표에 제공된다. 일부 구현예에서, 본 명세서에 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 구현예에서, 본 명세서에 기재된 아넬로벡터는, 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 구현예에서, 본 명세서에 기재된 아넬로벡터는, 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.

[0575] 일부 구현예에서, 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 열거된 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 구현예에서, ORF1 분자는 상이한 아넬로바이러스 유래의 복수의 하위서열(예를 들어, 표 N 내지 Z에 열거된 알파토르크바이러스 클레이드 1 내지 7 하위서열로부터 선택되는 ORF1 하위서열의 임의의 조합)을 포함한다. 구현예에서, ORF1 분자는 하나의 아넬로바이러스 유래의 Arg-풍부 도메인, 젤리-롤 도메인, N22 도메인, 및 CTD 중 하나 이상, 및 또 다른 아넬로바이러스 유래의 HVR을 포함한다. 구현예에서, ORF1 분자는 하나의 아넬로바이러스 유래의 젤리-롤 도메인, HVR, N22 도메인, 및 CTD 중 하나 이상, 및 또 다른 아넬로바이러스 유래의 Arg-풍부 도메인을 포함한다. 구현예에서, ORF1 분자는 하나의 아넬로바이러스 유래의 Arg-풍부 도메인, HVR, N22 도메인, 및 CTD 중 하나 이상, 및 또 다른 아넬로바이러스 유래의 젤리-롤 도메인을 포함한다. 구현예에서, ORF1 분자는 하나의 아넬로바이러스 유래의 Arg-풍부 도메인, 젤리-롤 도메인, HVR, 및 CTD 중 하나 이상, 및 또 다른 아넬로바이러스 유래의 N22 도메인을 포함한다. 구현예에서, ORF1 분자는 하나의 아넬로바이러스 유래의 Arg-풍부 도메인, 젤리-롤 도메인, HVR, 및 N22 도메인 중 하나 이상, 및 또 다른 아넬로바이러스 유래의 CTD를 포함한다.

[0576] ORF1 분자, 또는 이의 스플라이스 변이체 또는 기능적 단편이 (예를 들어, 유전 요소를 봉입함으로써, 예를 들어, 아넬로벡터의 단백질성 외부를 형성하기 위해) 본 명세서에 기재된 조성물 및 방법에 이용될 수 있는 추가의 예시적인 아넬로바이러스는, 예를 들어, PCT 출원 PCT/US2018/037379 및 PCT/US19/65995(본 명세서에 전문이 참조로 포함됨)에 기재되어 있다.

[0577] [표 N] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스, 클레이드 3)

명칭	Ring1
속/클레이드	*알파토르크바이러스*, 클레이드 3
수탁번호	AJ620231.1
단백질 수탁번호	CAF05750.1

전체 서열: 743 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGR
RKTRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTF
ATNFTSHINDRIMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLE
LTRYLGASVKIYRHPDQDFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHK
ILVPSLQTRPKGRKAIRLRIAPPTLFTDKWYFQKDIADLTLFNIMAVEAD
LRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFNNDNSDSKLKEFLNKAF
PTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQYFAPLDALW
GDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFCHL

TGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNI
YKEGQSKCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFP
KLYNEKVKDYGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVM
EDISRSGPFAPKVEKPSTQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYE
IPGTGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETS
DLVFSGPKKPRVDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQ
EVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCLR
```

(서열번호 185)

[0578]

주석:

추정의 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

[0579]

[0580] [표 0] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스, 클레이드 3)

Ring1 ORF1 (*알파토르크바이러스* 클레이드 3)	
Arg-풍부 영역	MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRR RWRRGRRKTRTYRRRRRFRRRGRK (서열번호 186)
젤리-롤 도메인	AKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDRIMKGPFGG GHSTMRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQ DFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRL RIAPPTLFTDKWYFQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQ VLSSVYNNYLSI (서열번호 187)
초가변 도메인	NTFNNDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKK PNPQINKPLESQYFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMIT YHAKLREFPNSYQGNKAFCHLTGIYSPPYLNQGR (서열번호 188)
N22	ISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQSKCLLTD MPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKD YGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSG PFAPKVEKPSTQLVMKYCFNFN (서열번호 189)
C-말단 도메인	WGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSFRSWD MRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQR ESRPWETEEESETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQ LIRTQQGVHVNPCLR (서열번호 190)

[0581]

[0582] [표 P] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(베타토르크바이러스)

명칭	Ring2
속/클레이드	*베타토르크바이러스*
수탁번호	JX134045.1
단백질 수탁번호	AGG91484.1

전체 서열: 666 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQ
PPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSML
TLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTE
LPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQ
NRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNP
RINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSL
KSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGED
TQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID
TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQV
QYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVEN
PAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQ
DWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQ
QQQLYRERIISLLKDQ (  서열번호    215)
```

[0583]

주석:

추정의 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

[0584]

[0585] [표 Q] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(베타토르크바이러스)

Ring2 ORF1 (*베타토르크바이러스*)	
Arg-풍부 영역	MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (서열번호 216)
젤리-롤 도메인	PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPV HWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKIT FYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTR RKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKP DKLSNNVTLWSLNT (서열번호 217)
초가변 도메인	ISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLL PLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLD MILYSLKSPEAIKNEWTTENMKWNQLNNAG (서열번호 218)
N22	TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELIL EGFPLWLIYWGFADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVI LNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFTPNI QGQLSDNISMFYKFYFK (서열번호 219)
C-말단 도메인	WGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRH GNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEK KEKEVQQLISNLRQQQQLYRERIISLLKDQ (서열번호 220)

[0586]

[0587] [표 D1] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

명칭	Ring 3.1
속/클레이드	*감마토르크바이러스*
수탁번호	
단백질 수탁번호	

전체 서열: 677 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRRRRKR
KKVRRKLKKITIKQWQPDSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDY
VQPKAPQGGGFGCEVFNLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQII
LYRHPDVDFIVSWDNQPPFLLNKYTYPELQPQNLLLARRKRIILSQKSNP
KGKLRIKLRIPPPKQMITKWFFQRDFCDVNLFKLCASAASFRYPGISHGA
QSTIFSAYALNTDFYQCSDWCQTNTETGYLNIKTQQMPLWFHYREGGKEK
WYKYTNKEHRPYTNTYLKSISYNDGLFSPKAMFAFEVKAGGEGTTEPPQG
AQLIANLPLIALRYNPHEDTGHGNEIYLTSTFKGTYDKPKVTDALYFNNV
PLWMGFYGYWDFILQETKNKGVFDQHMFVVKCPALRPISQVTKQVYYPLV
DMDFCSGRLPFDEYLSKDIKSHWYPTAERQTVTINNFVTAGPYMPKFEPT
DKDSTWQLNYHYKFFFKWGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIK
NPEKLHPATLFHDWDLRRGFITQAAIKRMSENLQIDSSFESDGTESPKKK
KRCTKEIPTQNQKQEEIQECLLSLCEEPTCQEETEDLQLFIQQQQQQQQYK
    LRKNLFKLLTHLKKGQRISQLQTGLLE (  서열번호    919)
```

주석:

추정의 도메인	*AA 범위*
Arg-풍부 영역	1 – 59
젤리-롤 도메인	60 - 260
초가변 영역	261 - 356
N22	357 – 517
C-말단 도메인	518 – 677

[0588]

[0589] [표 D2] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

Ring3.1 (*감마토르크바이러스*)	
Arg-풍부 영역	MPFWWRRRNKRWWGRRFRYRRYNKYKTRRRRRIPRRRNRRFTKTRR RRKRKKVRRKLKK (서열번호 920)
젤리-롤 도메인	ITIKQWQPDSVKKCKIKGYSTLVMGAQGKQYNCYTNQASDYVQPKAP QGGGFGCEVFNLKWLYQEYTAHRNIWTKTNEYTDLCRYTGAQIILYR HPDVDFIVSWDNQPPFLLNKYTYPELQPQNLLLARRKRIILSQKSNPKG KLRIKLRIPPPKQMITKWFFQRDFCDVNLFKLCASAASFRYPGISHGAQS TIFSAYAL (서열번호 921)
초가변 도메인	NTDFYQCSDWCQTNTETGYLNIKTQQMPLWFHYREGGKEKWYKYTN KEHRPYTNTYLKSISYNDGLFSPKAMFAFEVKAGGEGTTEPPQGAQLIA N (서열번호 922)
N22	LPLIALRYNPHEDTGHGNEIYLTSTFKGTYDKPKVTDALYFNNVPLWM GFYGYWDFILQETKNKGVFDQHMFVVKCPALRPISQVTKQVYYPLVD MDFCSGRLPFDEYLSKDIKSHWYPTAERQTVTINNFVTAGPYMPKFEPT DKDSTWQLNYHYKFFFK (서열번호 923)
C-말단 도메인	WGGPQVTDPTVEDPCSRNKYPVPDTMQQTIQIKNPEKLHPATLFHDWD LRRGFITQAAIKRMSENLQIDSSFESDGTESPKKKKRCTKEIPTQNQKQE EIQECLLSLCEEPTCQEETEDLQLFIQQQQQQQYKLRKNLFKLLTHLKK GQRISQLQTGLLE (서열번호 924)

[0590]

[0591] [표 R] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

명칭	Ring4
속/클레이드	*감마토르크바이러스*
수탁번호	
단백질 수탁 번호	

전체 서열: 662 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKL
RKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLIN
VPPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLI
FYRDNKTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKP
NGRLTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCN
ENPQVFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNIN
IKSHEYEKSISYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPV
KDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCRQIK
ADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTL
LQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKW
GGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRR
GFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEET
LSCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKEN
          QQMLQLHTGMLP ( 서열번호     925)
```

주석:

추정의 도메인	*AA 범위*
Arg-풍부 영역	1 – 58
젤리-롤 도메인	59 - 260
초가변 영역	261 - 339
N22	340 – 499
C-말단 도메인	500 – 662

[0592]

[0593] [표 S] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(감마토르크바이러스)

Ring4 (*감마토르크바이러스*)	
Arg-풍부 영역	MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRR RRKLRKVKRKKK (서열번호 926)
젤리-롤 도메인	SLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPPKTPY GGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFYRDN KTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGR LTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNE NPQVFFYYL (서열번호 927)
초가변 도메인	NHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHEYEK SISYENGYFQSSFLQTQCIYTSERGEACIAE (서열번호 928)
N22	KPLGIAIYNPVKDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLG LFGYLDYCRQIKADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQS FINGNGPFNQPPTLLQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTES NWELKYKYVFTFK (서열번호 929)
C-말단 도메인	WGGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDY RRGFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKE EETLSCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKL KENQQMLQLHTGMLP (서열번호 930)

[0594]

[0595] [표 D5] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스) 클레이드 1

명칭	Ring 5.2
속/클레이드	*알파토르크바이러스* 클레이드 1
수탁번호	
단백질 수탁번호	

전체 서열: 728 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRRPYRR
RRRRGRRRRRRRRRHKPTLILRQWQPDCIRHCKITGWMPLIICGKGSTQF
NYITHADDITPRGASYGGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLEL
CRYKGTTLKLYRHPEVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHI
VVPSLKTKPRGRKAIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLEL
RNPWLRMSTLSPCIGFNVLKNSIYTNLSNLPQYKNERLNIINNILHPQEI
TGTNNKKWQYTYTKLMAPIYYSANRASTYDWENYSKETNYNNTYVKFTQK
RQEKLTKIRKEWQMLYPQQPTALPDSYDLLQEYGLYSPYYLNPTRINLDW
MTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLLQDMPLFFM
CYGYIDWAIKNTGVSSLVKDARICIRCPYTEPQLVGSTEDIGFVPISETF
MRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLESIISCSPFMPRDQDMNGW
DITIGYKMDFLWGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKL
LGPRTVFHKWDIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAF
RGENREQKECYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRHQ
   RVLRRGLKLVFTDILRLRQGVHWNPELT (  서열번호    931)
```

주석:

추정의 도메인	*AA 범위*
Arg-풍부 영역	1 – 66
젤리-롤 도메인	67 - 277
초가변 영역	278 - 395
N22	396 – 561
C-말단 도메인	562 – 728

[0596]

[0597] [표 D6] 예시적인 아넬로바이러스 ORF1 아미노산 하위서열(알파토르크바이러스) 클레이드 1

Ring5.2 (알파토르크바이러스) 클레이드 1	
Arg-풍부 영역	TAWWWGRWRRRWRRRRPYTTRLRRRRARRAFPRRRRRRFVSRRWRR PYRRRRRRGRRRRRRRRRHK (서열번호 932)
젤리-롤 도메인	PTLILRQWQPDCIRHCKITGWMPLIICGKGSTQFNYITHADDITPRGASY GGNFTNMTFSLEAIYEQFLYHRNRWSASNHDLELCRYKGTTLKLYRHP EVDYIVTYSRTGPFEISHMTYLSTHPMLMLLNKHHIVVPSLKTKPRGRK AIKVRIRPPKLMNNKWYFTRDFCNIGLFQLWATGLELRNPWLRMSTLS PCIGFNVLKNSIYTNL (서열번호 933)
초가변 도메인	SNLPQYKNERLNIINNILHPQEITGTNNKKWQYTYTKLMAPIYYSANRA STYDWENYSKETNYNNTYVKFTQKRQEKLTKIRKEWQMLYPQQPTAL PDSYDLLQEYGLYSPYYLNPTR (서열번호 934)
N22	INLDWMTPYTHVRYNPLVDKGFGNRIYIQWCSEADVSYNRTKSKCLL QDMPLFFMCYGYIDWAIKNTGVSSLVKDARICIRCPYTEPQLVGSTEDI GFVPISETFMRGDMPVLAPYIPLSWFCKWYPNIAHQKEVLESIISCSPFM PRDQDMNGWDITIGYKMDFL (서열번호 935)
C-말단 도메인	WGGSPLPSQPIDDPCQQGTHPIPDPDKHPRLLQVSNPKLLGPRTVFHKW DIRRGQFSKRSIKRVSEYSSDDESLAPGLPSKRNKLDSAFRGENREQKE CYSLLKALEEEETPEEEEPAPQEKAQKEELLHQLQLQRRHQRVLRRGL KLVFTDILRLRQGVHWNPELT (서열번호 936)

[0598]

[0599] 일부 구현예에서, 제1 영역은 핵산 분자(예를 들어, DNA)에 결합할 수 있다. 일부 구현예에서, 염기성 잔기는 아르기닌, 히스티딘 또는 라이신 또는 이의 조합으로부터 선택된다. 일부 구현예에서, 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 구현예에서, 제1 영역은 약 30 내지 120개의 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 구현예에서, 제1 영역은 바이러스 ORF1 아르기닌-풍부 영역(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 아르기닌-풍부 영역)의 구조 또는 활성을 포함한다. 일부 구현예에서, 제1 영역은 핵 국소화 신호를 포함한다.

[0600] 일부 구현예에서, 제2 영역은 젤리-롤 도메인, 예를 들어, 바이러스 ORF1 젤리-롤 도메인(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 단백질로부터의 젤리-롤 도메인)의 구조 또는 활성을 포함한다. 일부 구현예에서, 제2 영역은 또 다른 ORF1 분자의 제2 영역에 결합하여, 예를 들어, 단백질성 외부(예를 들어, 캡시드) 또는 이의 부분을 형성할 수 있다.

[0601] 일부 구현예에서, 제4 영역은 단백질성 외부(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1 분자의 다량체를 포함하는 단백질성 외부)의 표면 상에 노출된다.

[0602] 일부 구현예에서, 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.

[0603] 일부 구현예에서, 제1 영역, 제2 영역, 제3 영역, 제4 영역, 및/또는 HVR 중 하나 이상은 이종 아미노산 서열(예를 들어, 이종 ORF1 분자로부터의 상응하는 영역)에 의해 대체될 수 있다. 일부 구현예에서, 이종 아미노산 서열은 예를 들어, 본 명세서에 기재된 바와 같은, 요망되는 작용성을 갖는다.

[0604] 일부 구현예에서, ORF1 분자는 (예를 들어, PCT/US19/65995의 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, 또는 그 이상의 아미노산을 포함하는 모티프)를 포함한다. 일부 구현예에서, 보존된 모티프

는 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. 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 보존된 모티프는 표 19에 열거된 바와 같은 하나 이상의 아미노산 서열을 포함한다.

[0605] 일부 구현예에서, ORF1 분자는 예를 들어, 본 명세서에 기재된 바와 같은, 야생형 ORF1 단백질에 비하여 적어도 하나의 차이(예를 들어, 돌연변이, 화학적 변형 또는 후생적 변화)를 포함한다.

[0606] *N22 도메인 내의 보존된 ORF1 모티프*

[0607] 일부 구현예에서, 본 명세서에 기재된 폴리펩티드(예를 들어, ORF1 분자)는 아미노산 서열 $YNPX^2DXGX^2N$(서열번호 829)을 포함하며, X^n은 임의의 n개의 아미노산의 인접 서열이다. 예를 들어, X^2는 임의의 2개의 아미노산의 인접 서열을 나타낸다. 일부 구현예에서, $YNPX^2DXGX^2N$(서열번호 829)은 예를 들어, 본 명세서에 기재된 바와 같은, ORF1 분자의 N22 도메인 내에 포함된다. 일부 구현예에서, 본 명세서에 기재된 유전 요소는 아미노산 서열 $YNPX^2DXGX^2N$(서열번호 829)을 인코딩하는 핵산 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1 분자를 인코딩하는 핵산 서열)을 포함하며, X^n은 임의의 n개의 아미노산의 인접 서열이다.

[0608] 일부 구현예에서, 폴리펩티드(예를 들어, ORF1 분자)는 예를 들어, N22 도메인 내의, 예를 들어, $YNPX^2DXGX^2N$(서열번호 829) 모티프의 부분을 포함하고/포함하거나 이에 측접하는 보존된 2차 구조를 포함한다. 일부 구현예에서, 보존된 2차 구조는 제1 베타 가닥 및/또는 제2 베타 가닥을 포함한다. 일부 구현예에서, 제1 베타 가닥은 약 5 내지 6개(예를 들어, 3, 4, 5, 6, 7 또는 8개)의 아미노산 길이이다. 일부 구현예에서, 제1 베타 가닥은 $YNPX^2DXGX^2N$(서열번호 829) 모티프의 N-종말단에 티로신(Y) 잔기를 포함한다. 일부 구현예에서, $YNPX^2DXGX^2N$(서열번호 829) 모티프는 랜덤 코일(예를 들어, 약 8 내지 9개 아미노산의 랜덤 코일)을 포함한다. 일부 구현예에서, 제2 베타 가닥은 약 7 내지 8개(예를 들어, 5, 6, 7, 8, 9 또는 10개) 아미노산 길이이다. 일부 구현예에서, 제2 베타 가닥은 $YNPX^2DXGX^2N$(서열번호 829) 모티프의 C-종말단에 아스파라긴(N) 잔기를 포함한다.

[0609] 예시적인 $YNPX^2DXGX^2N$(서열번호 829) 모티프-측접 2차 구조는 PCT/US19/65995(본 명세서에 전문이 참조로 포함됨)의 실시예 47 및 도 48에 기재되어 있다. 일부 구현예에서, ORF1 분자는 PCT/US19/65995의 도 48에 나타낸 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 구현예에서, ORF1 분자는 (예를 들어, 본 명세서에 기재된 바와 같은) $YNPX^2DXGX^2N$(서열번호 829) 모티프에 측접하는, PCT/US19/65995의 도 48에 나타낸 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.

[0610] *ORF1 젤리-롤 도메인 내의 보존된 2차 구조 모티프*

[0611] 일부 구현예에서, 본 명세서에 기재된 폴리펩티드(예를 들어, ORF1 분자)는 (예를 들어, 본 명세서에 기재된 바와 같은) 아넬로바이러스 ORF1 단백질에 의해 포함되는 하나 이상의 2차 구조 요소를 포함한다. 일부 구현예에서, ORF1 분자는 (예를 들어, 본 명세서에 기재된 바와 같은) 아넬로바이러스 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 구현예에서, 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.

[0612] 일부 구현예에서, 보존된 2차 구조 요소(즉, 베타 가닥 및/또는 알파 나선)의 쌍은 예를 들어, 랜덤 코일 서열, 베타 가닥 또는 알파 나선 또는 이의 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, 또는 그 이상의 아미노산을 포함할 수 있다. 일부 구현예에서, ORF1 분자는 하나 이상의 추가의 베타 가닥 및/또는 알파 나선을 (예를 들어, 젤리-롤 도메인 내에) 추가로 포함할 수 있다. 일부 구현예에서, 연속 베타 가닥 또는 연속 알파 나선이 조합될 수 있다. 일부 구현예에서, 제1 베타 가닥 및 제2 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 구현예에서, 제3 베타 가닥 및 제4 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 구현예에서, 제4 베타 가닥 및 제5 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 구현예에서, 제6 베타 가닥 및 제7 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 구현예에서, 제7 베타 가닥 및 제8 베타 가닥은 더 큰 베타 가닥 내에 포함된다. 일부 구현예에서, 제8 베타 가닥 및 제9 베타 가닥은 더 큰 베타 가닥 내에 포함된다.

[0613] 일부 구현예에서, 제1 베타 가닥은 약 5 내지 7개(예를 들어, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다. 일부 구현예에서, 제2 베타 가닥은 약 15 내지 16개(예를 들어, 13, 14, 15, 16, 17, 18 또는 19개)의 아미노산 길이이다. 일부 구현예에서, 제1 알파 나선은 약 15 내지 17개(예를 들어, 13, 14, 15, 16, 17, 18, 19 또는 20개)의 아미노산 길이이다. 일부 구현예에서, 제3 베타 가닥은 약 3 내지 4개(예를 들어, 1, 2, 3, 4, 5 또는 6개)의 아미노산 길이이다. 일부 구현예에서, 제4 베타 가닥은 약 10 내지 11개(예를 들어, 8, 9, 10, 11, 12 또는 13개)의 아미노산 길이이다. 일부 구현예에서, 제5 베타 가닥은 약 6 내지 7개(예를 들어, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다. 일부 구현예에서, 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 구현예에서, 제2 알파 나선은 (예를 들어, 랜덤 코일 서열에 의해 분리된) 2개의 더 작은 알파 나선으로 파단될 수 있다. 일부 구현예에서, 2개의 더 작은 알파 나선의 각각은 약 4 내지 6개(예를 들어, 2, 3, 4, 5, 6, 7 또는 8개)의 아미노산 길이이다. 일부 구현예에서, 제6 베타 가닥은 약 4 내지 5개(예를 들어, 2, 3, 4, 5, 6 또는 7개)의 아미노산 길이이다. 일부 구현예에서, 제7 베타 가닥은 약 5 내지 6개(예를 들어, 3, 4, 5, 6, 7, 8 또는 9개)의 아미노산 길이이다. 일부 구현예에서, 제8 베타 가닥은 약 7 내지 9개(예를 들어, 5, 6, 7, 8, 9, 10, 11, 12 또는 13개)의 아미노산 길이이다. 일부 구현예에서, 제9 베타 가닥은 약 5 내지 7개(예를 들어, 3, 4, 5, 6, 7, 8, 9 또는 10개)의 아미노산 길이이다.

[0614] 예시적인 젤리-롤 도메인 2차 구조는 PCT/US19/65995의 실시예 47 및 본 명세서의 도 25에 기재되어 있다. 일부 구현예에서, ORF1 분자는 본 명세서의 도 25에 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.

[0615] *컨센서스 ORF1 도메인 서열*

[0616] 일부 구현예에서, 예를 들어, 본 명세서에 기재된 바와 같은, ORF1 분자는 젤리-롤 도메인, N22 도메인, 및/또는 C-말단 도메인(CTD) 중 하나 이상을 포함한다. 일부 구현예에서, 젤리-롤 도메인은 본 명세서에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 어느 하나에 열거된 바와 같은) 젤리-롤 도메인 컨센서스 서열을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, N22 도메인은 본 명세서에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 어느 하나에 열거된 바와 같은) N22 도메인 컨센서스 서열을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, CTD 도메인은 본 명세서에 기재된 바와 같은(예를 들어, 표 37A 내지 37C 중 어느 하나에 열거된 바와 같은) CTD 도메인 컨센서스 서열을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 표 37A 내지 37C 중 어느 하나에 형식 "(Xa-b)"로 열거된 아미노산은 인접 아미노산의 연속물을 포함하며, 여기서, 연속물은 적어도 a개 및 최대 b개의 아미노산을 포함한다. 특정 구현예에서, 연속물에서 모든 아미노산은 동일하다. 다른 구현예에서, 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.

[0617] [표 37A] 알파토르크바이러스 ORF1 도메인 컨센서스 서열

도메인	서열	서열번호
젤리-롤	LVLTQWQPNTVRRCYIRGYLPLIICGEN(X_{0-3})TTSRNYATHS DDTIQKGPFGGGMSTTTFSLRVLYDEYQRFMNRWTYSNED LDLARYLGCKFTFYRHPDXDFIVQYNTNPPFKDTKLTAPSIH P(X_{1-5})GMLMLSKRKILIPSLKTRPKGKHYVKVRIGPPKLFED KWYTQSDLCDVPLVXLYATAADLQHPFGSPQTDNPCVTFQ VLGSXYNKHLSISP; 여기서 X = 임의의 아미노산임.	227
N22	SNFEFPGAYTDITYNPLTDKGVGNMVWIQYLTKPDTIXDKT QS(X_{0-3})KCLIEDLPLWAALYGYVDFCEKETGDSAIIXNXGRV LIRCPYTKPPLYDKT(X_{0-4})NKGFVPYSTNFGNGKMPGGSGY VPIYWRARWYPTLFHQKEVLEDIVQSGPFAYKDEKPSTQLV MKYCFNFN; 여기서 X = 임의의 아미노산임.	228
CTD	WGGNPISQQVVRNPCKDSG(X_{0-3})SGXGRQPRSVQVVDPKY MGPEYTFHSWDWRRGLFGEKAIKRMSEQPTDDEIFTGGXPK RPRRDPPTXQXPEE(X_{1-4})QKESSSFR(X_{2-14})PWESSSQEXESES QEEEE(X_{0-30})EQTVQQQLRQQLREQRRLRVQLQLLFQQLLKT (X_{0-4})QAGLHINPLLLSQA(X_{0-40})*; 여기서 X = 임의의 아미노산임.	229

[0618]

[0619] [표 37B] 베타토르크바이러스 ORF1 도메인 컨센서스 서열

도메인	서열	서열번호
젤리-롤	LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESIVPH HEPGGGGWSIQQFTLGALYEEHLKLRNWWTKSNDGLPLVR YLGCTIKLYRSEDTDYIVTYQRCYPMTATKLTYLSTQPSRM LMNKHKIIVPSKXT(X_{1-4})NKKKKPYKKIFIKPPSQMQNKWYF QQDIANTPLLQLTXTACSLDRMYLSSDSISNNITFTSLNTNFF QNPNFQ; 여기서 X = 임의의 아미노산임.	230
N22	(X_{4-10})TPLYFECRYNPFKDKGTGNKVYLVSNN(X_{1-8})TGWDPP TDPDLIIEGFPLWLLLWGWLDWQKKLGKIQNIDTDYILVIQS XYYIPP(X_{1-3})KLPYYVPLDXD(X_{0-2})FLHGRSPY(X_{3-16})PSDKQH WHPKVRFQXETINNIALTGPGTPKLPNQKSIQAHMKYKFYF K; 여기서 X = 임의의 아미노산임.	231
CTD	WGGCPAPMETITDPCKQPKYPIPNNLLQTTSLQXPTTPIETYL YKFDERRGLLTKKAAKRIKKDXTTETTLFTDTGXXTSTTLPT XXQTETTQEEXTSEEE(X_{0-5})ETLLQQLQQLRRKQKQLRXRIL QLLQLLXLL(X_{0-26})*; 여기서 X = 임의의 아미노산임.	232

[0620]

[0621] [표 37C] 감마토르크바이러스 ORF1 도메인 컨센서스 서열

도메인	서열	서열번호
젤리-롤	TIPLKQWQPESIRKCKIKGYGTLVLGAEGRQFYCYTNEKDE YTPPKAPGGGGFGVELFSLEYLYEQWKARNNIWTKSNXYK DLCRYTGCKITFYRHPTTDFIVXYSRQPPFEIDKXTYMXXHP QXLLLRKHKKIILSKATNPKGKLKKKIKIKPPKQMLNKWFF QKQFAXYGLVQLQAAACBLRYPRLGCCNENRLITLYYLN; 여기서 X = 임의의 아미노산임.	233
N22	LPIVVARYNPAXDTGKGNKXWLXSTLNGSXWAPPTTDKDL IIEGLPLWLALYGYWSYJKKVKKDKGILQSHMFVVKSPAIQP LXTATTQXTFYPXIDNSFIQGKXPYDEPJTXNQKKLWYPTLE HQQETINAIVESGPYVPKLDNQKNSTWELXYXYTFYFK; 여기서 X = 임의의 아미노산임.	234
CTD	WGGPQIPDQPVEDPKXQGTYPVPDTXQQTIQIXNPLKQKPE TMFHDWDYRRGIITSTALKRMQENLETDSSFXSDSEETP(X_{0-2})KKKKRLTXELPXPQEETEEIQSCLLSLCEESTCQEE(X_{1-6})ENL QQLIHQQQQQQQQLKHNILKLLSDLKZKQRLLQLQTGILE(X_{1-10})*; 여기서 X = 임의의 아미노산임.	235

[0622]

[0623] 일부 구현예에서, 젤리-롤 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 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 구현예에서, 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 구현예에서, 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.

[0624] *ORF1 단백질 서열의 확인*

[0625] 일부 구현예에서, 아넬로바이러스 ORF1 단백질 서열, 또는 ORF1 단백질을 인코딩하는 핵산 서열은 아넬로바이러스의 게놈(예를 들어, 핵산 서열결정 기법, 예를 들어, 심층 서열결정 기법에 의해 확인된, 예를 들어, 추정의 아넬로바이러스 게놈)으로부터 확인될 수 있다. 일부 구현예에서, ORF1 단백질 서열은 하기 선택 기준 중 하나 이상(예를 들어, 1, 2개, 또는 3개 전부)에 의해 확인된다:

[0626] *(i) 길이 선택:* 단백질 서열(예를 들어, 하기 (ii) 또는 (iii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)은 추정 아넬로바이러스 ORF1 단백질을 확인하기 위해 약 600개 초과의 아미노산 잔기에 대해 크기-선택될 수 있다. 부 구현예에서, 아넬로바이러스 ORF1 단백질 서열은 길이가 적어도 약 600, 650, 700, 750, 800, 850, 900, 950, 또는 1000개의 아미노산 잔기이다. 일부 구현예에서, 알파토르크바이러스 ORF1 단백질 서열은 길이가 적어도 약 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 900, 또는 1000개 아미노산 잔기이다. 일부 구현예에서, 베타토르크바이러스 ORF1 단백질 서열은 길이가 적어도 약 650, 660, 670, 680, 690, 700, 750, 800, 900, 또는 1000개 아미노산 잔기이다. 일부 구현예에서, 감마토르크바이러스 ORF1 단백질 서열은 길이가 적어도 약 650, 660, 670, 680, 690, 700, 750, 800, 900, 또는 1000개의 아미노산 잔기이다. 일부 구현예에서, 아넬로바이러스 ORF1 단백질을 인코딩하는 핵산 서열은 길이가 적어도 약 1800, 1900, 2000,

2100, 2200, 2300, 2400, 또는 2500개 뉴클레오티드이다. 일부 구현예에서, 알파토르크바이러스 ORF1 단백질 서열을 인코딩하는 핵산 서열은 길이가 적어도 약 2100, 2150, 2200, 2250, 2300, 2400, 또는 2500개 뉴클레오티드이다. 일부 구현예에서, 베타토르크바이러스 ORF1 단백질 서열을 인코딩하는 핵산 서열은 길이가 적어도 약 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, 또는 2500개 또는 1000개 뉴클레오티드이다. 일부 구현예에서, 감마토르크바이러스 ORF1 단백질 서열을 인코딩하는 핵산 서열은 길이가 적어도 약 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, 또는 2500 또는 1000개 뉴클레오티드이다.

[0627] *(ii) ORF1 모티프의 존재:* 단백질 서열(예를 들어, 상기 (i) 또는 하기 (iii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)을 필터링하여 상기 기재된 N22 도메인에 보존된 ORF1 모티프를 함유하는 것들을 확인할 수 있다. 일부 구현예에서, 추정 아넬로바이러스 ORF1 서열은 서열 YNPXXDXGXXN을 포함한다. 일부 구현예에서, 추정 아넬로바이러스 ORF1 서열은 서열 Y[NCS]P*XX*D*X*[GASKR]*XX*[NTSVAK]을 포함한다.

[0628] *(iii) 아르기닌-풍부 영역의 존재:* 단백질 서열(예를 들어, 상기 (i) 및/또는 (ii)에 기재된 기준을 통과하는 추정 아넬로바이러스 ORF1 서열)은 아르기닌-풍부 영역(예를 들어, 본 명세서에 기재된 바와 같음)을 포함하는 것에 대해 필터링될 수 있다. 일부 구현예에서, 추정 아넬로바이러스 ORF1 서열은 적어도 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 70개 아미노산의 인접 서열을 포함한다. 일부 구현예에서, 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 70개 아미노산의 인접 서열을 포함한다. 일부 구현예에서, 아르기닌-풍부 영역은 추정 아넬로바이러스 ORF1 단백질의 시작 코돈의 적어도 약 30, 40, 50, 60, 70 또는 80개 아미노산 하류에 위치한다. 일부 구현예에서, 아르기닌-풍부 영역은 추정 아넬로바이러스 ORF1 단백질의 시작 코돈의 적어도 약 50개 아미노산 하류에 위치한다.

[0629] *ORF2 분자*

[0630] 일부 구현예에서, 아넬로벡터는 ORF2 분자 및/또는 ORF2 분자를 인코딩하는 핵산을 포함한다. 일반적으로, ORF2 분자는 아넬로바이러스 ORF2 단백질(예를 들어, 본 명세서에 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 어느 하나에 열거된 바와 같은 아넬로바이러스 ORF2 단백질)의 구조적 특징 및/또는 활성을 갖는 폴리펩티드 또는 이의 기능적 단편을 포함한다. 일부 구현예에서, ORF2 분자는 표 A2, A4, A6, A8, A10, A12, C1 내지 C5, 2, 4, 6, 8, 10, 12, 14, 16 또는 18 중 어느 하나에 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.

[0631] 일부 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, ORF2 분자는 나선-회전-나선 모티프(예를 들어, 회전 영역에 측접하는 2개의 알파 나선을 포함하는 나선-회전-나선 모티프)를 포함한다. 일부 구현예에서, ORF2 분자는 TTV 분리주 TA278 또는 TTV 분리주 SANBAN의 ORF2 단백질의 아미노산 서열을 포함하지 않는다. 일부 구현예에서, ORF2 분자는 단백질 포스파타제 활성을 갖는다. 일부 구현예에서, ORF2 분자는 예를 들어, 본 명세서에 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 어느 하나에 나타낸 바와 같은) 야생형 ORF2 단백질에 비하여 적어도 하나의 차이(예를 들어, 돌연변이, 화학적 변형 또는 후생적 변화)를 포함한다.

[0632] *보존된 ORF2 모티프*

[0633] 일부 구현예에서, 본 명세서에 기재된 폴리펩티드(예를 들어, ORF2 분자)는 아미노산 서열 [W/F]X^7HX^3CX^1CX^5H(서열번호 949)를 포함하며, 여기서, X^n은 임의의 n개의 아미노산의 인접 서열이다. 구현예에서, X^7은 임의의 7개의 아미노산의 인접 서열을 나타낸다. 구현예에서, X^3은 임의의 3개의 아미노산의 인접 서열을 나타낸다. 구현예에

서, X^1은 임의의 단일의 아미노산을 나타낸다. 구현예에서, X^5는 임의의 5개의 아미노산의 인접 서열을 나타낸다. 일부 구현예에서, [W/F]는 트립토판 또는 페닐알라닌 중 어느 하나일 수 있다. 일부 구현예에서, $[W/F]X^7HX^3CX^1CX^5H$(서열번호 949)는 예를 들어, 본 명세서에 기재된 바와 같은 ORF2 분자의 N22 도메인 내에 포함된다. 일부 구현예에서, 본 명세서에 기재된 유전 요소는 아미노산 서열 $[W/F]X^7HX^3CX^1CX^5H$(서열번호 949)를 인코딩하는 핵산 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF2 분자를 인코딩하는 핵산 서열)을 포함하며, 여기서, X^n은 임의의 n개의 아미노산의 인접 서열이다.

[0634] *유전 요소, 예를 들어, 비-아넬로바이러스 서열을 포함하는 유전 요소*

[0635] 일부 구현예에서, 아넬로벡터는 유전 요소를 포함한다. 일부 구현예에서, 유전 요소는 아넬로바이러스 이외의 바이러스 유래의 핵산 서열(예를 들어, 길이가 적어도 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, 또는 4000개 뉴클레오티드인 인접 핵산 서열), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 아넬로바이러스 이외의 바이러스는 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV)이다. 일부 구현예에서, 아넬로바이러스 이외의 바이러스는 AAV(예를 들어, AAV1, AAV2, 또는 AAV5)이다. 일부 구현예에서, 아넬로바이러스 이외의 바이러스 유래의 핵산 서열은 비-아넬로바이러스 복제 원점(예를 들어, AAV, 예를 들어, AAV1, AAV2, 또는 AAV5로부터 유래된 복제 원점)을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은 AAV Rep-결합 모티프(RBM), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은 AAV 말단 분해 부위(TRS), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은 회전환 복제에 의해 복제하는 바이러스로부터 유래된다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은 회전 헤어핀 복제에 의해 복제하는 바이러스로부터 유래된다.

[0636] 일부 구현예에서, 유전 요소는 하나 이상의 역위 말단 반복부(ITR)를 포함한다. 일부 구현예에서, 유전 요소는 하나의 ITR을 포함한다. 일부 구현예에서, 유전 요소는 본 명세서에 기재된 바와 같은 이펙터 또는 이펙터-인코딩 서열에 대해 5'에 위치하는 ITR을 포함한다. 일부 구현예에서, 유전 요소는 본 명세서에 기재된 바와 같은 이펙터 또는 이펙터-인코딩 서열에 대해 3'에 위치하는 ITR을 포함한다. 일부 구현예에서, 유전 요소는, 예를 들어, 본 명세서에 기재된 바와 같은 이펙터 또는 이펙터-인코딩 서열에 측접하는 2개의 ITR을 포함한다. 일부 구현예에서, 비-아넬로바이러스 복제 원점은, 예를 들어, 본 명세서에 기재된 바와 같은, ITR, 예를 들어, AAV ITR에 포함된다.

[0637] 일부 구현예에서, 유전 요소는 AAV(예를 들어, AAV1, AAV2, AAV3, AAV4, AAV5, 또는 AAV6) 유래의 ITR 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, AAV ITR은, 예를 들어, 문헌[Grimm et al., 2005, *J. Virol.*, DOI: 10.1128/JVI.80.1.426-439.2006](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같은, 예를 들어, 문헌[Grimm et al., 상기함]의 도 1A에 나타낸 바와 같은, 서열을 갖는다. 구현예에서, AAV ITR은 문헌[Chiorini et al., 1999, *J. Virol* 73(5): 4293-4298](본 명세서에 전문이 참조로 포함됨)에 기재된 바와 같은 서열을 갖는다.

[0638] 일부 구현예에서, 유전 요소는 (예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, AAV 유래의) ITR 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열의 하위서열을 포함한다. 구현예에서, 유전 요소는 AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCC(서열번호 1051)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 CGGGCGGGTGGTGGCGGCGGTTGGGGCTCGGCGCTCGCTCGCTCGCTGGGCGGGCGGGCGGT(서열번호 1052)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다.

[0639] 일부 구현예에서, 유전 요소는 (예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, AAV 유래의) RBM 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 (GMGY)x4(서열번호 1053)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 (GMGY)x5(서열번호 1054)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 GCGCGCTCGCTCGCTC(서열번호 1055)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 GCTCGCTCGCTCGCTG(서열번호 1056)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다.

[0640] 일부 구현예에서, 유전 요소는 (예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, AAV 유래의) TRS 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 XGTTGG(서열번호 1057)(여기서 X는 G, C, T, 또는 A로부터 선택됨)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 AGTTGG(서열번호 1058)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 유전 요소는 GGTTGG(서열번호 1059)의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다.

[0641] 일부 구현예에서, 유전 요소 작제물(예를 들어, 본 명세서에 기재된 바와 같음)은 하기 표 61에 나타낸 바와 같은, 또는 도 10에 도해된 바와 같은 구조를 갖는 핵산 서열을 포함한다.

[0642] 일부 구현예에서, 유전 요소(예를 들어, 본 명세서에 기재된 바와 같음)는 하기 표 61에 나타낸 바와 같은, 또는 도 10에 도해된 바와 같은 구조를 갖는 핵산 서열을 포함한다. 구현예에서, 유전 요소는 (i) 하나 이상(예를 들어, 1 또는 2개)의 비-아넬로바이러스(예를 들어, AAV) ITR 서열; (ii) 외인성 이펙터를 인코딩하는 서열; 및/또는 (iii) 아넬로바이러스 게놈 유래의 서열(예를 들어, 인접 또는 비-인접 서열)(또는 이에 대하여 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열), 또는 길이가 적어도 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, 또는 4000개 뉴클레오티드인 이의 인접 부분 중 1, 2개, 또는 전부를 포함한다.

[0643] 구현예에서, 유전 요소는 아넬로바이러스 게놈 또는 이의 일부 내에 위치한 비-아넬로바이러스(예를 들어, AAV) ITR 서열을 포함한다. 일 구현예에서, 비-아넬로바이러스 ITR 서열은, 아넬로바이러스 게놈 서열 또는 이의 일부의 3' 말단보다, 아넬로바이러스 게놈 서열 또는 이의 일부의 5' 말단에 더 가깝게 위치한다. 일 구현예에서, 비-아넬로바이러스 ITR 서열은, 아넬로바이러스 게놈 서열 또는 이의 일부의 5' 말단보다, 아넬로바이러스 게놈 서열 또는 이의 일부의 3' 말단에 더 가깝게 위치한다.

[0644] 구현예에서, 유전 요소는 아넬로바이러스 게놈 서열 또는 이의 일부의 5' 말단에 위치한 비-아넬로바이러스(예를 들어, AAV) ITR 서열을 포함한다. 구현예에서, 유전 요소는 아넬로바이러스 게놈 서열 또는 이의 일부의 3' 말단에 위치한 비-아넬로바이러스(예를 들어, AAV) ITR 서열을 포함한다.

[0645] 구현예에서, 비-아넬로바이러스 ITR 서열은 아넬로바이러스 게놈 서열 또는 이의 일부와 동일한 배향을 공유한다. 구현예에서, 비-아넬로바이러스 ITR 서열은 아넬로바이러스 게놈 서열 또는 이의 일부로부터 역 배향을 갖는다.

[0646] 구현예에서, 유전 요소는 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터)를 인코딩하는 서열을 포함한다. 구현예에서, 이펙터를 인코딩하는 서열은 비-아넬로바이러스 ITR 서열의 상류에 위치한다. 구현예에서, 이펙터를 인코딩하는 서열은 비-아넬로바이러스 ITR 서열의 하류에 위치한다.

[0647] 구현예에서, 유전 요소는 복수(예를 들어, 2개)의 비-아넬로바이러스 ITR 서열을 포함한다. 구현예에서, 복수의 비-아넬로바이러스 ITR 서열은 동일한 서열을 공유한다. 구현예에서, 복수의 비-아넬로바이러스 ITR 서열은 상이한 서열을 갖는다. 구현예에서, 복수의 비-아넬로바이러스 ITR 서열은 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 핵산 서열 동일성을 공유한다. 구현예에서, 유전 요소는 동일한 배향을 공유하는 2개의 비-아넬로바이러스 ITR 서열을 포함한다. 구현예에서, 유전 요소는 반대 배향을 갖는 2개의 비-아넬로바이러스 ITR 서열을 포함한다. 구현예에서, 유전 요소는 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터)를 인코딩하

는 서열을 포함하며, 여기서 이펙터를 인코딩하는 서열은 비-아넬로바이러스 ITR 서열 중 하나 이상과 동일한 배향을 공유한다. 구현예에서, 유전 요소는 이펙터(예를 들어, 내인성 이펙터 또는 외인성 이펙터)를 인코딩하는 서열을 포함하며, 여기서 이펙터를 인코딩하는 서열은 비-아넬로바이러스 ITR 서열 중 하나 이상과 반대 배향으로 있다.

[0648] **[표 61] 예시적인 AAV-아넬로바이러스 유전 요소 구조**

플라스미드 번호	*플라스미드 명칭*	*설명*	*개략도*	*복제/패키징?*
pRTx-1260	pRing2-5'NCR-AAV-Ori-FWD	IVC로 양성 가닥을 만들기 위해, FWD 배향의 5'NCR에 단일 AAV-TRS-RBM이 있는 Ring2	도 10a	예
pRTx-1261	pRing2-5'NCR-AAV-Ori-Rev	IVC로 음성 가닥을 만들기 위해, Rev 배향의 5'NCR에 단일 AAV-TRS-RBM이 있는 Ring2	도 10b	예

[0649]

pRTx-1262	pRing2-3'NCR-AAV-Ori-FWD	IVC로 양성 가닥을 만들기 위해, FWD 배향의 3'NCR에 단일 AAV-TRS-RBM이 있는 Ring2	도 10c	예
pRTx-1263	pRing2-3'NCR-AAV-Ori-Rev	IVC로 음성 가닥을 만들기 위해, Rev 배향의 3'NCR에 단일 AAV-TRS-RBM이 있는 Ring2	도 10d	예
진행 중	pRing2-5'AAV-Ori	IVC로 양성 가닥을 만들기 위해, FWD 배향의 5'NCR 앞에 단일 AAV-TRS-RBM이 있는 Ring2	도 10e	
진행 중	pRing2-5'AAV-Ori-Rev	IVC로 음성 가닥을 만들기 위해, Rev 배향의 5'NCR 앞에 단일 AAV-TRS-RBM이 있는 Ring2	도 10f	
진행 중	pRing2ΔORF::hEF1a_EGFP-5'AAV-Ori	IVC로 양성 가닥을 만들기 위해, FWD 배향의 5'NCR 앞에 단일 AAV-TRS-RBM이 있는, ORF가 hEF1a-EGFP 유전자로 대체된 Ring2 벡터	도 10g	
진행 중	pRing2ΔORF::hEF1a_EGFP-5'AAV-Ori-Rev	IVC로 음성 가닥을 만들기 위해, Rev 배향의 5'NCR 앞에 단일 AAV-TRS-RBM이 있는, ORF가 hEF1a-EGFP 유전자로 대체된 Ring2 벡터	도 10h	

[0650]

진행 중	pRing2-2xAAV-Ori	플라스미드의 양성 가닥을 만들기 위해, FWD 배향으로 AAV-TRS-RBM이 측접된 Ring2	도 10i	
진행 중	pRing2-2xAAV-Ori-Rev	플라스미드의 음성 가닥을 만들기 위해, Rev 배향으로 AAV-TRS-RBM이 측접된 Ring2	도 10j	
pRTx-1472	pRing2ΔORF::hEF1a_EGFP-2xAAV-Ori	플라스미드의 양성 가닥을 만들기 위해, FWD 배향으로 ORF가 AAV-TRS-RBM이 측접된 hEF1a-EGFP 유전자로 대체된 Ring2 벡터	도 10k	예
진행 중	pRing2ΔORF::hEF1a_EGFP-2xAAV-Ori-Rev	플라스미드의 음성 가닥을 만들기 위해, FWD 배향으로 ORF가 AAV-TRS-RBM이 측접된 hEF1a-EGFP 유전자로 대체된 Ring2 벡터	도 10l	

[0651]

[0652] 일부 구현예에서, 유전 요소는 비-아넬로바이러스 Rep 분자, 예를 들어, 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV) 유래의 Rep 단백질; 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 폴리펩티드의 존재 하에 복제를 거칠 수 있다. 일부 구현예에서, 유전 요소는 AAV Rep 분자, 예를 들어, AAV Rep 단백질(예를 들어, AAV1, AAV2, 또는 AAV5 Rep 단백질), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 폴리펩티드의 존재 하에 복제를 거칠 수 있다.

[0653] 일부 구현예에서, 유전 요소는 선형이다. 일부 구현예에서, 유전 요소는 환형이다. 일부 구현예에서, 유전 요소는 단일-가닥이다. 일부 구현예에서, 유전 요소는 이중-가닥이다. 일부 구현예에서, 유전 요소는 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% DNA로 이루어진다. 일부 구현예에서, 유전 요소는 100% DNA이다.

[0654] 일부 구현예에서, 유전 요소는 하기의 특성 중 하나 이상을 갖는다: 숙주 세포의 게놈과 실질적으로 비-통합성임, 에피솜 핵산임, 단일 가닥 DNA임, 환형임, 약 1 내지 10 kb임, 세포의 핵 내에 존재함, 내인성 단백질에 의해 결합될 수 있음, 이펙터, 예컨대 숙주 또는 표적 세포의 유전자, 활성 또는 기능을 표적화하는 폴리펩티드 또는 핵산(예를 들어, RNA, iRNA, 마이크로RNA)를 생성함. 일 구현예에서, 유전 요소는 실질적으로 비-통합성 DNA이다. 일부 구현예에서, 유전 요소는 패키징 신호, 예를 들어, 캡시드 단백질에 결합하는 서열을 포함한다. 일부 구현예에서, 패키징 또는 캡시드-결합 서열의 외측에서, 유전 요소는 야생형 아넬로바이러스 핵산 서열에 대하여 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% 미만의 서열 동일성을 가지며, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 핵산 서열에 대하여 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% 미만의 서열 동일성을 갖는다. 일부 구현예에서, 패키징 또는 캡시드-결합 서열의 외측에서, 유전 요소는 아넬로바이러스 핵산 서열과 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. 특정 구현예에서, 유전 요소는 프로모터 서열, 치료

용 이펙터 및 캡시드 결합 단백질을 인코딩하는 서열을 포함하는 환형, 단일 가닥 DNA이다.

[0655] 일부 구현예에서, 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, 10 kb, 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 구현예에서, 유전 요소는 독립적으로 또는 부가적으로 1000 b 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 또는 그 초과)의 길이를 갖는다. 일부 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 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 구현예에서, 유전 요소는 약 3.6 내지 3.9 kb의 길이를 갖는다. 일부 구현예에서, 유전 요소는 약 2.8 내지 2.9 kb의 길이를 갖는다. 일부 구현예에서, 유전 요소는 약 2.0 내지 3.2 kb의 길이를 갖는다.

[0656] 일부 구현예에서, 유전 요소는 본 명세서에 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.

[0657] 구현예에서, 유전 요소를 (예를 들어, 시험관 내 환화에 의해 생성되는) 이중-가닥 환형 DNA로부터 생성하였다. 일부 구현예에서, 유전 요소를 이중-가닥 환형 DNA로부터 회전환 복제에 의해 생성하였다. 구현예에서, 회전환 복제는 세포(예를 들어, 숙주 세포, 예를 들어, 포유동물 세포, 예를 들어, 인간 세포, 예를 들어, HEK293T 세포, A549 세포 또는 Jurkat 세포)에서 발생한다. 구현예에서, 유전 요소는 세포에서 회전환 복제에 의해 지수적으로 증폭될 수 있다. 구현예에서, 유전 요소는 세포에서 회전환 복제에 의해 선형적으로 증폭될 수 있다. 구현예에서, 이중-가닥 환형 DNA 또는 유전 요소는 세포에서 회전환 복제에 의해 원래의 양의 적어도 2, 4, 8, 16, 32, 64, 128, 256, 518, 1024배 이상을 제공할 수 있다. 구현예에서, 이중-가닥 환형 DNA를, 예를 들어, 본 명세서에 기재된 바와 같이, 세포 내로 도입하였다.

[0658] 일부 구현예에서, 이중-가닥 환형 DNA 및/또는 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 구현예에서, 이중-가닥 환형 DNA 및/또는 유전 요소는 박테리아 플라스미드 백본을 포함하지 않는다.

[0659] 일 구현예에서, 본 발명은 (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, DNA 서열)을 포함하는 유전 요소를 포함한다. 이러한 구현예에서, 유전 요소는 고유 바이러스 서열(예를 들어, 본 명세서에 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.

[0660] *단백질 결합 서열*

[0661] 많은 바이러스에 의해 사용되는 전략은 바이러스 캡시드 단백질이 이의 게놈 내의 특정 단백질 결합 서열을 인식하는 것이다. 예를 들어, 비세그먼트화된 게놈을 갖는 바이러스, 예컨대 효모의 L-A 바이러스에서, 게놈의 5' 말단에 이차 구조(스템-루프) 및 특정 서열이 존재하며, 이는 둘 다 바이러스 캡시드 단백질에 결합하는 데 사용된다. 그러나, 세그먼트화된 게놈을 갖는 바이러스, 예컨대 레오비리대(Reoviridae), 오르토믹소비리대(Orthomyxoviridae)(인플루엔자), 부니아비리대(Bunyaviruse) 및 아레나바이러스(Arenaviruse)는 게놈 세그먼트의 각각을 패키징할 필요가 있다. 일부 바이러스는 세그먼트의 상보성 영역을 이용하여, 바이러스가 각각의 게놈 분자 중 하나를 포함시키는 것을 보조한다. 다른 바이러스는 상이한 세그먼트의 각각에 대하여 특이적인 결합 부위를 갖는다. 예를 들어, 문헌[Curr Opin Struct Biol. 2010 Feb; 20(1): 114-120]; 및 문헌[Journal

of Virology (2003), 77(24), 13036-13041]을 참조한다.

[0662] 일부 구현예에서, 유전 요소는 실질적으로 비-병원성인 단백질에 결합하는 단백질 결합 서열을 인코딩한다. 일부 구현예에서, 단백질 결합 서열은 단백질성 외부로의 유전 요소의 패키징을 용이하게 한다. 일부 구현예에서, 단백질 결합 서열은 실질적으로 비-병원성인 단백질의 아르기닌-풍부 영역에 특이적으로 결합한다. 일부 구현예에서, 유전 요소는 PCT/US19/65995의 실시예 8에 기재된 바와 같은 단백질 결합 서열을 포함한다.

[0663] 일부 구현예에서, 유전 요소는, 예를 들어, 본 명세서에 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.

[0664] 구현예에서, 단백질 결합 서열은, 예를 들어, 본 명세서에 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.

[0665] *5' UTR 영역*

[0666] 일부 구현예에서, 본 명세서에 기재된 바와 같은 핵산 분자(예를 들어, 유전 요소, 유전 요소 작제물, 또는 유전 요소 영역)는 5' UTR 서열, 예를 들어, 본 명세서에 기재된 바와 같은 5' UTR 보존된 도메인 서열(예를 들어, 표 A1, B1, 또는 C1의 임의의 것), 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다.

[0667] 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX_1CAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX_1CAGTCT, 또는 이에 대하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실, 또는 추가)를 갖는 핵산 서열을 포함한다. 구현예에서, X_1은 A이다. 구현예에서, X_1은 존재하지 않는다.

[0668] 일부 구현예에서, 5' UTR 서열은 알파토르크바이러스(예를 들어, Ring1)의 5' UTR의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 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 구현예에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95.775% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 A1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, 또는 이에 대하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실, 또는 추가)를 갖는 핵산 서열을 포함한다.

[0669] 일부 구현예에서, 5' UTR 서열은 베타토르크바이러스(예를 들어, Ring2)의 5' UTR의 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. 구현예에서, 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 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 85% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 87% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 87.324% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 88% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 88.732% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 91% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 91.549% 서열 동일성을 갖는 핵

산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 92% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 92.958% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 94% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 94.366% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 95.775% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 B1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, 또는 이에 대하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실, 또는 추가)를 갖는 핵산 서열을 포함한다.

[0670] 일부 구현예에서, 5' UTR 서열은 감마토르크바이러스(예를 들어, Ring4)의 5' UTR의 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 서열을 포함한다. 구현예에서, 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 구현예에서, 핵산 분자는 표 C1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 핵산 분자는 표 C1에 열거된 5' UTR 보존된 도메인에 대하여 적어도 97.183% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 또는 이에 대하여 적어도 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함한다. 일부 구현예에서, 5' UTR 서열은 핵산 서열 AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, 또는 이에 대하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 이하의 뉴클레오티드 차이(예를 들어, 치환, 결실, 또는 추가)를 갖는 핵산 서열을 포함한다.

[0671] 일부 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 아넬로바이러스 5' UTR 서열, 예를 들어, 표 38에 나타낸 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 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 컨센서스 5' UTR 서열의 핵산 서열을 포함하며, X_1, X_2, X_3, X_4, 및 X_5는 각각 독립적으로, 임의의 뉴클레오티드이며, 예를 들어, X_1 = G 또는 T이며, X_2 = C 또는 A이며, X_3 = G 또는 A이며, X_4 = T 또는 C이며, X_5 = A, C 또는 T이다. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 컨센서스 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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.

[0672] 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 알파토르크바이러스 컨센

서스 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 38에 나타낸 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.

[0673] **[표 38] 아넬로바이러스 유래의 예시적인 5' UTR 서열**

공급원	서열	서열번호
컨센서스	CGGGTGCCGX$_1$AGGTGAGTTTACACACCGX$_2$AGT CAAGGGGCAATTCGGGCTCX$_3$GGACTGGCCGGG CX$_4$X$_5$TGGG X$_1$ = G 또는 T X$_2$ = C 또는 A X$_3$ = G 또는 A X$_4$ = T 또는 C X$_5$ = 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
TTV-TJN02	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	110
TTV-tth8	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	111
알파토르크바이러스 컨센서스 5' UTR	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGC X$_1$X$_2$TGGG; 여기서 X$_1$은 T 또는 C를 포함하고, X$_2$는 A, C, 또는 T를 포함함.	112
알파토르크바이러스 클레이드 1 5' UTR (예: TTV-CT30F)	CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	113
알파토르크바이러스 클레이드 2 5' UTR (예: TTV-P13-1)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CGGG	114
알파토르크바이러스 클레이드 3 5' UTR (예: TTV-tth8)	CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG	115
알파토르크바이러스 클레이드 4 5' UTR (예: TTV-HD20a)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGAGGCCGGGCCAT GGG	116
알파토르크바이러스 클레이드 5 5' UTR (예: TTV-16)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CCGGG	117
알파토르크바이러스 클레이드 6 5' UTR (예: TTV-TJN02)	CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	118
알파토르크바이러스 클레이드 7 5' UTR (예: TTV-HD16d)	CGGGTGCCGAAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG	119

[0674]

[0675] *5' UTR 서열의 확인*

[0676] 일부 구현예에서, 아넬로바이러스 5' UTR 서열은 아넬로바이러스의 게놈(예를 들어, 핵산 서열결정 기법, 예를 들어, 심층 서열결정 기법에 의해 확인된, 예를 들어, 추정의 아넬로바이러스 게놈) 내에서 확인될 수 있다. 일부 구현예에서, 아넬로바이러스 5' UTR 서열은 하기 단계 중 하나 또는 둘 다에 의해 확인된다:

[0677] *(i) 환화 접합점의 확인:* 일부 구현예에서, 5' UTR은 전장의 환화된 아넬로바이러스 게놈의 환화 접합점 근처에 위치할 것이다. 예를 들어, 서열의 중첩 영역을 확인함으로써 환화 접합점을 확인할 수 있다. 일부 구현예에서, 서열의 중첩 영역은 환화된 전장 아넬로바이러스 게놈 서열을 생성하기 위해 서열로부터 트리밍될 수 있다. 일부 구현예에서, 게놈 서열은 소프트웨어를 사용하여 이러한 방식으로 환화된다. 이론에 구속되고자 하지 않지만, 게놈을 컴퓨터로 환화하면 서열의 시작 위치가 비-생물학적으로 배향될 수 있다. 시퀀스 내의 표지점을 사용하여 서열을 올바른 방향으로 재배향할 수 있다. 예를 들어, 표지점 서열은 본 명세서에 기재된 바와 같은 아넬로바이러스 게놈 내의 하나 이상의 요소(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이

러스의 TATA 박스, cap 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, 또는 GC-풍부 영역 중 하나 이상)에 실질적인 상동성을 갖는 서열을 포함할 수 있다.

[0678] *(ii) 5' UTR 서열의 확인:* 일단 추정 아넬로바이러스 게놈 서열이 얻어지면, 서열(또는 이의 부분, 예를 들어, 약 40 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 또는 90 내지 100개 뉴클레오티드의 길이를 가짐)을 하나 이상의 아넬로바이러스 5' UTR 서열(예를 들어, 본 명세서에 기재된 바와 같음)과 비교하여 이에 대한 실질적인 상동성을 갖는 서열을 확인할 수 있다. 일부 구현예에서, 추정 아넬로바이러스 5' UTR 영역은 본 명세서에 기재된 바와 같은 아넬로바이러스 5' UTR에 대하여 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는다.

[0679] *GC-풍부 영역*

[0680] 일부 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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.

[0681] 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 바와 같은 36-뉴클레오티드 GC-풍부 서열(예를 들어, 36-뉴클레오티드 컨센서스 GC-풍부 영역 서열 1, 36-뉴클레오티드 컨센서스 GC-풍부 영역 서열 2, TTV 클레이드 1 36-뉴클레오티드 영역, TTV 클레이드 3 36-뉴클레오티드 영역, TTV 클레이드 3 분리주 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 바와 같은 36-뉴클레오티드 GC-풍부 서열(예를 들어, 36-뉴클레오티드 컨센서스 GC-풍부 영역 서열 1, 36-뉴클레오티드 컨센서스 GC-풍부 영역 서열 2, TTV 클레이드 1 36-뉴클레오티드 영역, TTV 클레이드 3 36-뉴클레오티드 영역, TTV 클레이드 3 분리주 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 인접 뉴클레오티드를 포함하는 핵산 서열을 포함한다.

[0682] 구현예에서, 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. 구현예에서, 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 인접 뉴클레오티드를 포함하는 핵산 서열을 포함한다.

[0683] 구현예에서, 36-뉴클레오티드 GC-풍부 서열은 하기로부터 선택된다:

[0684] (i) CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC(서열번호 160),

[0685] (ii) GCGCTX$_1$CGCGCGCGCGCCGGGGGGCTGCGCCCCCCC(서열번호 164), 여기서, X$_1$은 T, G 또는 A로부터 선택됨;

[0686] (iii) GCGCTTCGCGCGCCGCCCACTAGGGGGCGTTGCGCG(서열번호 165);

[0687] (iv) GCGCTGCGCGCGCCGCCCAGTAGGGGGCGCAATGCG(서열번호 166);

[0688] (v) GCGCTGCGCGCGCGGCCCCCGGGGGAGGCATTGCCT(서열번호 167);

[0689] (vi) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC(서열번호 168);

[0690] (vii) GCGCTTCGCGCGCGCGCCGGGGGGCTCCGCCCCCCC(서열번호 169);

[0691] (viii) GCGCTTCGCGCGCGCGCCGGGGGGCTGCGCCCCCCC(서열번호 170);

[0692] (ix) GCGCTACGCGCGCGCGCCGGGGGGCTGCGCCCCCCC(서열번호 171); 또는

[0693] (x) GCGCTACGCGCGCGCGCCGGGGGGCTCTGCCCCCCC(서열번호 172).

[0694] 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 핵산 서열 CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC(서열번호 160)를 포함한다.

[0695] 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 컨센서스 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 구현예에서, X_1 내지 X_{34} 중 하나 이상(예를 들어, 이들 전부)는 각각 독립적으로 표 39에 특정된 뉴클레오티드(또는 부재)이다. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 예시적인 TTV GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3 또는 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 TTV-HD23a GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2, 단편 3, 단편 4, 단편 5, 단편 6 또는 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 TTV-JA20 GC-풍부 서열(예를 들어, 전체 서열, 단편 1, 단편 2 또는 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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 이의 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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. 구현예에서, 유전 요소(예를 들어, 유전 요소의 단백질-결합 서열)는 표 39에 나타낸 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.

[0696] [표 39] 아넬로바이러스 유래의 예시적인 GC-풍부 서열

공급원	서열	서열번호
컨센서스	CGGCGGX$_1$GGX$_2$GX$_3$X$_4$X$_5$CGCGCTX$_6$CGCGC GCX$_7$X$_8$X$_9$X$_{10}$CX$_{11}$X$_{12}$X$_{13}$X$_{14}$GGGGX$_{15}$X$_{16}$X$_{17}$X$_{18}$ X$_{19}$X$_{20}$X$_{21}$GCX$_{22}$X$_{23}$X$_{24}$X$_{25}$CCCCCCCX$_{26}$CGCGC ATX$_{27}$X$_{28}$GCX$_{29}$CGGGX$_{30}$CCCCCCCCCX$_{31}$X$_{32}$X $_{33}$GGGGGGCTCCGX$_{34}$CCCCCCGGCCCCCC X$_1$ = G 또는 C X$_2$ = G, C, 또는 부재 X$_3$ = C 또는 부재 X$_4$ = G 또는 C X$_5$ = G 또는 C X$_6$ = T, G, 또는 A	120

[0697]

		X_7 = G 또는 C X_8 = G 또는 부재 X_9 = C 또는 부재 X_{10} = C 또는 부재 X_{11} = G, A, 또는 부재 X_{12} = G 또는 C X_{13} = C 또는 T X_{14} = G 또는 A X_{15} = G 또는 A X_{16} = A, G, T, 또는 부재 X_{17} = G, C, 또는 부재 X_{18} = G, C, 또는 부재 X_{19} = C, A, 또는 부재 X_{20} = C 또는 A X_{21} = T 또는 A X_{22} = G 또는 C X_{23} = G, T, 또는 부재 X_{24} = C 또는 부재 X_{25} = G, C, 또는 부재 X_{26} = G 또는 C X_{27} = G 또는 부재 X_{28} = C 또는 부재 X_{29} = G 또는 A X_{30} = G 또는 T X_{31} = C, T, 또는 부재 X_{32} = G, C, A, 또는 부재 X_{33} = G 또는 C X_{34} = C 또는 부재	
예시적인 TTV 서열	전체 서열	GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCATGCGCGGGKCC CCCCCCCNNCGGGGGGCTCCGCCCCCCGGC CCCCCCCCGTGCTAAACCCACCGCGCATGC GCGACCACGCCCCCGCCGCC	121

[0698]

	단편 1	GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCAT	122
	단편 2	GCGCGGGKCCCCCCCCCNNCGGGGGGCTC CG	123
	단편 3	CCCCCCGGCCCCCCCCCGTGCTAAACCCAC CGCGCATGCGCGACCACGCCCCCGCCGCC	124
TTV-CT30F	전체 서열	GCGGCGG-GGGGGCG-GCCGCG-TTCGCGCGCCGCCCACCAGGGGGTG--CTGCG-CGCCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC--GGGGGGGCTCCGCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCGCCGCC	125
	단편 1	GCGGCGG	126
	단편 2	GGGGGCG	127
	단편 3	GCCGCG	128
	단편 4	TTCGCGCGCCGCCCACCAGGGGGTG	129
	단편 5	CTGCG	130
	단편 6	CGCCCCCCCCCGCGCAT	131
	단편 7	GCGCGGGGCCCCCCCCC	132
	단편 8	GGGGGGGCTCCGCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCGCCGCC	133
TTV-HD23a	전체 서열	CGGCGGCGGCGGCG-CGCGCGCTGCGCGCGCG---CGCCGGGGGGGCGCCAGCG-CCCCCCCCCCCCGCGCAT GCACGGGTCCCCCCCCCCACGGGGGGCTCC G CCCCCCGGCCCCCCCCC	134
	단편 1	CGGCGGCGGCGGCG	135
	단편 2	CGCGCGCTGCGCGCGCG	136
	단편 3	CGCCGGGGGGGCGCCAGCG	137

[0699]

	단편 4	CCCCCCCCCCCGCGCAT	138
	단편 5	GCACGGGTCCCCCCCCCCACGGGGGGCTCCG	139
	단편 6	CCCCCCGGCCCCCCCCC	140
TTV-JA20	전체 서열	CCGTCGGCGGGGGGGCCGCGCGCTGCGCGCGCGGCCC-CCGGGGGAGGCACAGCCTCCCCCCCCCGCGCGCATGCGCGCGGGTCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCCCCC	141
	단편 1	CCGTCGGCGGGGGGGCCGCGCGCTGCGCGCGCGGCCC	142
	단편 2	CCGGGGGAGGCACAGCCTCCCCCCCCCGCGCGCATGCGCGCGGGTCCCCCCCCCTCCGGGGGGCTCCGCCCCCCGGCCCCCCCC	143
TTV-TJN02	전체 서열	CGGCGGCGGCG-CGCGCGCTACGCGCGCG---CGCCGGGGGG----CTGCCGC-CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCC	144
	단편 1	CGGCGGCGGCG	145
	단편 2	CGCGCGCTACGCGCGCG	146
	단편 3	CGCCGGGGGG	147
	단편 4	CTGCCGC	148
	단편 5	CCCCCCCCCGCGCAT	149
	단편 6	GCGCGGGGCCCCCCCCC	150
	단편 7	GCGGGGGGCTCCG	151
	단편 8	CCCCCCGGCCCCCC	152
TTV-tth8	전체 서열	GCCGCCGCGGCGGCGGGGG-GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCG---CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCCCCG	153

[0700]

	단편 1	GCCGCCGCGGCGGCGGGGG	154
	단편 2	GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCG	155
	단편 3	CCCCCCCCCGCGCAT	156
	단편 4	GCGCGGGGCCCCCCCCC	157
	단편 5	GCGGGGGGCTCCG	158
	단편 6	CCCCCCGGCCCCCCCCG	159
	단편 7	CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC	160
	단편 8	CCGCCATCTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAATGGTGG	161
	단편 9	CTTAAGTAGTTGAGGCGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAATGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACACGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACTTCCTTCC	162
추가의 GC-풍부 서열	36-뉴클레오티드 컨센서스 GC-풍부 영역 서열 1	CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC	163
	36-뉴클레오티드 영역 컨센서스 서열 2	GCGCTX$_1$CGCGCGCGCCGGGGGGCTGCGCCCCCCC, 여기서, X$_1$은 T, G, 또는 A로부터 선택됨	164
	TTV 클레이드 1 36-뉴클레오티드 영역	GCGCTTCGCGCGCCGCCCACTAGGGGGCGTTGCGCG	165
	TTV 클레이드 3 36-뉴클레오티드 영역	GCGCTGCGCGCGCCGCCCAGTAGGGGGCGCAATGCG	166

[0701]

	TTV 클레이드 3 단리 GH1 36-뉴클레오티드 영역	GCGCTGCGCGCGCGGCCCCCGGGGGAGGC ATTGCCT	167
	TTV 클레이드 3 sle1932 36-뉴클레오티드 영역	GCGCTGCGCGCGCGCCGGGGGGGCGCC AGCGCCC	168
	TTV 클레이드 4 ctdc002 36-뉴클레오티드 영역	GCGCTTCGCGCGCGCCGGGGGGCTCCGC CCCCCC	169
	TTV 클레이드 5 36-뉴클레오티드 영역	GCGCTTCGCGCGCGCCGGGGGGCTGCGC CCCCCC	170
	TTV 클레이드 6 36-뉴클레오티드 영역	GCGCTACGCGCGCGCCGGGGGGCTGCG CCCCCCC	171
	TTV 클레이드 7 36-뉴클레오티드 영역	GCGCTACGCGCGCGCCGGGGGGCTCTGC CCCCCC	172
추가의 알파토르크바이러스 GC-풍부 영역 서열	TTV-CT30F	GCGGCGGGGGGGCGGCCGCGTTCGCGCGC CGCCCACCAGGGGGTGCTGCGCGCCCCCCC CCGCGCATGCGCGGGGCCCCCCCCCGGGG GGGCTCCGCCCCCCCGGCCCCCCCCCGTGC TAAACCCACCGCGCATGCGCGACCACGCCC CCGCCGCC	801
	TTV-P13-1	CCGAGCGTTAGCGAGGAGTGCGACCCTACC CCCTGGGCCCACTTCTTCGGAGCCGCGCGC TACGCCTTCGGCTGCGCGCGGCACCTCAGA CCCCCGCTCGTGCTGACACGCTTGCGCGTG TCAGACCACTTCGGGCTCGCGGGGGTCGGG	802

[0702]

	TTV-tth8	GCCGCCGCGGCGGCGGGGGGCGGCGCGCT GCGCGCGCCGCCCAGTAGGGGGAGCCATG CGCCCCCCCCCGCGCATGCGCGGGGCCCCC CCCCGCGGGGGGCTCCGCCCCCCGGCCCCC CCCG	803
	TTV-HD20a	CGGCCCAGCGGCGGCGCGCGCTTCGCGC GCGCGCCGGGGGGCTCCGCCCCCCCCCGCG CATGCGCGGGGCCCCCCCCCCGCGGGGGGCT CCGCCCCCCGGTCCCCCCCCG	804
	TTV-16	CGGCCGTGCGGCGGCGCGCGCTTCGCGC GCGCGCCGGGGGCTGCCGCCCCCCCCCCGCG CATGCGCGCGGGGCCCCCCCCCCGCGGGGG GCTCCGCCCCCCGGCCCCCCCCCCCG	805
	TTV-TJN02	CGGCGGCGGCGCGCGCGCTACGCGCGCGC GCCGGGGGGCTGCCGCCCCCCCCCCGCGCA TGCGCGGGGCCCCCCCCCCGCGGGGGGCTCC GCCCCCCGGCCCCCC	806
	TTV-HD16d	GGCGGCGGCGCGCGCGCTACGCGCGCGCG CCGGGGAGCTCTGCCCCCCCCCGCGCATGC GCGCGGGTCCCCCCCCCCGCGGGGGGCTCCG CCCCCCGGTCCCCCCCCCG	807

[0703]

[0704] *이펙터*

[0705] 일부 구현예에서, 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 단백질을 인코딩하는 하나 이상의 서열을 포함할 수 있다. 일부 구현예에서, 기능적 핵산은 비-코딩 RNA이다. 일부 구현예에서, 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, 프룩토스 2,6-비스포스페이트는 포스포프룩토키나제 1을 활성화시키고, 인슐린에 반응하여 당분해 속도를 증가시킨다. 또 다른 예에서, 이펙터는 수용체로의 기질 결합을 저해하고, 이의 활성화를 저해할 수 있으며, 예를 들어, 날트렉손(naltrexone) 및 날록손(naloxone)은 오피오이드 수용체를 활성화시키지 않고 결합하며, 수용체가 오피오이드에 결합하지 못하게 한다. 이펙터 활성은 또한, 단백질 안정성/분해 및/또는 전사물 안정성/분해를 조절하는 것을 포함할 수 있다. 예를 들어, 단백질은 분해를 위해 폴리펩티드 보조-인자인 유비퀴틴에 의해 단백질 상에 표적화되어, 분해에 대해 표시된다. 또 다른 예에서, 이펙터는 효소의 활성 부위를 차단함으로써 효소 활성을 저해하며, 예를 들어, 메토트렉세이트는 천연 기질보다 1000배 더 단단하게 디하이드로폴레이트 환원효소에 결합하고, 뉴클레오티드 염기 합성을 저해하는, 효소 디하이드로폴레이트 환원효소에 대한 보조효소인 테트라하이드로폴레이트의 구조적 유사체이다.

[0706] 일부 구현예에서, 이펙터를 인코딩하는 서열은 유전 요소의 부분이며, 예를 들어, 상기 서열은 본 명세서에 기재된 바와 같이 삽입 부위에서 삽입될 수 있다. 구현예에서, 이펙터를 인코딩하는 서열은 비코딩 영역, 예를 들어, 오픈-리딩 프레임의 3'에, 그리고 유전 요소의 GC-풍부 영역의 5'에 배치된 비코딩 영역에서, TATA 박스의 상류의 5' 비코딩 영역에서, 5' UTR에서, 폴리-A 신호의 하류 또는 GC-풍부 영역의 상류의 3' 비코딩 영역에서 유전 요소 내로 삽입된다. 구현예에서, 이펙터를 인코딩하는 서열은 예를 들어, 본 명세서에 기재된 바와 같이 TTV-tth8 플라스미드의 뉴클레오티드 약 3588에서, 또는 예를 들어, 본 명세서에 기재된 바와 같이 TTMV-LY2 플라스미드의 뉴클레오티드 약 2843에서 유전 요소 내로 삽입된다. 구현예에서, 이펙터를 인코딩하는 서열은 예를 들어, 본 명세서에 기재된 바와 같이 TTV-tth8 플라스미드의 뉴클레오티드 336 내지 3015에 또는 그 내에, 또는

예를 들어, 본 명세서에 기재된 바와 같이 TTV-LY2 플라스미드의 뉴클레오티드 242 내지 2812에 또는 그 내에 유전 요소 내로 삽입된다. 일부 구현예에서, 이펙터를 인코딩하는 서열은 오픈-리딩 프레임(예를 들어, 본 명세서에 기재된 바와 같은 ORF, 예를 들어, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3 및/또는 ORF2t/3)의 부분 또는 전부를 대체한다.

[0707] 일부 구현예에서, 이펙터를 인코딩하는 서열은 100 내지 2000, 100 내지 1000, 100 내지 500, 100 내지 200, 200 내지 2000, 200 내지 1000, 200 내지 500, 500 내지 1000, 500 내지 2000, 또는 1000 내지 2000개 뉴클레오티드를 포함한다. 일부 구현예에서, 이펙터는, 예를 들어, 본 명세서에 기재된 바와 같은 핵산 또는 단백질 페이로드이다.

[0708] *조절 핵산*

[0709] 일부 구현예에서, 이펙터는 조절 핵산이다. 조절 핵산은 내인성 유전자 및/또는 외인성 유전자의 발현을 변경시킨다. 일 구현예에서, 조절 핵산은 숙주 유전자를 표적화한다. 조절 핵산은 내인성 유전자(예를 들어, 본 명세서의 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. 구현예에서, 조절 핵산은 miRNA를 인코딩한다. 일부 구현예에서, 조절 핵산은 야생형 아넬로바이러스에 대해 내인성이다. 일부 구현예에서, 조절 핵산은 야생형 아넬로바이러스에 대해 외인성이다.

[0710] 일부 구현예에서, 조절 핵산은 통상적으로 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.

[0711] 일부 구현예에서, 조절 핵산은 핵산 서열, 예를 들어, 가이드 RNA(gRNA)를 포함한다. 일부 구현예에서, DNA 표적화 모이어티는 가이드 RNA 또는 가이드 RNA를 인코딩하는 핵산을 포함한다. gRNA 짧은 합성 RNA는 불완전한 이펙터 모이어티에 대한 결합에 필요한 "스캐폴드" 서열 및 게놈 표적에 대한 사용자-정의된 약 20개 뉴클레오티드 표적화 서열로 구성될 수 있다. 실제로, 가이드 RNA 서열은 일반적으로 17 내지 24개의 뉴클레오티드(예를 들어, 19, 20 또는 21개의 뉴클레오티드)의 길이를 갖고 표적화된 핵산 서열에 대하여 상보성이 되도록 설계된다. 효과적인 가이드 RNA의 설계에 사용하기 위한 맞춤형 gRNA 생성기 및 알고리즘이 상업적으로 구매 가능하다. 유전자 편집은 또한, 천연 발생 crRNA-tracrRNA 복합체를 모방하고 (뉴클레아제에 결합하기 위한) tracrRNA 및 (편집을 위해 뉴클레아제를 표적화된 서열로 가이드하는) 적어도 하나의 crRNA 둘 다를 함유하는 조작된(합성) 단일 RNA 분자인, 키메라 "단일 가이드 RNA"("sgRNA")를 사용하여 달성되었다. 화학적으로 변형된 sgRNA는 또한 게놈 편집에서 효과적인 것으로 입증되었으며; 예를 들어, 문헌[Hendel et al. (2015) Nature Biotechnol., 985 - 991]을 참조한다.

[0712] 조절 핵산은 특정 DNA 서열(예를 들어, 유전자의 프로모터, 인핸서, 침묵자 또는 억제자에 인접하거나 그 내의 서열)을 인식하는 gRNA를 포함한다.

[0713] 특정 조절 핵산은 RNA 간섭(RNAi)의 생물학적 과정을 통하여 유전자 발현을 저해할 수 있다. RNAi 분자는 통상적으로 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), 마이크로RNA(miRNA), 짧은 헤어핀 RNA(shRNA), 메로듀플렉스(meroduplex), 및 다이서(dicer) 기질(미국 특허 8,084,599, 8,349,809 및 8,513,207)을 포함하지만, 이에 제한되지 않는다.

[0714] 긴 비-코딩 RNA(lncRNA)는 100개 초과의 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.

[0715] 유전 요소는 내인성 유전자 또는 유전자 생성물(예를 들어, mRNA)의 전부 또는 이의 단편에 실질적으로 상보성이거나, 완전히 상보성인 서열을 갖는 조절 핵산을 인코딩할 수 있다. 조절 핵산은 인트론과 엑손 사이의 경계

에서의 서열에 상보성이어서, 특정 유전자의 새로 생성된 핵 RNA 전사물이 전사를 위해 mRNA로 성숙되는 것을 방지할 수 있다. 특정 유전자에 상보성인 조절 핵산은 유전자에 대한 mRNA와 혼성화하고, 이의 번역을 방지할 수 있다. 안티센스 조절 핵산은 DNA, RNA 또는 이의 유도체 또는 하이브리드일 수 있다.

[0716] 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 또는 그 이상의 뉴클레오티드일 수 있다. 표적화된 전사물에 대한 조절 핵산의 동일성의 정도는 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 또는 적어도 95%여야 한다.

[0717] 유전 요소는 조절 핵산, 예를 들어, 표적 유전자의 약 5 내지 약 25개의 인접 뉴클레오티드와 동일한 마이크로 RNA(miRNA) 분자를 인코딩할 수 있다. 일부 구현예에서, 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 도입될 포유동물의 게놈 내의 표적 이외의 임의의 뉴클레오티드 서열에 대하여 높은 동일성 백분율을 갖지 않는다.

[0718] 일부 구현예에서, 조절 핵산은 적어도 하나, 예를 들어, 2, 3, 4, 5, 6개, 또는 그 이상의 miRNA이다. 일부 구현예에서, 유전 요소는, 본 명세서, 예를 들어, 표 40에 기재된 서열에 대하여 상동성인 뉴클레오티드 서열들 또는 서열 중 임의의 하나에 대하여 적어도 약 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99% 또는 100% 뉴클레오티드 서열 동일성의 miRNA를 인코딩하는 서열을 포함한다.

[0719] [표 40] 조절 핵산, 예를 들어, miRNA의 예.

균주의 수탁번호	예시적인 하위서열 뉴클레오티드	Pre_miRNA	서열번호	miRNA_5prime_per_MiRdup	서열번호	miRNA_3prime_per_MiRdup	서열번호
AB008394.1	AB008394_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGU	300	AGUAGCUGACGUCAAGGAUUGAC(5')	395	CAUCCUCGGCGGAAGCUACACAA(3')	490
AB008394.1	AB008394_3579_3657	GCGUACGUCACAAGUCACGUGGAGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGUGUA	301	CAAGUCACGUGGAGGGGACCCG(5')	396	GGCCCCGUCACGUGACUUACCAC(3')	491
AB017613.1	AB017613_3462_3539	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUGAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGUG	302	AAGUAGCUGACGUCAAGGAUUGACG(5')	397	UCAUCCUCGGCGGAAGCUACACAA(3')	492
AB017613.1	AB017613_3566_3644	GCACACGUCAUAAGUCACGUGGUGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGAUUUGUCACGUGUGUA	303	AUAAGUCACGUGGUGGGGACCCG(5')	398	GGCCCCGUCACGUGAUUUGUCAC(3')	493
AB025946.1	AB025946_3534_3600	CUUCCGGGUCAUAGGUCACACCUACGUCACAAGUCACGUGGGGAGGGUUGGCGUAUAGCCCGGAAG	304	UGGGGAGGGUUGGCGUAUAGCCCGGA(3')	399	CCGGGUCAUAGGUCACACCUACGUCAC(5')	494
AB025946.1	AB025946_3730_3798	GCCGGGGGGCUGCCGCCCCCCCCGGGGAAAGGGGGGGGCCCCCCCCGGGGGGGGGUUUGCCCCCCGGC	305	CCCCCCCCGGGGGGGGUUUGCCC(3')	400	GGCUGCCGCCCCCCCGGGGAAAGGGGG(5')	495
AB028668.1	AB028668_3537_3615	AUACGUCAUCAGUCACGUGGGGGAAGGCGUGCCUAAACCCGGAAGCAUCCUCGUCCACGUGACUGUGACGUGUGUGGC	306	AUCAGUCACGUGGGGGAAGGCGUGC(5')	401	AUCCUCGUCCACGUGACUGUGA(3')	496
AB028669.1	AB028669_3440_3513	CAUUUUAAGUAAGGCGGAAGCAGCUCGGCGUACACAAAAUGGCGGCGGAGCACUUCCGGCUUGCCCAAAAUGG	307	AAGUAAGGCGGAAGCAGCUCGG(5')	402	GAGCACUUCCGGCUUGCCCAA(3')	497

[0720]

AB028669.1	AB028669_3548_3619	GUCACAAGUCACGUGGGGAGGGUUGGCGUUUAACCCGGAAGCCAAUCCUCUUACGUGGCCUGUCACGUGAC	308	AGUCACGUGGGGAGGGUUGGC(5')	403	CAAUCCUCUUACGUGGCCUG(3')	498
AB037926.1	AB037926_162_232	CGACCGCGUCCCGAAGGCGGGUACCCGAGGUGAGUUUACACACCGAGGUUAAGGGCCAAUUCGGGCUUGG	309	CCCGAAGGCGGGUACCCGAGGU(5')	404	CGAGGUUAAGGGCCAAUUCGGGCU(3')	499
AB037926.1	AB037926_3454_3513	CGCGGUAUCGUAGCCGACGCGGACCCGUUUUCGGGGCCCCCGCGGGGCUCUCGGCGCG	310	UAUCGUAGCCGACGCGGACCCCG(5')	405	GGGCCCCCGCGGGGCUCUCGGCG(3')	500
AB037926.1	AB037926_3531_3609	CGCCAUUUUGUGAUACGCGCGUCCCCUCCCGGCUUCCGUACAACGUCAGGCGGGGCGUGGCCGUAUCAGAAAAUGGCG	311	AUUUUGUGAUACGCGCGUCCCCUCCC(5')	406	GCGGGGCGUGGCCGUAUCAGAAAAUGG(3')	501
AB037926.1	AB037926_3637_3714	GCUACGUCAUAAGUCACGUGACUGGGCAGGUACUAAACCCGGAAGUAUCCUCGGUCACGUGGCCUGUCACGUAGUUG	312	AAGUCACGUGACUGGGCAGGU(5')	407	CCUCGGUCACGUGGCCUGU(3')	502
AB038621.1	AB038621_3511_3591	GGCUSUGACGUCAAAGUCACGUGGGRAGGGUGGCGUUAAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACAGCC	313	UGACGUCAAAGUCACGUGGGRAGGGU(5')	408	CCUCGUCACGUGACCUGACGUCACAG(3')	503
AB038622.1	AB038622_227_293	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGAAGGU	314	GAUCGAGCGUCCCGUGGGCGGGU(3')	409	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	504
AB038622.1	AB038622_3510_3591	GGUUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUUAAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGGCC	315	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	410	AUCCUCGUCACGUGACCUGACGUCACG(3')	505

[0721]

AB038623.1	AB038623_228_295	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	316	GAUCGAGCGUCCCGUGGGCGGGU(3')	411	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	506
AB038624.1	AB038624_228_295	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	317	GAUCGAGCGUCCCGUGGGCGGGU(3')	412	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	507
AB038624.1	AB038624_3511_3592	GGCUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUUAAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGGCC	318	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	413	AUCCUCGUCACGUGACCUGACGUCACG(3')	508
AB041957.1	AB041957_3414_3493	AGACCACGUGGUAAGUCACGUGGGGGCAGCUGCUGUAAACCCGGAAGUAGCUGACCCGCGUGACUGGUCACGUGACCUG	319	ACGUGGUAAGUCACGUGGGGGCAGCU(5')	414	CUGACCCGCGUGACUGGUCACGUGA(3')	509
AB049608.1	AB049608_3199_3277	CGCCAUUUUAUAAUACGCGCGUCCCCUCCCGGCUUCCGUACUACGUCAGGCGGGGCGUGGCCGUAUUAGAAAAUGGUG	320	AUUUUAUAAUACGCGCGUCCCCUCC(5')	415	CGGGGCGUGGCCGUAUUAGAAAAUGG(3')	510
AB050448.1	AB050448_3393_3465	UAAGUAAGGCGGAACCAGGCUGUCACCCUGUGUCAAAGGUCAAGGGACAGCCUUCCGGCUUGCACAAAAUGG	321	AAGGGACAGCCUUCCGGCUUGC(3')	416	AGUAAGGCGGAACCAGGCUGUCACCCUGU(5')	511
AB054647.1	AB054647_3537_3615	UGCCUACGUCAUAAGUCACGUGGGGACGGCUGCUGUAAACACGGAAGUAGCUGACCCGCGUGACUUGUCACGUGAGCA	322	CAUAAGUCACGUGGGGACGGCUGCU(5')	417	UAGCUGACCCGCGUGACUUGUCAC(3')	512

[0722]

[0723]

AB054648.1	AB054648_3439_3511	UUGUGUAAGGCGGAACAGGCUGACACCCCGUGUCAAAGGUCAGGGGUCAGCCUCCGCUUUGCACCAAAUGGU	323	UAAGGCGGAACAGGCUGACACCCC(5')	418	GGUCAGCCUCCGCUUUGCA(3')	513
AB054648.1	AB054648_3538_3617	UACCUACGUCAUAAGUCACGUGGGAAGAGCUGCUGUGAACCUGGAAGUAGCUGACCCGCGUGGCUUGUCACGUGAGUGC	324	UACGUCAUAAGUCACGUGGGAAGAGCUG(5')	419	GCUGACCCGCGUGGCUUGUCACGUGAGU(3')	514
AB064595.1	AB064595_116_191	UUUUCCUGGCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGGGCGUCCCGAGGGCGGGUGCCGGAGGUG	325	UCGGGCGUCCCGAGGGCGGGUG(3')	420	GGCCCGUCCGCGGCGAGAGCGCGAG(5')	515
AB064595.1	AB064595_3283_3351	AAAGUGAGUGGGGCCAGACUUCGCCAUAGGCCUUUAACUUCCGGGUGCGUCUGGGGGCCGCCAUUUU	326	AAAGUGAGUGGGGCCAGACUUCGCC(5')	421	UCCGGGUGCGUCUGGGGGCCGCCAUUU(3')	516
AB064595.1	AB064595_3427_3500	GUGACGUUACUCUCACGUGAUGGGGGCGUGCUCUAACCCGGAAGCAUCCUCGACCACGUGACUGUGACGUCAC	327	CUCUCACGUGAUGGGGGCGUGC(5')	422	AUCCUCGACCACGUGACUGUG(3')	517
AB064595.1	AB064595_41_116	AGCGUCUACUACGUACACUUCCUGGGGUGUGUCCUGCCACUGUAUAUAAACCAGAGGGGUGACGAAUGGUAGAGU	328	UCUACUACGUACACUUCCUGGGGUGUGU(5')	423	AUAAACCAGAGGGGUGACGAAUGGUAGAGU(3')	518
AB064596.1	AB064596_3424_3497	GUGACGUCAAAGUCACGUGGUGACGGCCAUUUUAACCCGGAAGUGGCUGUUGUCACGUGACUUGACGUCACGG	329	UGGCUGUUGUCACGUGACUUGA(3')	424	CAAAGUCACGUGGUGACGGCCAU(5')	519

AB064597.1	AB064597_3191_3253	GCUUUAGACGCCAUUUUAGGCCCUCGCGGGCACCCGUAGGCGCGUUUUAAUGACGUCACGGC	330	AGACGCCAUUUUAGGCCCUCGCGG(5')	425	GUAGGCGCGUUUUAAUGACGUCACGG(3')	520
AB064597.1	AB064597_3221_3294	CACCCGUAGGCGCGUUUUAAUGACGUCACGGCAGCCAUUUUGUCGUGACGUUUGAGACACGUGAUGGGGGCGU	331	UGUCGUGACGUUUGAGACACGUGAU(3')	426	UAGGCGCGUUUUAAUGACGUCACGGCAG(5')	521
AB064597.1	AB064597_3262_3342	GUCGUGACGUUUGAGACACGUGAUGGGGGCGCCUAAACCCGGAAGCAUCCCUGGUCACGUGACUCUGACGUCACGGCG	332	UGACGUUUGAGACACGUGAUGGGGGCGUGC(5')	427	AUCCCUGGUCACGUGACUCUGACGUCACG(3')	522
AB064598.1	AB064598_3179_3256	CGAAAGUGAGUGGGGCCAGACUUCGCCAUAAGGCCUUUAACUUCCGGGUGCGUGUGGGGGCCGCCAUUUUAGCUUCG	333	AGUGAGUGGGGCCAGACUUCGC(5')	428	GCGUGUGGGGGCCGCCAUUUUAGCUU(3')	523
AB064598.1	AB064598_3323_3399	CUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	334	UGUGACGUCAAAGUCACGUGGGGAGGGCGG(5')	429	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	524
AB064598.1	AB064598_3412_3485	CUGUCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAGAGGAAGUAUGACGUAGCGGCGGGGGGC	335	AAAAGAGGAAGUAUGACGUAGCGGCGG(3')	430	CGCCAUCUUGUGACUUCCUUCCGCUUUUU(5')	525
AB064599.1	AB064599_108_175	GGUAGAGUUUUUUCCGCCCGUCCGCAGCGAGGACGCGAGCGCAGCGAGCGGCCGAGCGACCCGUGGG	336	AGCGAGCGGCCGAGCGACCCG(3')	431	UAGAGUUUUUUCCGCCCGUCCG(5')	526
AB064599.1	AB064599_3389_3469	GCUGUGACGUUUCAGUCACGUGGGGAGGGAACGCCUAAACCCGGAAGCGUCCCUGGUCACGUGAUUGUGACGUCACGGCC	337	UUCAGUCACGUGGGGAGGGAACGC(5')	432	GUCCCUGGUCACGUGAUUGUGAC(3')	527

[0724]

[0725]

AB064599.1	AB064599_3483_3546	CCGCCAUUUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUGUGACGUAGCGGCGG	338	AAAAGAGGAAGUGUGACGUAGCGG(3')	433	CAUUUUGUGACUUCCUUCCGCUUUUU(5')	528
AB064600.1	AB064600_3378_3456	GACUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	339	UGUGACGUCAAAGUCACGUGGGAGGGCGG(5')	434	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	529
AB064600.1	AB064600_3469_3542	CUGUCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUAUGACGUGGCGGCGGGGGGGC	340	AAAAGAGGAAGUAUGACGUGGCGG(3')	435	CCGCCAUCUUGUGACUUCCUUCCGCUUUUU(5')	530
AB064601.1	AB064601_3318_3398	GGUUGUGACGUCAAAGUCACGUGGGGAGGGCGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGGCC	341	UGACGUCAAAGUCACGUGGGGAGGGCGG(5')	436	AUCCUCGUCACGUGACCUGACGUCACG(3')	531
AB064601.1	AB064601_3412_3477	CCCGCCAUCUUGUGACUUCCUUCCGCUUUUUCAAAAAAAAAGAGGAAGUGUGACGUAGCGGCGGG	342	AAAAAAGAGGAAGUGUGACGUAGCGGCGG(3')	437	CGCCAUCUUGUGACUUCCUUCCGCUUUUUC(5')	532
AB064602.1	AB064602_125_192	GCCCGUCCGCGGCGAGAGCGCGAGCGAAGCGAGCGAUCGAGCGUCCCGUGGGCGGGUGCCGUAGGUG	343	GAUCGAGCGUCCCGUGGGCGGGU(3')	438	CCGUCCGCGGCGAGAGCGCGAGCGA(5')	533
AB064602.1	AB064602_3368_3446	GACUGUGACGUCAAAGUCACGUGGGGAGGAGGGCGUGUAACCCGGAAGUCAUCCUCGUCACGUGACCUGACGUCACGG	344	UGUGACGUCAAAGUCACGUGGGAGGAGGG(5')	439	UCAUCCUCGUCACGUGACCUGACGUCACG(3')	534

AB064603.1	AB064603_3385_3447	UCGCGUCUUAGUGACGUCACGGCAGCCAUCUUGGUCCUGACGUCACUGUCACGUGGGGAGGG	345	UUGGUCCUGACGUCACUGUCA(3')	440	CUUAGUGACGUCACGGCAGCCAU(5')	535
AB064603.1	AB064603_3422_3498	UGACGUCACUGUCACGUGGGGAGGGAACACGUGAACCCGGAAGUGUCCCUGGUCACGUGACAUGACGUCACGGCCG	346	CGUCACUGUCACGUGGGGAGGGAACAC(5')	441	GUCCCUGGUCACGUGACAUGACGUC(3')	536
AB064604.1	AB064604_3436_3514	CGCCAUUUUAAGUAAGCAUGGCGGGCGGUGAUGUCAAAUGUUAAAGGUCACAGCCGGUCAUGCUUGCACAAAAUGGCG	347	UAAGUAAGCAUGGCGGGCGGUGAU(5')	442	CACAGCCGGUCAUGCUUGCACAAA(3')	537
AB064605.1	AB064605_3440_3518	CGCCAUUUUAAGUAAGCAUGGCGGGCGGUGACGUGCAAUGUCAAAGGUCACAGCCUGUCAUGCUUGCACAAAAUGGCG	348	AAGUAAGCAUGGCGGGCGGUGA(5')	443	ACAGCCUGUCAUGCUUGCACAA(3')	538
AB064606.1	AB064606_3377_3449	CCAUCUUAAGUAGUUGAGGCGGACGGUGGCGUCGGUUCAAAGGUCACCAUCAGCCACACCUACUCAAAAUGG	349	UAAGUAGUUGAGGCGGACGGUGGC(5')	444	CACCAUCAGCCACACCUACUCAAA(3')	539
AB064607.1	AB064607_3502_3569	GCCUGUCAUGCUUGCACAAAAUGGCGACUUCCGCUUCCGGGUCGCCGCCAUAUUUGGUCACGUGAC	350	UCAUGCUUGCACAAAAUGGCGGACUUCCG(5')	445	CGGGUCGCCGCCAUAUUUGGUCACGUGA(3')	540
AF079173.1	AF079173_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGU	351	AGUAGCUGACGUCAAGGAUUGAC(5')	446	CAUCCUCGGCGGAAGCUACACAA(3')	541
AF116842.1	AF116842_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGU	352	AGUAGCUGACGUCAAGGAUUGAC(5')	447	CAUCCUCGGCGGAAGCUACACAA(3')	542

[0726]

AF116842.1	AF116842_3579_3657	GCAUACGUCACAAGUCACGUGGGGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGUGUA	353	ACAAGUCACGUGGGGGGGACCCG(5')	448	GGCCCCGUCACGUGACUUACCAC(3')	543
AF122913.1	AF122913_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUGAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGU	354	AAGUAGCUGACGUCAAGGAUUGACG(5')	449	UCAUCCUCGGCGGAAGCUACACAA(3')	544
AF122913.1	AF122913_3579_3657	GCACACGUCAUAAGUCACGUGGUGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGAUUUGUCACGUGUGUA	355	AUAAGUCACGUGGUGGGGACCCG(5')	450	GGCCCCGUCACGUGAUUUGUCAC(3')	545
AF122914.1	AF122914_3476_3552	GCCAUUUUAAGUCAGCUCUGGGGAGGCGCGUGACUUCCAGUUCAAAGGUCAUCCUCACCAUAACUGGCACAAAAUGGC	356	AAGUCAGCUCUGGGGAGGCGUGACUU(5')	451	GUCAUCCUCACCAUAACUGGCACAA(3')	546
AF122915.1	AF122915_3475_3551	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAAAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGU	357	AGUAGCUGACGUCAAGGAUUGAC(5')	452	CAUCCUCGGCGGAAGCUACACAA(3')	547
AF122915.1	AF122915_3579_3657	GCAUACGUCACAAGUCACGUGGAGGGGACACGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGUGUA	358	CAAGUCACGUGGAGGGGACACG(5')	453	GGCCCCGUCACGUGACUUACCAC(3')	548
AF122916.1	AF122916_3458_3537	GCGCCAUGUUAAGUGGCUGUCGCCGAGGAUUGACGUCACAGUUCAAAGGUCAUCCUCGACGGUAACCGCAAACAUGGCG	359	UGUUAAGUGGCUGUCGCCGAGGAUUGA(5')	454	AUCCUCGACGGUAACCGCAAACAUG(3')	549

[0727]

AF122916.1	AF122916_3565_3641	CAUGCGUCAUAAGUCACAUGACAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	360	UAAGUCACAUGACAGGGGUCCA(5')	455	GGCCCCGACAUGUGACUCGUC(3')	550
AF122916.1	AF122916_91_164	UGGCAGCACUUCCGAAUGGCUGAGUUUUCCACGCCCGUCCGCGGAGAGGGAGCCACGGAGGUGAUCCCGAACG	361	CGGAGAGGGAGCCACGGAGGUG(3')	456	AGCACUUCCGAAUGGCUGAGUUUUCCA(5')	551
AF122917.1	AF122917_3369_3447	GCCAUUUUAAGUCAGCGCUGGGGAGGCAUGACUGUAAGUUCAAAGGUCAUCCUCACCGGAACUGACACAAAAUGGCCG	362	AAGUCAGCGCUGGGGAGGCAUGA(5')	457	AUCCUCACCGGAACUGACACAA(3')	552
AF122918.1	AF122918_3460_3540	GCCAUCUUAAGUGGCUGUCGCCGAGGAUUGACGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGUC	363	UCUUAAGUGGCUGUCGCCGAGGAUUGAC(5')	458	CAUCCUCGGCGGUAACCGCAAAGAUG(3')	553
AF122918.1	AF122918_3566_3642	AUACGUCAUAAGUCACAUGUCUAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	364	AAGUCACAUGUCUAGGGGUCCACU(5')	459	UAGGCCCCGACAUGUGACUCGU(3')	554
AF122919.1	AF122919_3370_3447	CCAUUUUAAGUAAGGCGGAAGCAGCUGUCCCUGUAACAAAAUGGCGGCGACAGCCUUCCGCUUUGCACAAAAUGGAG	365	AAGUAAGGCGGAAGCAGCUGUCC(5')	460	ACAGCCUUCCGCUUUGCACAA(3')	555
AF122920.1	AF122920_3460_3540	GCCAUCUUAAGUGGCUGUCGCUGAGGAUUGACGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGUC	366	AUCUUAAGUGGCUGUCGCUGAGGAUUGAC(5')	461	CAUCCUCGGCGGUAACCGCAAAGAUGG(3')	556

[0728]

[0729]

AF122920.1	AF122920_3565_3641	CAUACGUCAUAAGUCACAUGACAGGAGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	367	UAAGUCACAUGACAGGAGUCCACU(5')	462	UAGGCCCCGACAUGUGACUCGUC(3')	557
AF122921.1	AF122921_3459_3540	CGCCAUCUUAAGUGGCUGUCGCCGAGGAUUGGCGUCACAGUUCAAAGGUCAUCCUCGGCGGUAACCGCAAAGAUGGCGGU	368	AAGUGGCUGUCGCCGAGGAUUG(5')	463	UCCUCGGCGGUAACCGCAAA(3')	558
AF122921.1	AF122921_3565_3641	CAUACGUCAUAAGUCACAUGACAGGGGUCCACUUAAACACGGAAGUAGGCCCCGACAUGUGACUCGUCACGUGUGU	369	UAAGUCACAUGACAGGGGUCCA(5')	464	GGCCCCGACAUGUGACUCGUC(3')	559
AF129887.1	AF129887_3579_3657	GCAUACGUCACAAGUCACGUGGGGGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGGUGU	370	ACAAGUCACGUGGGGGGGACCCG(5')	465	GGCCCCGUCACGUGACUUACCAC(3')	560
AF247137.1	AF247137_3453_3530	CCGCCAUUUUAGGCUGUUGCCGGGCGUUUGACUUCCGUGUUAAAGGUCAAACACCCAGCGACACCAAAAAAUGGCCG	371	AUUUUAGGCUGUUGCCGGGCGUUUGACU(5')	466	UCAAACACCCAGCGACACCAAAAAAUGG(3')	561
AF247137.1	AF247137_3559_3636	CUACGUCAUAAGUCACGUGACAGGGAGGGGGCGACAAACCCGGAAGUCAUCCUCGCCCACGUGACUUACCACGUGGUG	372	AUAAGUCACGUGACAGGGAGGGG(5')	467	CCUCGCCCACGUGACUUACCAC(3')	562
AF247138.1	AF247138_3455_3532	GCCAUUUUAAGUAGGUGACGUCCAGGACUGACGUAAAGUUCAAAGGUCAUCCUCGGCGGAACCUAUACAAAAUGGCG	373	AAGUAGGUGACGUCCAGGACU(5')	468	CCUCGGCGGAACCUAUACAA(3')	563
AF247138.1	AF247138_3561_3637	CUACGUCAUAAGUCACGUGGGGACGGCUGUACUUAAACACGGAAGUAGGCCCCGUCACGUGAUUUACCACGUGGUG	374	CAUAAGUCACGUGGGGACGGCUGU(5')	469	GCCCCGUCACGUGAUUUACCAC(3')	564

[0730]

AF261761.1	AF261761_3431_3504	GCCAUUUUAAGUAAGGCGGAAGAGCUCUAGCUAUACAAAAUGGCGGCGGAGCACUUCCGCUUUGCCCAAAAUG	375	UAAGUAAGGCGGAAGAGCUCUAGCUA(5')	470	GCGGCGGAGCACUUCCGCUUUGCCCAAA(3')	565
AF351132.1	AF351132_3475_3552	GCCAUUUUAAGUAGCUGACGUCAAGGAUUGACGUAGAGGUUAAAGGUCAUCCUCGGCGGAAGCUACACAAAAUGGUG	376	AGUAGCUGACGUCAAGGAUUGAC(5')	471	CAUCCUCGGCGGAAGCUACACAA(3')	566
AF351132.1	AF351132_3579_3657	GCAUACGUCACAAGUCACGUGGGGGGGGACCCGCUGUAACCCGGAAGUAGGCCCCGUCACGUGACUUACCACGUGUGUA	377	ACAAGUCACGUGGGGGGGACCCG(5')	472	GGCCCCGUCACGUGACUUACCAC(3')	567
AF435014.1	AF435014_3344_3426	GGCGCCAUUUUAAGUAAGCAUGGCGGGCGGCGACGUCACAUGUCAAAGGUCACCGCACUUCCGUGCUUGCACAAAAUGGC	378	UAAGUAAGCAUGGCGGGCGGCGAC(5')	473	CACCGCACUUCCGUGCUUGCACAAA(3')	568
AF435014.1	AF435014_3453_3526	UGCUACGUCAUCGAGACACGUGGUGCCAGCAGCUGUAAACCCGGAAGUCGCUGACACACGUGUCUUGUCACGU	379	AUCGAGACACGUGGUGCCAGCAGCU(5')	474	UCGCUGACACACGUGUCUUGUCAC(3')	569
AJ620212.1	AJ620212_3360_3438	GCCAUUUUAAGUAGCACCGCCUAGGGAUGACGUAUAAGUUCAAAGGUCAUCCUCAGCCGGAACUUACACAAAAUGGU	380	UCAUCCUCAGCCGGAACUUACACAAAAUGG(3')	475	CAUUUUAAGUAAGCACCGCCUAGGGAUGAC(5')	570
AJ620212.1	AJ620212_3470_3542	ACGUCAUAUGUCACGUGGGGAGGCCCUGCUGCGCAAACGCGGAAGUAGGCCCCGUCACGUGUCAUACCACGU	381	AUAUGUCACGUGGGGAGGCCCUGCUG(5')	476	GUAGGCCCCGUCACGUGUCAUACCAC(3')	571

AJ620218.1	AJ620218_3381_3458	CCAUUUUAAGUAAGGCGGAAGCAGCUCCACUUUCUCACAAAAUGGCGGCGGGGCACUUCCGGCUUGCCCAAAAUGGC	382	AAGUAAGGCGGAAGCAGCUCCACUUU(5')	477	GGCGGGGCACUUCCGGCUUGCCCAA(3')	572
AJ620226.1	AJ620226_3451_3523	CCAUUUUAAGUAAGGCGGAAGUUUCUCCACUAUACAAAAUGGCGGCGGAGCACUUCCGGCUUGCCCAAAAUG	383	AAGUAAGGCGGAAGUUUCUCCACU(5')	478	CGGCGGAGCACUUCCGGCUUGCCCAA(3')	573
AJ620227.1	AJ620227_3379_3451	CCAUCUUAAGUAGUUGAGGCGGACGGUGGCGUGAGUUCAAGGUCACCAUCAGCCACACCUACUCAAAAUGG	384	UAAGUAGUUGAGGCGGACGGUGGC(5')	479	CACCAUCAGCCACACCUACUCAAA(3')	574
AJ620231.1	AJ620231_3429_3505	CGCCAUCUUAAGUAGUUGAGGCGGACGGGUGGCGUGAGUUCAAAGGUCACCAUCAGCCACACCUACUCUCAAAAUGGUG	385	UAAGUAGUUGAGGCGGACGGUGG(5')	480	ACCAUCAGCCACACCUACUCAAA(3')	575
AY666122.1	AY666122_3163_3236	UUUCGGACCUUCGGCGUCGGGGGGGUCGGGGGCUUUACUAAACAGACUCCGAGAUGCCAUUGGACACUGAGGG	386	GACCUUCGGCGUCGGGGGGGUCGGGGGG(5')	481	GACUCCGAGAUGCCAUUGGACACUGAGG(3')	576
AY666122.1	AY666122_3388_3464	CCAUUUUAAGUAGGUGCCGUCCAGCACUGCUGUUCCGGGUUAAAGGGCAUCCUCGGCGGAACCUAUACAAAAUGGC	387	AUCCUCGGCGGAACCUAUA(3')	482	AGUAGGUGCCGUCCAGCA(5')	577
AY666122.1	AY666122_3494_3567	CUACGUCAUCGAUGACGUGGGGAGGCGUACUAUGAAACGCGGAAGUAGGCCCCGCUACGUCAUCAUCACGUGG	388	AUCGAUGACGUGGGGAGGCGUACUAU(5')	483	AAGUAGGCCCCGCUACGUCAUCAC(3')	578
AY823988.1	AY823988_3452_3525	CCAUUUUAAGUAAGGCGGAAGAGCUGCUCUAUAUACAAAAUGGCGGAGGAGCACUUCCGGCUUGCCCAAAAUG	389	UGGCGGAGGAGCACUUCCGGCUUG(3')	484	AAGGCGGAAGAGCUGCUCUAUAU(5')	579

[0731]

AY823988.1	AY823988_3554_3629	UGCCUACGUAACAAGUCACGUGGGGAGGGUUGGCGUAUAACCCGGAAGUCAAUCCUCCCACGUGGCCUGUCACGU	390	AACAAGUCACGUGGGGAGGGUUGGC(5')	485	CAAUCCUCCCACGUGGCCUGUCAC(3')	580
AY823989.1	AY823989_3551_3623	UAAGUAAGGCGGAACCAGGCUGUCACCCCGUGUCAAAGGUCAGGGGUCAGCCUUCCGCUUUACACAAAAUGG	391	AGGGGUCAGCCUUCCGCUUUA(3')	486	AAGGCGGAACCAGGCUGUCACCCCGU(5')	581
AY823989.1	AY823989_3551_3623	UAAGUAAGGCGGAACCAGGCUGUCACCCCGUGUCAAAGGUCAGGGGUCAGCCUUCCGCUUUACACAAAAUGG	392	AGGGGUCAGCCUUCCGCUUUA(3')	487	AAGGCGGAACCAGGCUGUCACCCCGU(5')	582
DQ361268.1	DQ361268_3413_3494	GCAGCCAUUUUAAGUCAGCUUCGGGGAGGGUCACGCAAAGUUCAAAGGUCAUCCUCACCGGAACUGGUACAAAAUGGCCG	393	UAAGUCAGCUUCGGGGAGGGUCAC(5')	488	CAUCCUCACCGGAACUGGUACAAA(3')	583
DQ361268.1	DQ361268_3519_3593	UGCUACGUCAUAAGUGACGUAGCUGGUGUCUGCUGUAAACACGGAAGUAGGCCCCGCCACGUCACUUGUCACGU	394	UCAUAAGUGACGUAGCUGGUGUCUGCU(5')	489	UAGGCCCCGCCACGUCACUUGUCACG(3')	584

[0732]

[0733] siRNA 및 shRNA는 내인성 마이크로RNA(miRNA) 유전자의 가공 경로 내의 중간체와 유사하다(문헌[Bartel, Cell 116:281-297, 2004]). 일부 구현예에서, 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 배출물을 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 시드 영역으로 공지되어 있다. siRNA 및 miRNA가 상호 교환 가능하기 때문에, 외인성 siRNA는 siRNA에 대하여 시드 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]).

[0734] 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 결합 부위는 관련 문헌에서 잘 나타나 있다. RNAi 분자는 당업계에 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]).

[0735] 조절 핵산은 유전자에 의해 인코딩된 RNA의 발현을 조절할 수 있다. 다수의 유전자가 서로 어느 정도의 서열 상동성을 공유할 수 있기 때문에, 일부 구현예에서, 조절 핵산은 충분한 서열 상동성을 갖는 유전자의 부류를 표적화하도록 설계될 수 있다. 일부 구현예에서, 조절 핵산은 상이한 유전자 표적 간에 공유되는 또는 특정 유전자 표적에 특유한 서열에 상보성을 갖는 서열을 함유할 수 있다. 일부 구현예에서, 조절 핵산은 몇몇의 유전자 간에 상동성을 갖는 RNA 서열의 보존된 영역을 표적화하여, 그에 의해, 유전자 패밀리 내의 몇몇의 유전자(예를 들어, 상이한 유전자 아이소폼, 슬라이스 변이체, 돌연변이 유전자 등)를 표적화하도록 설계될 수 있다. 일부 구현예에서, 조절 핵산은 단일 유전자의 특정 RNA 서열에 특유한 서열을 표적화하도록 설계될 수 있다.

[0736] 일부 구현예에서, 유전 요소는 하나 이상의 유전자의 발현을 조절하는 조절 핵산을 인코딩하는 하나 이상의 서

열을 포함할 수 있다.

[0737] 일 구현예에서, 본 명세서의 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 17개 뉴클레오티드의 gRNA 서열은 일반적으로 Cas9-매개된 DNA 절단이 발생하게 하며; Cpf1에 있어서, 적어도 약 16개 뉴클레오티드의 gRNA 서열이 검출 가능한 DNA 절단을 달성하는 데 필요하다.

[0738] *치료용 이펙터(예를 들어, 펩티드 또는 폴리펩티드)*

[0739] 일부 구현예에서, 유전 요소는 치료용 발현 서열, 예를 들어, 치료용 펩티드 또는 폴리펩티드, 예를 들어, 세포내 펩티드 또는 세포내 폴리펩티드, 분비 폴리펩티드, 또는 단백질 대체 치료제를 인코딩하는 서열을 포함한다. 일부 구현예에서, 유전 요소는 단백질, 예를 들어, 치료용 단백질을 인코딩하는 서열을 포함한다. 치료용 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 이의 구성성분을 포함할 수 있지만, 이에 제한되지 않는다.

[0740] 일부 구현예에서, 유전 요소는 펩티드, 예를 들어, 치료용 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 이들 사이의 임의의 범위의 길이를 갖는다.

[0741] 일부 구현예에서, 치료용 발현 서열에 의해 인코딩된 폴리펩티드는 상기 중 임의의 것의 기능적 변이체 또는 이의 단편, 예를 들어, UniProt ID를 참조하여 본 명세서의 표에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질일 수 있다.

[0742] 일부 구현예에서, 치료용 발현 서열은 상기 중 임의의 것에 결합하는 항체 또는 항체 단편, 예를 들어, UniProt ID를 참조하여 본 명세서의 표에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질에 대한 항체를 인코딩할 수 있다. 본 명세서에서 용어 "항체"는 가장 넓은 의미로 사용되며, 요망되는 항원-결합 활성을 나타내는 한 모노클로날 항체, 폴리클로날 항체, 다중특이적 항체(예를 들어, 이중특이적 항체), 및 항체 단편을 포함하지만 이에 제한되지 않는 다양한 항체 구조를 포함한다. "항체 단편"은 적어도 하나의 중쇄 또는 경쇄를 포함하고 항원에 결합하는 분자를 지칭한다. 항체 단편의 예는 Fv, Fab, Fab', Fab'-SH, $F(ab')_2$; 디아바디; 선형 항체; 단쇄 항체 분자(예를 들어, scFv); 및 항체 단편으로부터 형성된 다중특이적 항체를 포함하지만 이에 제한되지 않는다.

[0743] **예시적인 세포내 폴리펩티드 이펙터**

[0744] 일부 구현예에서, 이펙터는 세포질 폴리펩티드 또는 세포질 펩티드를 포함한다. 일부 구현예에서, 이펙터는 DPP-4 저해제, GLP-1 신호전달의 활성화제, 또는 호중구 엘라스타제의 저해제인 세포질 펩티드를 포함한다. 일부 구현예에서, 이펙터는 성장 인자 또는 이의 수용체(예를 들어, FGF 수용체, 예를 들어, FGFR3)의 수준 또는 활성을 증가시킨다. 일부 구현예에서, 이펙터는 n-myc 상호작용 단백질 활성의 저해제(예를 들어, n-myc 상호작용 단백질 저해제); EGFR 활성의 저해제(예를 들어, EGFR 저해제); IDH1 및/또는 IDH2 활성의 저해제(예를 들어, IDH1 저해제 및/또는 IDH2 저해제); LRP5 및/또는 DKK2 활성의 저해제(예를 들어, LRP5 및/또는 DKK2 저해제); KRAS 활성의 저해제; HTT 활성의 활성화제; 또는 DPP-4 활성의 저해제(예를 들어, DPP-4 저해제)를 포함한다.

[0745] 일부 구현예에서, 이펙터는 조절 세포내 폴리펩티드를 포함한다. 일부 구현예에서, 조절 세포내 폴리펩티드는 표적 세포에 대해 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)에 결합한다. 일부 구현예에서, 조절 세포내 폴리펩티드는 표적 세포에 대해 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)의 수준 또는 활성을 증가시킨다. 일부 구현예에서, 조절 세포내 폴리펩티드는 표적 세포에 대해 내인성인 하나 이상의 분자(예를 들어, 단백질 또는 핵산)의 수준 또는 활성을 감소시킨다.

[0746] **예시적인 분비 폴리펩티드 이펙터**

[0747] 예시적인 분비 치료제는 본 명세서에, 예를 들어, 하기 표에 기재되어 있다.

[0748] [표 50] 예시적인 사이토카인 및 사이토카인 수용체

사이토카인	사이토카인 수용체(들)	Entrez Gene ID	UniProt ID
IL-1α, IL-1β, 또는 이의 이종이량체	IL-1 타입 1 수용체, IL-1 타입 2 수용체	3552, 3553	P01583, P01584
IL-1Ra	IL-1 타입 1 수용체, IL-1 타입 2 수용체	3454, 3455	P17181, P48551
IL-2	IL-2R	3558	P60568
IL-3	IL-3 수용체 α + β c (CD131)	3562	P08700
IL-4	IL-4R 타입 I, IL-4R 타입 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
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-11Rα 1 gp130	3589	P20809
IL-12 (예: p35, p40, 또는 이의 이종이량체)	IL-12Rβ1 및 IL-12Rβ2	3593, 3592	P29459, P29460
IL-13	IL-13R1α1 및 IL-13R1α2	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

[0749]

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 (예: p19, p40, 또는 이의 이종이량체)	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-10R2 사슬 및 IL-20R1 사슬	55801	Q9NPH9
IL-27 (예: p28, EBI3, 또는 이의 이종이량체)	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/OSMRβ	386653	Q6EBC2
IL-32		9235	P24001
IL-33	ST2	90865	O95760
IL-34	콜로니-자극 인자 1 수용체	146433	Q6ZMJ4
IL-35 (예: p35, EBI3, 또는 이의 이종이량체)	IL-12Rβ2/gp130; IL-12Rβ2/IL-12Rβ2; gp130/gp130	10148	Q14213
IL-36	IL-36Ra	27179	Q9UHA7
IL-37	IL-18Rα 및 IL-18BP	27178	Q9NZH6
IL-38	IL-1R1, IL-36R	84639	Q8WWZ1
IFN-α	IFNAR	3454	P17181

[0750]

IFN-β	IFNAR	3454	P17181
IFN-γ	IFNGR1/IFNGR2	3459	P15260
TGF-β	TβR-I 및 TβR-II	7046, 7048	P36897, P37173
TNF-α	TNFR1, TNFR2	7132, 7133	P19438, P20333

[0751]

[0752] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 50의 사이토카인, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그(paralog)) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 UniProt ID를 참조하여 표 50에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대하여 상응하는 야생형 사이토카인의 Kd보다 10%, 20%, 30%, 40%, 또는 50% 이하로 더 높거나 더 낮은 Kd로 상응하는 사이토카인 수용체에 결합한다. 일부 구현예에서, 이펙터는 제1 영역(예를 들어, 표 50

의 사이토카인 폴리펩티드 또는 이의 기능적 변이체 또는 단편) 및 제2의 이종 영역을 포함하는 융합 단백질을 포함한다. 일부 구현예에서, 제1 영역은 표 50의 제1 사이토카인 폴리펩티드이다. 일부 구현예에서, 제2 영역은 표 50의 제2 사이토카인 폴리펩티드이며, 여기서 제1 및 제2 사이토카인 폴리펩티드는 야생형 세포에서 서로 사이토카인 이종이량체를 형성한다. 일부 구현예에서, 표 50의 폴리펩티드 또는 이의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대해 내인성인 신호 서열, 또는 이종 신호 서열을 포함한다. 일부 구현예에서, 표 50의 사이토카인, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 본 명세서에 기재된 질환 또는 장애의 치료에 사용된다.

[0753] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 50의 사이토카인에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 50의 사이토카인 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 항체 분자는 신호 서열을 포함한다.

[0754] 예시적인 사이토카인 및 사이토카인 수용체는, 예를 들어, 문헌[Akdis et al., "Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β, and TNF-α: Receptors, functions, and roles in diseases" October 2016 Volume 138, Issue 4, Pages 984-1010]에 기재되어 있으며, 이는 상기 문헌의 표 I을 포함하여 본 명세서에 전문이 참조로 포함된다.

[0755] [표 51] 예시적인 폴리펩티드 호르몬 및 수용체

호르몬	수용체	Entrez Gene ID	UniProt ID
나트륨이뇨 펩티드, 예를 들어, 심방 나트륨이뇨 펩티드(ANP)	NPRA, NPRB, NPRC	4878	P01160
뇌 나트륨이뇨 펩티드(BNP)	NPRA, NPRB	4879	P16860
C 형 나트륨이뇨 펩티드(CNP)	NPRB	4880	P23582
성장 호르몬(GH)	GHR	2690	P10912
인간 성장 호르몬(hGH)	hGH 수용체(인간 GHR)	2690	P10912
프롤락틴(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
황체 형성 호르몬(LH)	LHR	3973	P22888
항이뇨 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

[0756]

[0757] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 51의 호르몬, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 UniProt ID를 참조하여 표 51에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대하여 상응하는 야생형 호르몬의 Kd보다 10%, 20%, 30%, 40%, 또는 50% 이하로 더 높은 Kd로 상응하는 수용체에 결합한다. 일부 구현예에서, 표 51의 폴리펩티드 또는 이의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대해 내인성인 신호 서열, 또는 이종 신호 서열을 포함한다. 일부 구현예에서, 표 51의 호르몬, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 본 명세서에 기재된 질환 또는 장애의 치료에 사용된다.

[0758] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 51의 호르몬에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 51의 호르몬 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 항체 분자는 신호 서열을 포함한다.

[0759] [표 52] 예시적인 성장 인자

성장 인자		Entrez Gene ID	UniProt ID
PDGF 패밀리			
PDGF(예를 들어, PDGF-1, PDGF-2, 또는 이의 이종이량체)	PDGF 수용체, 예를 들어, PDGFRα, PDGFRβ	5156	P16234
CSF-1	CSF1R	1435	P09603
SCF	CD117	3815	P10721
VEGF 패밀리			
VEGF(예를 들어, 아이소폼 VEGF 121, VEGF 165, VEGF 189, 및 VEGF 206)	VEGFR-1, VEGFR-2	2321	P17948
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-α	EGFR	7039	P01135
앰피레귤린	EGFR	374	P15514
HB-EGF	EGFR	1839	Q99075
베타셀룰린	EGFR, ErbB-4	685	P35070
에피레귤린	EGFR, ErbB-4	2069	O14944

[0760]

헤레귤린	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
안지오포이에틴 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

[0761]

[0762] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 52의 성장 인자, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 UniProt ID를 참조하여 표 52에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 기능적 변이체는 동일한 조건 하에서 동일한 수용체에 대하여 상응하는 야생형 성장 인자의 Kd보다 10%, 20%, 30%, 40%, 또는 50% 이하로 더 높은 Kd로 상응하는 수용체에 결합한다. 일부 구현예에서, 표 52의 폴리펩티드 또는 이의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대해 내인성인 신호 서열, 또는 이종 신호 서열을 포함한다. 일부 구현예에서, 표 52의 성장 인자, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 본 명세서에 기재된 질환 또는 장애의 치료에 사용된다.

[0763] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 52의 성장 인자에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 52의 성장 인자 수용체에 결합하는 항체 분자(예를 들어, scFv)를 포함한다. 일부 구현예에서, 항체 분자는 신호 서열을 포함한다.

[0764] 예시적인 성장 인자 및 성장 인자 수용체는, 예를 들어, 문헌[Bafico et al., "Classification of Growth Factors and Their Receptors" Holland-Frei Cancer Medicine. 6th edition]에 기재되어 있으며, 이는 본 명세서에 전문이 참조로 포함된다.

[0765] [표 53] 응고-연관 인자

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

[0766]

[0767] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 53의 폴리펩티드, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 UniProt ID를 참조하여 표 53에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 기능적 변이체는, 예를 들어, 야생형 단백질보다 10%, 20%, 30%, 40%, 또는 50% 이상 더 낮은 속도로, 상응하는 야생형 단백질과 동일한 반응을 촉매한다. 일부 구현예에서, 표 53의 폴리펩티드 또는 이의 기능적 변이체는 신호 서열, 예를 들어, 이펙터에 대해 내인성인 신호 서열, 또는 이종 신호 서열을 포함한다. 일부 구현예에서, 표 53의 폴리펩티드, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 표 53의 질환 또는 장애의 치료에 사용된다.

[0768] **예시적인 단백질 대체 치료제**

[0769] 예시적인 단백질 대체 치료제는 본 명세서에, 예를 들어, 하기 표에 기재되어 있다.

[0770] [표 54] 예시적인 효소 이펙터 및 상응하는 적응증

이펙터	결핍증	Entrez Gene ID	UniProt ID
3-메틸크로토닐-CoA 카복실라제	3-메틸크로토닐-CoA 카복실라제 결핍증	56922, 64087	Q96RQ3, Q9HCC0
아세틸-CoA-글루코사미니드 N-아세틸트랜스퍼라제	점액다당류증 MPS III(산필리포 증후군) 유형 III-C	138050	Q68CP4
ADAMTS13	혈전성 혈소판 감소성 자반증	11093	Q76LX8
아데닌 포스포리보실트랜스퍼라제	아데닌 포스포리보실트랜스퍼라제 결핍증	353	P07741
아데노신 데아미나제	아데노신 데아미나제 결핍증	100	P00813
ADP-리보스 단백질 하이드롤라제	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	이염백색질장애	410	P15289
카텝신 K	피크노디소토시스	1513	P43235
세라미다제	파아버병(지방육아종증)	125981, 340485, 55331	Q8TDN7, Q5QJU3, Q9NUN7
시스타티오닌 B 신타제	호모시스틴뇨증	875	P35520

[0771]

[0772]

돌리콜-P-만노스 신타제	N-글리코실화 CDG Ie 의 선천성 장애	8813, 54344	O60762, Q9P2X0
돌리코-P-Glc:Man9GlcNAc2-PP-돌리콜 글루코실트랜스퍼라제	N-글리코실화 CDG Ic 의 선천성 장애	84920	Q5BKT4
돌리코-P-Man:Man5GlcNAc2-PP-돌리콜 만노실트랜스퍼라제	N-글리코실화 CDG Id 의 선천성 장애	10195	Q92685
돌리킬-P-글루코스:Glc-1-Man-9-GlcNAc-2-PP-돌리킬-α-3-글루코실트랜스퍼라제	N-글리코실화 CDG Ih 의 선천성 장애	79053	Q9BVK2
돌리킬-P-만노스:Man-7-GlcNAc-2-PP-돌리킬-α-6-만노실트랜스퍼라제	N-글리코실화 CDG Ig 의 선천성 장애	79087	Q9BV10
인자 II	인자 II 결핍증	2147	P00734
인자 IX	B 형 혈우병	2158	P00740
인자 V	오우렌병	2153	P12259
인자 VIII	A 형 혈우병	2157	P00451
인자 X	스튜어트-프라워 인자 결핍증	2159	P00742
인자 XI	C 형 혈우병	2160	P03951
인자 XIII	피브린 안정화 인자 결핍증	2162, 2165	P00488, P05160
갈락토사민-6-설페이트 설파타제	점액다당류증 MPS IV(모르키오 증후군) 유형 IV-A	2588	P34059
갈락토실세라마이드 β-갈락토시다제	크라베병	2581	P54803

[0773]

강글리오시드 β-갈락토시다제	전신 GM1 강글리오시드증	2720	P16278
강글리오시드 β-갈락토시다제	GM2 강글리오시드증	2720	P16278
강글리오시드 β-갈락토시다제	스핑고지질증 유형 I	2720	P16278
강글리오시드 β-갈락토시다제	스핑고지질증 유형 II (소아형)	2720	P16278
강글리오시드 β-갈락토시다제	스핑고지질증 유형 III (성인형)	2720	P16278
글루코시다제 I	N-글리코실화 CDG IIb 의 선천성 장애	2548	P10253
글루코실세라마이드 β-글루코시다제	고셰병	2629	P04062
헤파란-S-설페이트 설파미다제	점액다당류증 MPS III(산필리포 증후군) 유형 III-A	6448	P51688
호모겐티세이트 옥시다제	알캅톤뇨증	3081	Q93099
히알루로니다제	점액다당류증 MPS IX(히알루로니다제 결핍증)	3373, 8692, 8372, 23553	Q12794, Q12891, O43820, Q2M3T9
이두로네이트 설페이트 설파타제	점액다당류증 MPS II (헌터 증후군)	3423	P22304
레시틴-콜레스테롤 아실트랜스퍼라제 (LCAT)	완전 LCAT 결핍증, 물고기눈병, 죽상동맥경화증, 고콜레스테롤혈증	3931	606967
라이신 옥시다제	글루타르산혈증 유형 I	4015	P28300
리소좀산 리파제	콜레스테릴 에스테르 축적병(CESD)	3988	P38571
리소좀산 리파제	리소좀산 리파제 결핍증	3988	P38571
리소좀산 리파제	볼만병	3988	P38571

[0774]

리소좀 펩스타틴-비감수성 펩티다제	세로이드 리포푸신증 후기 영아형(CLN2, 얀스키-빌쇼스키병)	1200	O14773
만노스(Man) 포스페이트 (P) 아이소머라제	N-글리코실화 CDG Ib 의 선천성 장애	4351	P34949
만노실-α-1,6-당단백질-β-1,2-N-아세틸글루코사미닐 트랜스퍼라제	N-글리코실화 CDG IIa 의 선천성 장애	4247	Q10469
메탈로프로테이나제-2	윈체스터 증후군	4313	P08253
메틸말로닐-CoA 뮤타제	메틸말론산혈증(비타민 b12 비-반응성)	4594	P22033
N-아세틸 갈락토사민 α-4-설페이트 설파타제 (아릴설파타제 B)	점액다당류증 MPS VI (마로토-라미 증후군)	411	P15848
N-아세틸-D-글루코사미니다제	점액다당류증 MPS III(산필리포 증후군) 유형 III-B	4669	P54802
N-아세틸-갈락토사미니다제	쉰들러병 유형 I(영아 중증형)	4668	P17050
N-아세틸-갈락토사미니다제	쉰들러병 유형 II(칸자키병, 성인-발병형)	4668	P17050
N-아세틸-갈락토사미니다제	쉰들러병 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(거짓-헐러 다발디스트로피)	79158	Q3T906

N-아세틸글루코사미닐-1-포스포트랜스퍼라제 촉매 서브유닛	점액지질증 ML II (I-세포병)	79158	Q3T906
N-아세틸글루코사미닐-1-포스포트랜스퍼라제, 기질-인식 서브유닛	점액지질증 ML III (거짓-헐러 다발디스트로피) 유형 III-C	84572	Q9UJJ9
N-아스파르틸글루코사민니다제	아스파르틸글루코사민뇨	175	P20933
뉴라미니다제 1 (시알리다제)	시알산증	4758	Q99519
팔미토일-단백질 티오에스테라제-1	세로이드 리포푸신증 성인형(CLN4, 쿠프스병)	5538	P50897
팔미토일-단백질 티오에스테라제-1	세로이드 리포푸신증 영아형(CLN1, 산타부리-할티아병)	5538	P50897
페닐알라닌 하이드록실라제	페닐케톤뇨증	5053	P00439
포스포만노뮤타제-2	N-글리코실화 CDG Ia 의 선천성 장애(오직 신경계 및 신경계-다기관 형태)	5373	O15305
포르포빌리노겐 데아미나제	급성 간헐성 포르피린증	3145	P08397
퓨린 뉴클레오시드 포르포릴라제	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

[0775]

스테롤 27-하이드록실라제	뇌건황색종증(콜레스타놀 지질증)	1593	Q02318
티미딘 포스포릴라제	미토콘드리아 신경위장 뇌근병증(MNGIE)	1890	P19971
트리헥소실세라마이드 α-갈락토시다제	파브리병	2717	P06280
티로시나제, 예를 들어, OCA1	백색증, 예를 들어, 눈백색증	7299	P14679
UDP-GlcNAc:돌리킬-P NAcGlc 포스포트랜스퍼라제	N-글리코실화 CDG Ij 의 선천성 장애	1798	Q9H3H5
UDP-N-아세틸글루코사민-2-에피머라제/N-아세틸만노사민 키나제, 시알린	시알뇨증(Sialuria) 프랑스형	10020	Q9Y223
우리카제	레슈-니한 증후군, 통풍	391051	단백질 없음
우리딘 디포스페이트 글루쿠로닐-트랜스퍼라제(예를 들어, UGT1A1)	크리글러-나자르 증후군	54658	P22309
α-1,2-만노실트랜스퍼라제	N-글리코실화 CDG Il 의 선천성 장애(608776)	79796	Q9H6U8
α-1,2-만노실트랜스퍼라제	N-글리코실화의 선천성 장애, 유형 I(pre-Golgi 글리코실화 결함)	79796	Q9H6U8
α-1,3-만노실트랜스퍼라제	N-글리코실화 CDG Ii 의 선천성 장애	440138	Q2TAA5
α-D-만노시다제	α-만노시도시스, 유형 I(중증) 또는 II(경증)	10195	Q92685
α-L-푸코시다제	푸코시드축적증	4123	Q9NTJ4

[0776]

α-l-이두로니다제	점액다당류증 MPS IH/S(헐러-샤이에 증후군)	2517	P04066
α-l-이두로니다제	점액다당류증 MPS I-H(헐러 증후군)	3425	P35475
α-l-이두로니다제	점액다당류증 MPS I-S(샤이에 증후군)	3425	P35475
β-1,4-갈락토실트랜스퍼라제	N-글리코실화 CDG IId 의 선천성 장애	3425	P35475
β-1,4-만노실트랜스퍼라제	N-글리코실화 CDG Ik 의 선천성 장애	2683	P15291
β-D-만노시다제	β-만노시도시스	56052	Q9BT22
β-갈락토시다제	점액다당류증 MPS IV(모르키오 증후군) 유형 IV-B	4126	O00462
β-글루쿠로니다제	점액다당류증 MPS VII(슬라이 증후군)	2720	P16278
β-헥소미니다제 A	테이-삭스병	2990	P08236
β-헥소미니다제 B	샌드호프병	3073	P06865

[0777]

[0778] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 54의 효소, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는

UniProt ID를 참조하여 표 54에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 기능적 변이체는, 예를 들어, 야생형 단백질보다 10%, 20%, 30%, 40%, 또는 50% 이상 더 낮은 속도로, 상응하는 야생형 단백질과 동일한 반응을 촉매한다. 일부 구현예에서, 표 54의 효소, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 표 54의 질환 또는 장애의 치료에 사용된다. 일부 구현예에서, 아넬로벡터는 우리딘 디포스페이트 우리긴 글루쿠로닐-트랜스퍼라제 또는 이의 기능적 변이체를 표적 세포, 예를 들어, 간 세포로 운반하는 데 사용된다. 일부 구현예에서, 아넬로벡터는 OCA1 또는 이의 기능적 변이체를 표적 세포, 예를 들어, 망막 세포로 운반하는 데 사용된다.

[0779] **[표 55] 예시적인 비-효소 이펙터 및 상응하는 적응증**

이펙터	적응증	Entrez Gene 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 요독증후군	3075	P08603
시스틴증(리소좀 시스틴 수송체)	시스틴증	1497	O60931
부고환 분비 단백질 1(HE1; NPC2 단백질)	니만-픽병 유형 C2	10577	P61916
GDP-푸코스 수송체-1	N-글리코실화 CDG IIc 의 선천성 장애(람밤=-하샤론(Rambam-Hasharon) 증후군)	55343	Q96A29
GM2 활성화제 단백질	GM2 활성화제 단백질 결핍증(테이-삭스병 AB 변이형, GM2A)	2760	Q17900
리소좀 막횡단 CLN3 단백질	세로이드 리포푸신증 소아형(CLN3, 바텐병, 포그트-스필마이어병)	1207	Q13286
리소좀 막횡단 CLN5 단백질	세로이드 리포푸신증 변이형 후기 영아형, 핀란드형(CLN5)	1203	O75503

[0780]

Na 인산염 공동수송체, 시알린	영아 시알산 축적 장애	26503	Q9NRA2
Na 인산염 공동수송체, 시알린	시알뇨증 핀란드형(살라병)	26503	Q9NRA2
NPC1 단백질	니만-픽병 유형 C1/유형 D	4864	O15118
올리고머 골지 복합체-7	N-글리코실화 CDG IIe 의 선천성 장애	91949	P83436
프로사포신	프로사포신 결핍증	5660	P07602
보호 단백질/카텝신 A(PPCA)	갈락토시알리도증(골드버그 증후군, 조합된 뉴라미니다제 및 β-갈락토시다제 결핍증)	5476	P10619
만노스-P-돌리콜 이용에 수반되는 단백질	N-글리코실화 CDG If 의 선천성 장애	9526	O75352
사포신 B	사포신 B 결핍증(술파티드 활성화제 결핍증)	5660	P07602
사포신 C	사포신 C 결핍증(고셰 활성화제 결핍증)	5660	P07602
술파타제-변형 인자-1	뮤코술파티도시스(Mucosulfatidosis)(다발성 술파타제 결핍증)	285362	Q8NBK3
막횡단 CLN6 단백질	세로이드 리포푸신증 변이형 후기 영아형(CLN6)	54982	Q9NWW5
막횡단 CLN8 단백질	지적 장애를 갖는 세로이드 리포푸신증 진행성 간질	2055	Q9UBY8
vWF	폰 빌레브란트병	7450	P04275
인자 I(피브리노겐)	무피브리노겐혈증	2243, 2244, 2266	P02671, P02675, P02679
에리트로포이에틴(hEPO)			

[0781]

[0782] 일부 구현예에서, 본 명세서에 기재된 이펙터는 에리트로포이에틴(EPO), 예를 들어, 인간 에리트로포이에틴(hEPO), 또는 이의 기능적 변이체를 포함한다. 일부 구현예에서, 에리트로포이에틴, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 적혈구형성을 자극하는 데 사용된다. 일부 구현예에서, 에리트로포이에틴, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 질환 또는 장애, 예를 들어, 빈혈의 치료에 사용된다. 일부 구현예에서, 아넬로벡터는 EPO 또는 이의 기능적 변이체를 표적 세포, 예를 들어, 적혈구로 운반하는 데 사용된다.

[0783] 일부 구현예에서, 본 명세서에 기재된 이펙터는 표 55의 폴리펩티드, 또는 이의 기능적 변이체, 예를 들어, 이의 상동체(예를 들어, 오솔로그 또는 파라로그) 또는 단편을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 UniProt ID를 참조하여 표 55에 열거된 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 서열 동일성을 갖는 단백질을 포함한다. 일부 구현예에서, 표 55의 폴리펩티드, 또는 이의 기능적 변이체를 인코딩하는 아넬로벡터는 표 55의 질환 또는 장애의 치료에 사용된다. 일부 구현예에서, 아넬로벡터는 SMN 또는 이의 기능적 변이체를 표적 세포, 예를 들어, 척수 및/또는 운동 뉴런의 세포로 운반하는 데

사용된다. 일부 구현예에서, 아넬로벡터는 마이크로-디스트로핀을 표적 세포, 예를 들어, 근세포로 운반하는 데 사용된다.

[0784] 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.

[0785] 일부 구현예에서, 본 명세서에 기재된 이펙터는 응고 인자, 예를 들어, 본 명세서에서 표 54 또는 표 55에 열거된 응고 인자를 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 돌연변이되는 경우 리소좀 축적 장애를 야기하는 단백질, 예를 들어, 본 명세서에서 표 54 또는 표 55에 열거된 단백질을 포함한다. 일부 구현예에서, 본 명세서에 기재된 이펙터는 수송체 단백질, 예를 들어, 본 명세서에서 표 55에 열거된 수송체 단백질을 포함한다.

[0786] 일부 구현예에서, 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 촉매한다. 일부 구현예에서, 기능적 변이체는 야생형 단백질에 의해 결합되는 동일한 결합 파트너에, 예를 들어, 동일한 조건 하의 동일한 결합 파트너에 대한 상응하는 야생형 단백질의 Kd보다 10%, 20%, 30%, 40% 또는 50% 이하로 더 큰 Kd로 결합한다. 일부 구현예에서, 기능적 변이체는 야생형 폴리펩티드의 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일한 폴리펩티드 서열을 갖는다. 일부 구현예에서, 기능적 변이체는 상응하는 야생형 단백질의 상동체(예를 들어, 오솔로그(오솔로그) 또는 파라로그)를 포함한다. 일부 구현예에서, 기능적 변이체는 융합 단백질이다. 일부 구현예에서, 융합물은 상응하는 야생형 단백질에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일성을 갖는 제1 영역 및 제2 이종 영역을 포함한다. 일부 구현예에서, 기능적 변이체는 상응하는 야생형 단백질의 단편을 포함하거나, 이로 이루어진다.

[0787] **재생, 복구, 섬유증 인자**

[0788] 본 명세서에 기재된 치료용 폴리펩티드는 또한, 예를 들어, 표 56에 개시된 바와 같은 성장 인자, 또는 이의 기능적 변이체, 예를 들어, UniProt ID를 참조하여 표 56에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질을 포함한다. 또한 재생 및 복구를 촉진시키는 이러한 성장 인자, 또는 miRNA에 대한 항체 또는 이의 단편이 포함된다.

[0789] [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	Gene ID: 2246	NP_001341882
miR-199-3p	MIMAT0000232	
miR-590-3p	MIMAT0004801	
mi-17-92	MI0000071	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/
miR-222	MI0000299	
miR-302-367	MIR302A And MIR367	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/

[0790]

[0791] **형질전환 인자**

[0792] 본 명세서에 기재된 치료용 폴리펩티드는 또한, 예를 들어, 섬유아세포를 분화 세포로 형질전환시키는 단백질 인자, 예를 들어, 표 57에 개시된 인자 또는 이의 기능적 변이체, 예를 들어, UniProt ID를 참조하여 표 57에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질을 포함한다.

[0793] [표 57] 예시적인 형질전환 인자

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

[0794]

[0795] **세포 재생을 자극하는 단백질**

[0796] 본 명세서에 기재된 치료용 폴리펩티드는 또한 세포 재생을 자극하는 단백질, 예를 들어, 표 58에 개시된 단백질 또는 이의 기능적 변이체, 예를 들어, UniProt ID를 참조하여 표 58에 개시된 단백질 서열에 대하여 적어도 80%, 85%, 90%, 95%, 967%, 98%, 99% 동일성을 갖는 단백질을 포함한다.

[0797] [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

[0798]

[0799] STING 조절제 이펙터

[0800] 일부 구현예에서, 본 명세서에 기재된 분비 이펙터는 STING/cGAS 신호전달을 조절한다. 일부 구현예에서, STING 조절제는 폴리펩티드, 예를 들어, 바이러스 폴리펩티드 또는 이의 기능적 변이체이다. 예를 들어, 이펙터는 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 조절제(예를 들어, 저해제)를 포함할 수 있으며, 이는 본 명세서에 전문이 참조로 포함된다. 추가의 STING 조절제(예를 들어, 활성화제)는, 예를 들어, 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]; 문헌[Bose "cGAS/STING Pathway in Cancer: Jekyll and Hyde Story of Cancer Immune Response" Int J Mol Sci. 2017 Nov; 18(11): 2456]; 및 문헌[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 있으며, 이들 각각은 본 명세서에 전문이 참조로 포함된다.

[0801] 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, 이에 제한되지 않는다. 본 명세서에 기재된 본 발명에 유용한 펩티드는 또한, 항원-결합 펩티드, 예를 들어, 항원 결합 항체 또는 항체-유사 단편, 예컨대 단일 쇄 항체, 나노바디를 포함한다(예를 들어, 문헌[Steeland et al. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. Drug Discov Today: 21(7):1076-113] 참조). 이러한 항원 결합 펩티드는 시토졸 항원, 핵 항원, 또는 소기관-내 항원에 결합할 수 있다.

[0802] 일부 구현예에서, 유전 요소는 작은 펩티드, 펩티드모방체(예를 들어, 펩토이드(peptoid)), 아미노산, 및 아미노산 유사체를 인코딩하는 서열을 포함한다. 이러한 치료제는 일반적으로 몰당 약 5,000 그램 미만의 분자량, 몰당 약 2,000 그램 미만의 분자량, 몰당 약 1,000 그램 미만의 분자량, 몰당 약 500 그램 미만의 분자량, 및 이러한 화합물의 염, 에스테르 및 다른 약제학적으로 허용 가능한 형태를 갖는다. 이러한 치료제는 신경전달물

질, 호르몬, 약물, 독소, 바이러스 또는 미생물 입자, 합성 분자, 및 이의 효능제 또는 길항제를 포함할 수 있지만, 이에 제한되지 않는다.

[0803] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 특정 위치, 조직, 또는 세포를 표적화할 수 있는 리간드에 연결된 폴리펩티드를 포함한다.

[0804] *유전자 편집 구성성분*

[0805] 아넬로벡터의 유전 요소는 유전자 편집 시스템의 구성성분을 인코딩하는 하나 이상의 유전자를 포함할 수 있다. 예시적인 유전자 편집 시스템은 클러스터링된 규칙적으로 산재된 짧은 회문 반복부(CRISPR) 시스템, 아연 핑거 뉴클레아제(ZFN), 및 전사 활성화제-유사 이펙터-기반 뉴클레아제(TALEN)를 포함한다. ZFN, TALEN, 및 CRISPR-기반 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.

[0806] CRISPR 시스템은 원래 박테리아 및 고세균에서 발견된 적응 방어 시스템이다. CRISPR 시스템은 CRISPR-회합 또는 "Cas" 엔도뉴클레아제(예를 들어, Cas9 또는 Cpf1)로 명명되는 RNA-가이드된 뉴클레아제를 사용하여, 외래 DNA를 절단한다. 통상적인 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 단백질이 아닌) 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 통상적으로 표적 DNA 서열에 상응하는 약 20개 뉴클레오티드 RNA 서열인 "가이드 RNA"를 포함한다. crRNA는 또한, tracrRNA에 결합하여, RNase III에 의해 절단되는 부분 이중-가닥 구조를 형성하여, crRNA/tracrRNA 하이브리드를 초래하는 영역을 포함한다. 그 다음, crRNA/tracrRNA 하이브리드는 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 전체에 걸쳐 나타난다.

[0807] 일부 구현예에서, 아넬로벡터는 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 써모필러스(Streptococcus thermophilus) CRISPR1), 5'-NGGNG(스트렙토코커스 써모필러스 CRISPR3), 및 5'-NNNGATT(네이세리아 메닌지디티스(Neisseria meningiditis))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 평활-말단(blunt-end) 절단을 수행한다. 또 다른 부류 II CRISPR 시스템은 Cas9보다 더 작은 V형 엔도뉴클레아제 Cpf1을 포함하며; 예는 AsCpf1(아시드아미노코커스(Acidaminococcus) 종 유래) 및 LbCpf1(라크노스피라세아에(Lachnospiraceae) 종 유래)을 포함한다. 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(overhang)을 갖는 오프셋 또는 스태거형 이중-가닥 파단부를 도입함으로써, 예를 들어, 코딩 가닥 상의 PAM 부위(이로부터 3')로부터 18개 뉴클레오티드 하류, 및 상보성 가닥 상의 PAM 부위로부터 23개 뉴클레오티드 하류에 위치한 5-뉴클레오티드 오프셋 또는 스태거형 절단으로 표적 DNA를 절단함으로써 표적 DNA를 절단하며; 이러한 오프셋 절단으로부터 초래되는 5-뉴클레오티드 오버행은 상동성 재조합에 의한 DNA 삽입에 의해 평활-말단 절단된 DNA에서의 삽입에 의한 것보다 더 정밀한 게놈 편집을 가능하게 한다. 예를 들어, 문헌[Zetsche et al. (2015) Cell, 163:759 - 771]을 참조한다.

[0808] 다양한 CRISPR 회합(Cas) 유전자가 아넬로벡터 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 것이다. 일부 구현예에서, 아넬로벡터는, 다양한 원핵생물 종 중 임의의 것의 유래일 수 있는 Cas 단백질, 예를 들어, Cas9 단백질을 인코딩하는 유전자를 포함한다. 일부 구현예에서, 아넬로벡터는 특정 프로토스페이서-인접 모티프(PAM) 서열을 인식하도록 선택되는 특정 Cas 단백질, 예를 들어, 특정 Cas9 단백질을 인코딩하는 유전자를 포함한다. 일부 구현예에서, 아넬로벡터는 세포, 접합체, 배아 또는 동물 내로 도입되어, 예를 들어, 동일한, 유사한 또는 상이한 PAM 모티프를 포함하는 부위의 인식 및 변형을 가능하

게 할 수 있는 2가지 이상의 상이한 Cas 단백질 또는 2가지 이상의 Cas 단백질을 인코딩하는 핵산을 포함한다. 일부 구현예에서, 아넬로벡터는 비활성화된 뉴클레아제를 갖는 변형된 Cas 단백질, 예를 들어, 뉴클레아제-결핍 Cas9를 인코딩하는 유전자를 포함한다.

[0809] 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 Cas엔도뉴클레아제(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 개시되고, 이로부터 공개적으로 이용 가능한 수많은 CRISPR/Cas9 플라스미드를 참조한다. 각각 개별 가이드 RNA에 의해 유도되는, 2개의 별개 이중-가닥 파단을 도입하는 "이중 닉카제" Cas9는 더 정확한 게놈 편집을 달성하는 것으로 문헌[Ran et al. (2013) Cell, 154:1380 - 1389]에 기재되어 있다.

[0810] 진핵생물의 유전자를 편집하기 위한 CRISPR 기술은 미국 특허 출원 공개 2016/0138008A1 및 US2015/0344912A1, 및 미국 특허 8,697,359, 8,771,945, 8,945,839, 8,999,641, 8,993,233, 8,895,308, 8,865,406, 8,889,418, 8,871,445, 8,889,356, 8,932,814, 8,795,965, 및 8,906,616에 개시되어 있다. Cpf1 엔도뉴클레아제 및 상응하는 가이드 RNA 및 PAM 부위는 미국 특허 출원 공개 2016/0208243 A1에 개시되어 있다.

[0811] 일부 구현예에서, 아넬로벡터는 본 명세서에 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 일부(예를 들어, 이의 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.

[0812] 일부 구현예에서, 아넬로벡터는 dCas9와 하나 이상의 이펙터 도메인의 전부 또는 일부(예를 들어, 전장 야생형 이펙터 도메인, 또는 이의 단편 또는 변이체, 예를 들어, 이의 생물학적 활성 부분)의 융합물을 인코딩하는 유전자를 포함하여 본 명세서에 기재된 방법에 유용한 키메라 단백질을 생성한다. 따라서, 일부 구현예에서, 아넬로벡터는 dCas9-메틸라제 융합물을 인코딩하는 유전자를 포함한다. 다른 일부 구현예에서, 아넬로벡터는 내인성 유전자를 표적화하기 위해 부위-특이적 gRNA와 함께 dCas9-효소 융합물을 인코딩하는 유전자를 포함한다.

[0813] 다른 양태에서, 아넬로벡터는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개, 또는 그 이상의 dCas9와 융합된 이펙터 도메인(전체 또는 생물학적 활성 부분)을 인코딩하는 유전자를 포함한다.

[0814] *조절 서열*

[0815] 일부 구현예에서, 유전 요소는 이펙터를 인코딩하는 서열에 작동 가능하게 연결된 조절 서열, 예를 들어, 프로모터 또는 인핸서를 포함한다.

[0816] 일부 구현예에서, 프로모터는 발현 생성물을 인코딩하는 DNA 서열에 인접하여 위치한 DNA 서열을 포함한다. 프로모터는 인접 DNA 서열에 작동 가능하게 연결될 수 있다. 프로모터는 통상적으로 프로모터가 존재하지 않는 경우 발현되는 생성물의 양과 비교하여 DNA 서열로부터 발현되는 생성물의 양을 증가시킨다. 하나의 유기체로부터의 프로모터를 이용하여 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.

[0817] 일 구현예에서, 높은 수준의 구성적 발현이 요망된다. 이러한 프로모터의 예는 제한 없이, 레트로바이러스 라우스 육종 바이러스(RSV) 긴 말단 반복부(LTR) 프로모터/인핸서, 사이토메갈로바이러스(CMV) 즉시 조기 프로모터/인핸서(예를 들어, 문헌[Boshart et al, Cell, 41:521-530 (1985)] 참조), SV40 프로모터, 디하이드로폴레이트

환원효소 프로모터, 세포질 β-액틴 프로모터 및 포스포글리세롤 키나제(PGK) 프로모터를 포함한다.

[0818] 또 다른 구현예에서, 유도성 프로모터가 요망될 수 있다. 유도성 프로모터는 제한 없이, 아연-유도성 양 메탈로티오닌(MT) 프로모터; 덱사메타손(Dex)-유도성 마우스 유방 종양 바이러스(MMTV) 프로모터; T7 중합효소 프로모터 시스템(WO 98/10088); 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.

[0819] 일부 구현예에서, 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 구현예에서, 다른 고유 발현 제어 요소, 예컨대 인핸서 요소, 폴리아데닐화 부위 또는 코작(Kozak) 컨센서스 서열도 또한 고유 발현을 모방하기 위해 사용될 수 있다.

[0820] 일부 구현예에서, 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, 디스트로핀, 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 a 쇄), 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.

[0821] 유전 요소는 인핸서, 예를 들어, 유전자를 인코딩하는 DNA 서열에 인접하여 위치한 DNA 서열을 포함할 수 있다. 인핸서 요소는 통상적으로 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, 인핸서 요소는 생성물을 인코딩하는 DNA 서열의 100개 염기쌍, 200개 염기쌍 또는 300개 이상의 염기쌍 상류 또는 하류에 위치할 수 있다. 인핸서 요소는 DNA 서열로부터 발현되는 재조합 생성물의 양을 프로모터 요소에 의해 제공되는 증가된 발현을 초과하여 증가시킬 수 있다. 다수의 인핸서 요소는 당업자에게 용이하게 이용 가능하다.

[0822] 일부 구현예에서, 유전 요소는 본 명세서에 기재된 발현 생성물을 인코딩하는 서열을 측접하는 하나 이상의 역위 말단 반복부(ITR)를 포함한다. 일부 구현예에서, 유전 요소는 본 명세서에 기재된 발현 생성물을 인코딩하는 서열을 측접하는 하나 이상의 긴 말단 반복부(LTR)를 포함한다. 사용될 수 있는 프로모터 서열의 예는 시미안 바이러스 40(SV40) 조기 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 조기 프로모터, 및 라우스 육종 바이러스 프로모터를 포함하지만, 이에 제한되지 않는다.

[0823] *복제 단백질*

[0824] 일부 구현예에서, 아넬로벡터, 예를 들어, 합성 아넬로벡터의 유전 요소는 하나 이상의 복제 단백질을 인코딩하는 서열을 포함할 수 있다. 일부 구현예에서, 아넬로벡터는 회전환 복제 방법에 의해 복제할 수 있으며, 예를 들어, 선도 가닥 및 지연 가닥의 합성은 분리된다. 이러한 구현예에서, 아넬로벡터는 3가지의 추가의 요소를 포함한다: i) 개시 단백질을 인코딩하는 유전자, ii) 이중 가닥 원점, 및 iii) 단일 가닥 원점. 복제 단백질을 포

함하는 회전환 복제(RCR) 단백질 복합체는 선도 가닥에 결합하며, 복제 원점을 불안정화시킨다. RCR 복합체는 게놈을 절단하여, 유리 3'OH 말단을 생성한다. 세포 DNA 중합효소는 유리 3'OH 말단으로부터 바이러스 DNA 복제를 개시한다. 게놈이 복제된 후에, RCR 복합체는 루프를 공유적으로 폐쇄시킨다. 이는 양성 환형 단일-가닥 모 DNA 분자 및 음성 모체 가닥 및 새로 합성된 양성 가닥으로 구성된 환형 이중-가닥 DNA 분자의 방출을 야기한다. 단일-가닥 DNA 분자는 캡시드화되거나 제2차의 복제에 수반될 수 있다. 예를 들어, 문헌[Virology Journal 2009, 6:60 doi:10.1186/1743-422X-6-60]을 참조한다.

[0825] 유전 요소는 중합효소, 예를 들어, RNA 중합효소 또는 DNA 중합효소를 인코딩하는 서열을 포함할 수 있다.

[0826] *기타 서열*

[0827] 일부 구현예에서, 유전 요소는 생성물(예를 들어, 리보자임, 단백질을 인코딩하는 치료용 mRNA, 외인성 유전자)을 인코딩하는 핵산을 추가로 포함한다.

[0828] 일부 구현예에서, 유전 요소는 아넬로벡터의 종 및/또는 조직 및/또는 세포 향성(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 국재화, 엑소시토시스 조절, 증식 및 숙주 또는 숙주 세포에서의 아넬로벡터의 핵산 보호에 영향을 미치는 하나 이상의 서열을 포함한다.

[0829] 일부 구현예에서, 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, 이에 제한되지 않는다. 일 구현예에서, 유전 요소는 조절 핵산과 동일한 유전자 발현 생성물의 상이한 유전자좌를 표적화하기 위하여 siRNA를 인코딩하는 서열을 포함한다. 일 구현예에서, 유전 요소는 조절 핵산과 상이한 유전자 발현 생성물을 표적화하기 위하여 siRNA를 인코딩하는 서열을 포함한다.

[0830] 일부 구현예에서, 유전 요소는 하기의 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 치료용 mRNA 또는 단백질을 인코딩하는 서열.

[0831] 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.

[0832] *인코딩된 유전자*

[0833] 예를 들어, 유전 요소는 신호전달 생화학 경로와 연관된 유전자, 예를 들어, 신호전달 생화학 경로-연관 유전자 또는 폴리뉴클레오티드를 포함할 수 있다. 예에는 질환 연관 유전자 또는 폴리뉴클레오티드가 포함된다. "질환-연관" 유전자 또는 폴리뉴클레오티드는 비 질환 대조군의 조직 또는 세포와 비교하여, 질환-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 질환의 발생 및/또는 진행과 상호연관된다. 질환-연관 유전자는 또한, 질환의 병인의 직접적인 원인이 되거나, 질환의 병인을 담당하는 유전자(들)와 연관 불균형인 돌연변이(들) 또는 유전적 변이를 지니는 유전자를 지칭한다.

[0834] 질환-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. 질환-연관 유전자 및 폴리뉴클레오티드의 예는 본 명세서에 전문이 참조로 포함되는 미국 특허 8,697,359의 표 A 및 B에 열거되어 있다. 질환 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 본 명세서에 전문이 참조

로 포함되는 미국 특허 8,697,359의 표 A 내지 C에 열거되어 있다.

[0835] 또한, 유전 요소는 본 명세서의 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.

[0836] *바이러스 서열*

[0837] 일부 구현예에서, 유전 요소는 적어도 하나의 바이러스 서열을 포함한다. 일부 구현예에서, 서열은, 예를 들어, 본 명세서에 기재된 바와 같은, 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV) 유래의 하나 이상의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열에 대하여 상동성 또는 동일성을 갖는다. 일부 구현예에서, 유전 요소는, 예를 들어, 본 명세서에 기재된 바와 같은, 아넬로바이러스 게놈 유래의 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 서열은 하기 표 41에 열거된 바와 같은 아넬로바이러스 게놈 유래이다.

[0838] **[표 41] 아넬로바이러스 및 이들의 서열의 예.** 수탁번호 및 관련 서열 정보는 2018년 12월 11일에 참조한 바와 같이 www.ncbi.nlm.nih.gov/genbank/에서 수득될 수 있다.

[0839]

수탁번호	설명
AB017613.1	토르크 테노 바이러스 16 DNA, 완전한 게놈, 단리주: TUS01
AB026345.1	ORF1 및 ORF2에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TRM1
AB026346.1	ORF1 및 ORF2에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TK16
AB026347.1	ORF1 및 ORF2에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TP1-3
AB028669.1	ORF1 및 ORF2에 대한 TT 바이러스 유전자, 완전한 게놈, 단리주:TJN02
AB030487.1	PORF2a, pORF2b, pORF1에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 클론:JaCHCTC19
AB030488.1	PORF2a, pORF2b, pORF1에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 클론:JaBD89
AB030489.1	PORF2a, pORF2b, pORF1에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 클론:JaBD98
AB038340.1	ORF2s, ORF1, ORF3에 대한 TT 바이러스 유전자, 완전한 코딩 서열
AB038622.1	ORF2, ORF1, ORF3에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TTVyon-LC011
AB038623.1	ORF2, ORF1, ORF3에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TTVyon-KC186
AB038624.1	ORF2, ORF1, ORF3에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 단리주:TTVyon-KC197
AB041821.1	VP1에 대한 TT 바이러스 mRNA, 완전한 코딩 서열
AB050448.1	ORF1, ORF2, ORF3, ORF4에 대한 토르크 테노 바이러스 유전자, 완전한 코딩 서열, 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 코딩 서열, 클론: SAa-38
AB060595.1	ORF1, ORF2, ORF3, ORF4 에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 클론:SAj-30
AB060596.1	ORF1, ORF2, ORF3, ORF4 에 대한 TT 바이러스 유전자, 완전한 코딩 서열, 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
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 주 TTVCHN1, 완전한 게놈
AF116842.1	TT 바이러스 주 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 미공지 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 코딩 서열
AF298585.1	TT 바이러스 폴리시(Polish) 단리주 P/1C1, 완전한 게놈
AF315076.1	TTV-유사 바이러스 DXL1 미공지 유전자
AF315077.1	TTV-유사 바이러스 DXL2 미공지 유전자
AF345521.1	TT 바이러스 단리주 TCHN-G1 Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF345522.1	TT 바이러스 단리주 TCHN-E Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF345525.1	TT 바이러스 단리주 TCHN-D2 Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF345527.1	TT 바이러스 단리주 TCHN-C2 Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF345528.1	TT 바이러스 단리주 TCHN-F Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF345529.1	TT 바이러스 단리주 TCHN-G2 Orf2 및 Orf1 유전자, 완전한 코딩 서열
AF371370.1	TT 바이러스 ORF1, ORF3, 및 ORF2 유전자, 완전한 코딩 서열
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

[0840]

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
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 코딩 서열
AY026466.1	TT 바이러스 단리주 L02 Orf2 및 Orf1 유전자, 완전한 코딩 서열
DQ003341.1	토르크 테노 바이러스 클론 P2-9-02 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 코딩 서열
DQ003342.1	토르크 테노 바이러스 클론 P2-9-07 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 코딩 서열

[0841]

DQ003343.1	토르크 테노 바이러스 클론 P2-9-08 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 코딩 서열
DQ003344.1	토르크 테노 바이러스 클론 P2-9-16 ORF2 (ORF2), ORF1A (ORF1A), 및 ORF1B (ORF1B) 유전자, 완전한 코딩 서열
DQ186994.1	토르크 테노 바이러스 클론 P601 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ186995.1	토르크 테노 바이러스 클론 P605 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ186996.1	토르크 테노 바이러스 클론 BM1A-02 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ186997.1	토르크 테노 바이러스 클론 BM1A-09 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ186998.1	토르크 테노 바이러스 클론 BM1A-13 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ186999.1	토르크 테노 바이러스 클론 BM1B-05 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187000.1	토르크 테노 바이러스 클론 BM1B-07 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187001.1	토르크 테노 바이러스 클론 BM1B-11 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187002.1	토르크 테노 바이러스 클론 BM1B-14 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187003.1	토르크 테노 바이러스 클론 BM1B-08 ORF2 (ORF2) 유전자, 완전한 코딩 서열; 및 비기능적 ORF1 (ORF1) 유전자, 완전한 서열
DQ187004.1	토르크 테노 바이러스 클론 BM1C-16 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187005.1	토르크 테노 바이러스 클론 BM1C-10 ORF2 (ORF2) 및 ORF1 (ORF1) 유전자, 완전한 코딩 서열
DQ187007.1	토르크 테노 바이러스 클론 BM2C-25 ORF2 (ORF2) 유전자, 완전한 코딩 서열; 및 비기능적 ORF1 (ORF1) 유전자, 완전한 서열
DQ361268.1	토르크 테노 바이러스 단리주 ViPi04 ORF1 유전자, 완전한 코딩 서열
EF538879.1	토르크 테노 바이러스 단리주 CSC5 Orf2 및 Orf1 유전자, 완전한 코딩 서열
EU305675.1	토르크 테노 바이러스 단리주 LTT7 ORF1 유전자, 완전한 코딩 서열
EU305676.1	토르크 테노 바이러스 단리주 LTT10 ORF1 유전자, 완전한 코딩 서열
EU889253.1	토르크 테노 바이러스 단리주 ViPi08 비기능적 ORF1 유전자, 완전한 서열

[0842]

FJ392105.1	토르크 테노 바이러스 단리주 TW53A25 ORF2 유전자, 부분 코딩 서열; 및 ORF1 유전자, 완전한 코딩 서열
FJ392107.1	토르크 테노 바이러스 단리주 TW53A27 ORF2 유전자, 부분 코딩 서열; 및 ORF1 유전자, 완전한 코딩 서열
FJ392108.1	토르크 테노 바이러스 단리주 TW53A29 ORF2 유전자, 부분 코딩 서열; 및 ORF1 유전자, 완전한 코딩 서열
FJ392111.1	토르크 테노 바이러스 단리주 TW53A35 ORF2 유전자, 부분 코딩 서열; 및 ORF1 유전자, 완전한 코딩 서열
FJ392112.1	토르크 테노 바이러스 단리주 TW53A39 ORF2 유전자, 부분 코딩 서열; 및 ORF1 유전자, 완전한 코딩 서열
FJ392113.1	토르크 테노 바이러스 단리주 TW53A26 ORF2 유전자, 완전한 코딩 서열; 및 비기능적 ORF1 유전자, 완전한 서열
FJ392114.1	토르크 테노 바이러스 단리주 TW53A30 Orf2 및 Orf1 유전자, 완전한 코딩 서열
FJ392115.1	토르크 테노 바이러스 단리주 TW53A31 Orf2 및 Orf1 유전자, 완전한 코딩 서열
FJ392117.1	토르크 테노 바이러스 단리주 TW53A37 ORF1 유전자, 완전한 코딩 서열
FJ426280.1	토르크 테노 바이러스 주 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 zhenjiang, 완전한 게놈
LC381845.1	토르크 테노 바이러스 Human/Japan/KS025/2016 DNA, 완전한 게놈
MH648892.1	아넬로비리대 종 단리주 ctdc048, 완전한 게놈
MH648893.1	아넬로비리대 종 단리주 ctdh007, 완전한 게놈
MH648897.1	아넬로비리대 종 단리주 ctcb038, 완전한 게놈
MH648900.1	아넬로비리대 종 단리주 ctfc019, 완전한 게놈
MH648901.1	아넬로비리대 종 단리주 ctbb022, 완전한 게놈
MH648907.1	아넬로비리대 종 단리주 ctcf040, 완전한 게놈
MH648911.1	아넬로비리대 종 단리주 cthi018, 완전한 게놈
MH648912.1	아넬로비리대 종 단리주 ctea38, 완전한 게놈
MH648913.1	아넬로비리대 종 단리주 ctbg006, 완전한 게놈
MH648916.1	아넬로비리대 종 단리주 ctbg020, 완전한 게놈
MH648925.1	아넬로비리대 종 단리주 ctci019, 완전한 게놈

[0843]

MH648932.1	아넬로비리대 종 단리주 ctid031, 완전한 게놈
MH648946.1	아넬로비리대 종 단리주 ctdb017, 완전한 게놈
MH648957.1	아넬로비리대 종 단리주 ctch017, 완전한 게놈
MH648958.1	아넬로비리대 종 단리주 ctbh011, 완전한 게놈
MH648959.1	아넬로비리대 종 단리주 ctbc020, 완전한 게놈
MH648962.1	아넬로비리대 종 단리주 ctif015, 완전한 게놈
MH648966.1	아넬로비리대 종 단리주 ctei055, 완전한 게놈
MH648969.1	아넬로비리대 종 단리주 ctjg000, 완전한 게놈
MH648976.1	아넬로비리대 종 단리주 ctcj064, 완전한 게놈
MH648977.1	아넬로비리대 종 단리주 ctbj022, 완전한 게놈
MH648982.1	아넬로비리대 종 단리주 ctbf014, 완전한 게놈
MH648983.1	아넬로비리대 종 단리주 ctbd027, 완전한 게놈
MH648985.1	아넬로비리대 종 단리주 ctch016, 완전한 게놈
MH648986.1	아넬로비리대 종 단리주 ctbd020, 완전한 게놈
MH648989.1	아넬로비리대 종 단리주 ctga035, 완전한 게놈
MH648990.1	아넬로비리대 종 단리주 cthf001, 완전한 게놈
MH648995.1	아넬로비리대 종 단리주 ctbd067, 완전한 게놈
MH648997.1	아넬로비리대 종 단리주 ctce026, 완전한 게놈
MH648999.1	아넬로비리대 종 단리주 ctfb058, 완전한 게놈
MH649002.1	아넬로비리대 종 단리주 ctjj046, 완전한 게놈
MH649006.1	아넬로비리대 종 단리주 ctcf030, 완전한 게놈
MH649008.1	아넬로비리대 종 단리주 ctbg025, 완전한 게놈
MH649011.1	아넬로비리대 종 단리주 ctbh052, 완전한 게놈
MH649014.1	아넬로비리대 종 단리주 ctba003, 완전한 게놈
MH649017.1	아넬로비리대 종 단리주 ctbb016, 완전한 게놈
MH649022.1	아넬로비리대 종 단리주 ctch023, 완전한 게놈
MH649023.1	아넬로비리대 종 단리주 ctbd051, 완전한 게놈
MH649028.1	아넬로비리대 종 단리주 ctbf9, 완전한 게놈
MH649038.1	아넬로비리대 종 단리주 ctbi030, 완전한 게놈
MH649039.1	아넬로비리대 종 단리주 ctca057, 완전한 게놈
MH649040.1	아넬로비리대 종 단리주 ctch033, 완전한 게놈
MH649042.1	아넬로비리대 종 단리주 ctjd005, 완전한 게놈
MH649045.1	아넬로비리대 종 단리주 ctdc021, 완전한 게놈
MH649051.1	아넬로비리대 종 단리주 ctdg044, 완전한 게놈
MH649056.1	아넬로비리대 종 단리주 ctcc062, 완전한 게놈
MH649061.1	아넬로비리대 종 단리주 ctid009, 완전한 게놈

[0844]

MH649062.1	아넬로비리대 종 단리주 ctdc018, 완전한 게놈
MH649063.1	아넬로비리대 종 단리주 ctbf012, 완전한 게놈
MH649068.1	아넬로비리대 종 단리주 ctcc066, 완전한 게놈
MH649070.1	아넬로비리대 종 단리주 ctda011, 완전한 게놈
MH649077.1	아넬로비리대 종 단리주 ctbh034, 완전한 게놈
MH649083.1	아넬로비리대 종 단리주 ctdg028, 완전한 게놈
MH649084.1	아넬로비리대 종 단리주 ctii061, 완전한 게놈
MH649085.1	아넬로비리대 종 단리주 cteh021, 완전한 게놈
MH649092.1	아넬로비리대 종 단리주 ctbg012, 완전한 게놈
MH649101.1	아넬로비리대 종 단리주 ctif053, 완전한 게놈
MH649104.1	아넬로비리대 종 단리주 ctei657, 완전한 게놈
MH649106.1	아넬로비리대 종 단리주 ctca015, 완전한 게놈
MH649114.1	아넬로비리대 종 단리주 ctbf050, 완전한 게놈
MH649122.1	아넬로비리대 종 단리주 ctdc002, 완전한 게놈
MH649125.1	아넬로비리대 종 단리주 ctbb15, 완전한 게놈
MH649127.1	아넬로비리대 종 단리주 ctba013, 완전한 게놈
MH649137.1	아넬로비리대 종 단리주 ctbb000, 완전한 게놈
MH649141.1	아넬로비리대 종 단리주 ctbc019, 완전한 게놈
MH649142.1	아넬로비리대 종 단리주 ctid026, 완전한 게놈
MH649144.1	아넬로비리대 종 단리주 ctfj004, 완전한 게놈
MH649152.1	아넬로비리대 종 단리주 ctcj13, 완전한 게놈
MH649156.1	아넬로비리대 종 단리주 ctci006, 완전한 게놈
MH649157.1	아넬로비리대 종 단리주 ctbd025, 완전한 게놈
MH649158.1	아넬로비리대 종 단리주 ctbf005, 완전한 게놈
MH649161.1	아넬로비리대 종 단리주 ctcf045, 완전한 게놈
MH649165.1	아넬로비리대 종 단리주 ctcc29, 완전한 게놈
MH649169.1	아넬로비리대 종 단리주 ctib021, 완전한 게놈
MH649172.1	아넬로비리대 종 단리주 ctbh857, 완전한 게놈
MH649174.1	아넬로비리대 종 단리주 ctbj049, 완전한 게놈
MH649178.1	아넬로비리대 종 단리주 ctfc006, 완전한 게놈
MH649179.1	아넬로비리대 종 단리주 ctbe000, 완전한 게놈
MH649183.1	아넬로비리대 종 단리주 ctbb031, 완전한 게놈
MH649186.1	아넬로비리대 종 단리주 ctcb33, 완전한 게놈
MH649189.1	아넬로비리대 종 단리주 ctcc12, 완전한 게놈
MH649196.1	아넬로비리대 종 단리주 ctci060, 완전한 게놈
MH649199.1	아넬로비리대 종 단리주 ctbb017, 완전한 게놈

[0845]

MH649203.1	아넬로비리대 종 단리주 cthc018, 완전한 게놈
MH649204.1	아넬로비리대 종 단리주 ctbj003, 완전한 게놈
MH649206.1	아넬로비리대 종 단리주 ctbg010, 완전한 게놈
MH649208.1	아넬로비리대 종 단리주 ctid008, 완전한 게놈
MH649209.1	아넬로비리대 종 단리주 ctbg056, 완전한 게놈
MH649210.1	아넬로비리대 종 단리주 ctda001, 완전한 게놈
MH649212.1	아넬로비리대 종 단리주 ctcf004, 완전한 게놈
MH649217.1	아넬로비리대 종 단리주 ctbe029, 완전한 게놈
MH649223.1	아넬로비리대 종 단리주 ctci016, 완전한 게놈
MH649224.1	아넬로비리대 종 단리주 ctce11, 완전한 게놈
MH649228.1	아넬로비리대 종 단리주 ctcf013, 완전한 게놈
MH649229.1	아넬로비리대 종 단리주 ctcb036, 완전한 게놈
MH649241.1	아넬로비리대 종 단리주 ctda027, 완전한 게놈
MH649242.1	아넬로비리대 종 단리주 ctbf003, 완전한 게놈
MH649254.1	아넬로비리대 종 단리주 ctjb007, 완전한 게놈
MH649255.1	아넬로비리대 종 단리주 ctbb023, 완전한 게놈
MH649256.1	아넬로비리대 종 단리주 ctca002, 완전한 게놈
MH649258.1	아넬로비리대 종 단리주 ctcg010, 완전한 게놈
MH649263.1	아넬로비리대 종 단리주 ctgh3, 완전한 게놈
MK012439.1	아넬로비리대 종 단리주 cthe000, 완전한 게놈
MK012440.1	아넬로비리대 종 단리주 ctjd008, 완전한 게놈
MK012448.1	아넬로비리대 종 단리주 ctch012, 완전한 게놈
MK012457.1	아넬로비리대 종 단리주 ctda009, 완전한 게놈
MK012458.1	아넬로비리대 종 단리주 ctcd015, 완전한 게놈
MK012485.1	아넬로비리대 종 단리주 ctfd011, 완전한 게놈
MK012489.1	아넬로비리대 종 단리주 ctba003, 완전한 게놈
MK012492.1	아넬로비리대 종 단리주 ctbb005, 완전한 게놈
MK012493.1	아넬로비리대 종 단리주 ctcj014, 완전한 게놈
MK012500.1	아넬로비리대 종 단리주 ctcb001, 완전한 게놈
MK012504.1	아넬로비리대 종 단리주 ctcj010, 완전한 게놈
MK012516.1	아넬로비리대 종 단리주 ctcf003, 완전한 게놈
NC_038336.1	토르크 테노 바이러스 5 단리주 TCHN-C1 Orf2 및 Orf1 유전자, 완전한 코딩 서열
NC_038338.1	토르크 테노 바이러스 11 단리주 TCHN-D1 Orf2 및 Orf1 유전자, 완전한 코딩 서열
NC_038339.1	토르크 테노 바이러스 13 단리주 TCHN-A Orf2 및 Orf1 유전자, 완전한 코딩 서열
NC_038340.1	토르크 테노 바이러스 20 ORF4, ORF3, ORF2, ORF1 유전자, 완전한 코딩 서열, 클론: SAa-10

[0846]

NC_038341.1	토르크 테노 바이러스 21 단리주 TCHN-B Orf2 및 Orf1 유전자, 완전한 코딩 서열
NC_038342.1	토르크 테노 바이러스 23 ORF2, ORF1 유전자, 완전한 코딩 서열, 단리주: s-TTV CH65-2
NC_038343.1	토르크 테노 바이러스 24 ORF4, ORF3, ORF2, ORF1 유전자, 완전한 코딩 서열, 클론: SAa-01
NC_038344.1	토르크 테노 바이러스 29 ORF2, ORF1, ORF3 유전자, 완전한 코딩 서열, 단리주: TTVyon-KC009
NC_038345.1	토르크 테노 미니 바이러스 10 단리주 LIL-y1 ORF2, ORF1, ORF3, 및 ORF4 유전자, 완전한 코딩 서열
NC_038346.1	토르크 테노 미니 바이러스 11 단리주 LIL-y2 ORF2, ORF1, 및 ORF3 유전자, 완전한 코딩 서열
NC_038347.1	토르크 테노 미니 바이러스 12 단리주 LIL-y3 ORF2, ORF1, ORF3, 및 ORF4 유전자, 완전한 코딩 서열
NC_038350.1	토르크 테노 미디 바이러스 3 단리주 2PoSMA Orf2 및 Orf1 유전자, 완전한 코딩 서열
NC_038351.1	토르크 테노 미디 바이러스 4 단리주 6PoSMA ORF2, ORF1, 및 ORF3 유전자, 완전한 코딩 서열
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

[0847]

[0848] 일부 구현예에서, 유전 요소는 하나 이상의 비-아넬로바이러스, 예를 들어, 모노드나바이러스, 예를 들어, 쇼토쿠바이러스(예를 들어, 크레스드나비리코타[예를 들어, 레돈도바이러스, 서코바이러스{예를 들어, 돼지 서코바이러스, 예를 들어, PCV-1 또는 PCV-2; 또는 부리 깃털 질환 바이러스}, 제미니바이러스{예를 들어, 토마토 골든 모자이크 바이러스}, 또는 나노바이러스{예를 들어, BBTV, MDV1, SCSVF, 또는 FBNYV}]), 또는 파르보바이러스(예를 들어, 디펜도파라바이러스, 예를 들어, 보카바이러스 또는 AAV) 유래의 하나 이상의 서열에 대하여 상동성 또는 동일성을 갖는 하나 이상의 서열을 포함한다. 일부 구현예에서, 재조합 바이러스는 결함을 갖기 때문에, 감염성 입자를 생성하도록 지원이 제공될 수 있다. 이러한 지원은, 예를 들어, LTR 내의 조절 서열의 제어 하에 바이러스의 하나 이상의 유전자(예를 들어, Rep 유전자 및/또는 구조적 유전자)를 인코딩하는 플라스미드를 함유하는 헬퍼 세포주를 사용함으로써 제공될 수 있다. 본 명세서에 기재된 아넬로벡터를 복제하는 데 적합한 세포주는, 예를 들어, 본 명세서에 기재된 바와 같이 변형될 수 있는, 본 명세서에 기재된 바와 같은 숙주 세포주를 포함한다. 상기 유전 요소는 선택 가능한 마커를 인코딩하는 유전자를 추가로 함유하여, 요망되는 유전 요소가 확인되게 할 수 있다.

[0849] 일부 구현예에서, 유전 요소는 서열이 제1 뉴클레오티드 서열에 의해 인코딩된 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 비활성화시키지 않는 것으로 인식되는 특정 보존적 아미노산 치환이 이루어질 수 있다: 예컨대 양으로 하전된 아미노산(및 이의 역)에 관하여, 라이신, 아르기닌 및 히스티딘; 음으로 하전된 아미노산(및 이의 역)에 관하여, 아스파르트산 및 글루탐산; 및 특정 군의 중성으로 하전된 아미노산(및 모든 경우에, 또한 이의 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 당업계에 인식된다.

[0850] 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 매개변수와 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/ 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. 본 명세서에 기재된 바와 같이, 알고리즘은 갭 등을 설명한다. 동일성은 적어도 약 10개 아미노산 또는 뉴클레오티드 길이, 약 15개 아미노산 또는 뉴클레오티드 길이, 약 20개 아미노산 또는 뉴클레오티드 길이, 약 25개 아미노산 또는 뉴클레오티드 길이, 약 30개 아미노산 또는 뉴클레오티드 길이, 약 35개 아미노산 또는 뉴클레오티드 길이, 약 40개 아미노산 또는 뉴클레오티드 길이, 약 45개 아미노산 또는 뉴클레오티드 길이, 약 50개 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.

[0851] *단백질성 외부*

[0852] 일부 구현예에서, 아넬로벡터, 예를 들어, 합성 아넬로벡터는 유전 요소를 봉입하는 단백질성 외부를 포함한다. 단백질성 외부는 포유동물에서 원치 않는 면역 반응을 유도하지 못하는 실질적으로 비-병원성인 외부 단백질을 포함할 수 있다. 아넬로벡터의 단백질성 외부는 통상적으로 단백질성 외부를 구성하는 정20면체 형성으로 자가-조립될 수 있는 실질적으로 비-병원성인 단백질을 포함한다.

[0853] 일부 구현예에서, 단백질성 외부 단백질은 아넬로벡터의 유전 요소의 서열에 의해 인코딩된다(예를 들어, 유전 요소와 시스로 존재함). 다른 구현예에서, 단백질성 외부 단백질은 아넬로벡터의 유전 요소와 별개인 핵산에 의해 인코딩된다(예를 들어, 유전 요소와 트랜스로 존재함).

[0854] 일부 구현예에서, 단백질, 예를 들어, 실질적으로 비-병원성인 단백질 및/또는 단백질성 외부 단백질은 하나 이상, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 그 이상의 글리코실화된 아미노산을 포함한다.

[0855] 일부 구현예에서, 단백질, 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 폴리아르기닌 서열, 가변 영역, C-말단 폴리글루타민/글루탐산염 서열, 및 하나 이상의 이황화 가교를 포함한다.

[0856] 일부 구현예에서, 단백질은 캡시드 단백질이며, 예를 들어, 본 명세서에 기재된 캡시드 단백질을 인코딩하는 뉴클레오티드 서열, 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 아넬로바이러스 ORF1 분자 및/또는 캡시드 단백질 서열 중 어느 하나에 의해 인코딩된 단백질에 대하여 적어도 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 갖는다. 일부 구현예에서, 캡시드 단백질의 단백질 또는 기능적 단편은, 예를 들어, 본 명세서에 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.

[0857] 일부 구현예에서, 아넬로벡터는 캡시드 단백질 또는 캡시드 단백질의 기능적 단편 또는 본 명세서에 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 인코딩하는 뉴클레오티드를 포함한다.

[0858] 일부 구현예에서, 더 낮은 서열 동일성을 갖는 아미노산의 범위는 본 명세서에 기재된 특성 중 하나 이상 및 세포/조직/종 특이성(예를 들어, 향성)의 차이를 제공할 수 있다.

[0859] 일부 구현예에서, 아넬로벡터는 단백질성 외부 내에 지질이 결여된다. 일부 구현예에서, 아넬로벡터에는 지질 이중층, 예를 들어, 바이러스 외피가 결여된다. 일부 구현예에서, 아넬로벡터의 내부는 단백질성 외부에 의해 완전히 덮인다(예를 들어, 100% 커버리지). 일부 구현예에서, 아넬로벡터의 내부는 단백질성 외부에 의해 100% 미만, 예를 들어, 95%, 90%, 85%, 80%, 70%, 60%, 50% 이하의 커버리지로 덮인다. 일부 구현예에서, 단백질성 외부는 유전 요소가 아넬로벡터에 보유되는 한, 예를 들어, 물, 이온, 펩티드 또는 소분자에 대한 투과 가능성을 허용하는 갭 또는 불연속부를 포함한다.

[0860] 일부 구현예에서, 단백질성 외부는 숙주 세포 내로의 유전 요소의 유입을 매개하기 위하여 숙주 세포를 특이적으로 인식하고/인식하거나 이에 결합하는 하나 이상의 단백질 또는 폴리펩티드, 예를 들어, 상보성 단백질 또는 폴리펩티드를 포함한다.

[0861] 일부 구현예에서, 단백질성 외부는 하기 중 하나 이상을 포함한다: 예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, ORF1 분자의 아르기닌-풍부 영역, 젤리-롤 영역, N22 도메인, 초가변 영역, 및/또는 C-말단 도메인. 일부 구현예에서, 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, 본 명세서에 기재된 바와 같은, 아넬로바이러스 ORF1 핵산에 의해 인코딩된 단백질을 포함한다.

[0862] 일부 구현예에서, 단백질성 외부는 하기의 특성 중 하나 이상을 포함한다: 정20면체 대칭, 하나 이상의 숙주 세포 분자와 상호작용하는 분자를 인식하고/하거나 이에 결합하여, 숙주 세포로의 유입을 매개함, 지질 분자가 결여됨, 탄수화물이 결여됨, pH 및 온도 안정성임, 세제 저항성임, 및 숙주에서 실질적으로 비-면역원성이거나 비-병원성임.

[0863] 일부 구현예에서, 본 명세서에 기재된 바와 같은 단백질성 외부를 포함하는 제1 복수의 아넬로벡터가 대상체에게 투여된다. 일부 구현예에서, 본 명세서에 기재된 단백질성 외부를 포함하는 제2 복수의 아넬로벡터는 제1 복수의 투여 후 대상체에게 후속적으로 투여된다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터와 동일한 단백질성 외부를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터의 단백질성 외부에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 아미노산 서열 동일성을 갖는 단백질성 외부를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터의 ORF1 분자에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 아미노산 서열 동일성을 가진 ORF1 분자를 포함한다. 일부 구현예에서 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터에 의해 포함되는 ORF1 분자와 동일한 아미노산 서열을 갖는 ORF1 분자를 포함한다. 일부 구현예에서, 제1 복수의 아넬로벡터의 단백질성 외부는 제1 복수의 아넬로벡터의 단백질성 외부의 폴리펩티드, 예를 들어, ORF1 분자에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 아미노산 서열 동일성을 갖는 폴리펩티드, 예를 들어, ORF1 분자를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터의 단백질성 외부는 제1 복수의 아넬로벡터의 단백질성 외부의 폴리펩티드, 예를 들어, 캡시드 단백질에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 아미노산 서열 동일성을 갖는 폴리펩티드, 예를 들어, 캡시드 단백질을 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터와 공통인 적어도 하나의 표면 에피토프를 갖는 단백질성 외부를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터의 ORF1과 공통인 적어도 하나의 표면 에피토프를 갖는 ORF1 분자를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터의 단백질성 외부와 하나 이상의 아미노산 차이(예를 들어, 보존적 돌연변이)를 갖는 단백질성 외부를 포함한다. 일부 구현예에서, 제1 복수의 아넬로벡터의 단백질성 외부에 결합하는 항체, 예를 들어, 대상체 내의 항체는 또한 제2 복수의 아넬로벡터의 단백질성 외부에 결합한다. 일부 구현예에서, 항체는 제1 복수의 아넬로벡터의 단백질성 외부에 제2 복수의 아넬로벡터의 단백질성 외부와 동일한 친화도(예를 들어, 약 90 내지 110%, 예를 들어, 95 내지 105%의 KD를 가짐)로 결합한다.

[0864] 일부 구현예에서, 제1 복수의 아넬로벡터의 단백질성 외부는 제2 복수의 아넬로벡터의 단백질성 외부와 동일한 3차 구조를 포함한다. 일부 구현예에서, 제1 및 제2 복수의 아넬로벡터의 단백질성 외부의 구조, 예를 들어, 3차 구조는 저온 전자 현미경(cryo-EM), X-선 결정학, 또는 핵자기 공명(NMR)을 사용하여 결정될 수 있다. 일부 구현예에서, 제1 복수의 아넬로벡터의 단백질성 외부의 구조는 구조적 정렬 및 단백질 구조에서 원자의 원자 좌표의 측정, 예를 들어, 평균 제곱근 오차(RMSD)의 측정을 사용하여 제2 복수의 아넬로벡터의 단백질성 외부의 구조와 비교된다. 일부 구현예에서, RMSD는 비교되는 구조의 폴리펩티드 사슬의 백본, 비교되는 구조의 폴리펩티드 사슬의 알파 탄소, 또는 비교되는 구조의 모든 원자, 예를 들어, 제1 복수의 아넬로벡터의 단백질성 외부 및 제2 복수의 아넬로벡터의 단백질성 외부에 대하여 계산될 수 있다. 일부 구현예에서, 더 낮은 값, 예를 들어, 5 옹스트롬 이하의 RMSD는 제1 복수의 아넬로벡터의 단백질성 외부와 제2 복수의 아넬로벡터의 단백질성 외부 사이의 구조적 유사성을 나타낸다. 일부 구현예에서, 더 낮은 값, 예를 들어, 3 옹스트롬 이하의 RMSD는 제1 복수의 아넬로벡터의 단백질성 외부와 제2 복수의 아넬로벡터의 단백질성 외부 사이의 높은 구조적 유사성을 나타낸다. 일부 구현예에서, 0 옹스트롬의 RMSD는 2개 단백질이 동일한 구조를 포함하고, 예를 들어, 제1 복수의 아넬로벡터의 단백질성 외부의 구조가 제2 복수의 아넬로벡터의 단백질성 외부와 동일함을 나타낸다.

[0865] **III. 핵산 작제물**

[0866] 본 명세서에 기재된 유전 요소는 핵산 작제물(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 핵산 유전 요소 작제물)에 포함될 수 있다.

[0867] 일 양태에서, 본 발명은 (i) 외부 단백질(예를 들어, 비-병원성 외부 단백질, 예를 들어, 아넬로바이러스 ORF1 분자 또는 이의 스플라이스 변이체 또는 기능적 단편)을 인코딩하는 서열, (ii) 유전 요소를 비-병원성 외부 단백질에 결합시키는 외부 단백질 결합 서열, 및 (iii) 이펙터를 인코딩하는 서열을 포함하는 유전 요소를 포함하는 핵산 유전 요소 작제물을 포함한다.

[0868] 또 다른 양태에서, 본 발명은 (i) 유전 요소를 외부 단백질(예를 들어, 비-병원성 외부 단백질, 예를 들어, 아넬로바이러스 ORF1 분자 또는 이의 스플라이스 변이체 또는 기능적 단편)에 결합시키는 외부 단백질 결합 서열, (ii) 비-아넬로바이러스 서열(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 비-아넬로바이러스 복제 원점), 및 (iii) 이펙터를 인코딩하는 서열을 포함하는 유전 요소를 포함하는 핵산 유전 요소 작제물을 포함한다.

[0869] 유전 요소 또는 유전 요소 내의 임의의 서열은 임의의 적합한 방법을 사용하여 수득될 수 있다. 다양한 재조합 방법, 예를 들어, 바이러스 서열을 보유하는 세포로부터 라이브러리의 스크리닝, 이를 포함하는 것으로 알려져 있는 핵산 작제물로부터 서열의 유도, 또는 표준 기법을 사용한 이를 함유하는 세포 및 조직으로부터의 직접적인 단리가 당업계에 알려져 있다. 대안적으로 또는 조합하여, 유전 요소의 일부 또는 전부는 클로닝보다는 합성에 의해 생성될 수 있다.

[0870] 일부 구현예에서, 핵산 작제물은 조절 요소, 표적 유전자에 상동성인 핵산 서열, 및/또는 생존 가능한 세포 내에서 및/또는 세포내 분자가 표적 세포 내에 존재하는 경우 리포터 분자의 발현을 야기하기 위한 다양한 리포터 작제물을 포함한다.

[0871] 리포터 유전자는 잠재적으로 형질감염된 세포를 확인하고, 조절 서열의 기능성을 평가하는 데 사용된다. 일반적으로, 리포터 유전자는 수용자 유기체 또는 조직에 존재하거나 이에 의해 발현되지 않고, 발현이 다소 용이하게 검출 가능한 특성, 예를 들어, 효소 활성에 의해 나타나는 폴리펩티드를 인코딩하는 유전자이다. 리포터 유전자의 발현은 DNA가 수용자 세포 내로 도입된 후 적합한 시간에 검정된다. 적합한 리포터 유전자는 루시퍼라제, 베타-갈락토시다제, 클로람페니콜 아세틸 트랜스퍼라제, 분비 알칼리성 포스파타제, 또는 녹색 형광 단백질 유전자를 인코딩하는 유전자를 포함할 수 있다(예를 들어, 문헌[Ui-Tei et al., 2000 FEBS Letters 479: 79-82]). 적합한 발현 시스템은 널리 알려져 있으며, 알려진 기법을 사용하여 제조하거나 구매할 수 있다. 일반적으로, 가장 높은 리포터 유전자의 발현 수준을 나타내는 최소 5' 측접 영역을 갖는 작제물은 프로모터로서 확인된다. 이러한 프로모터 영역은 리포터 유전자에 연결되고, 프로모터-구동 전사를 조절하는 능력에 대하여 작용제를 평가하는 데 사용될 수 있다.

[0872] 일부 구현예에서, 핵산 작제물은 실질적으로 비-병원성이고/이거나 숙주 세포 내에 실질적으로 비-통합되거나, 숙주에서 실질적으로 비-면역원성이다.

[0873] 일부 구현예에서, 핵산 작제물은 이중-가닥이다. 일부 구현예에서, 핵산 작제물은 단일-가닥이다. 일부 구현예에서, 핵산 작제물은 환형(예를 들어, 본 명세서에 기재된 바와 같은, 예를 들어, 플라스미드 또는 미니서클)이다. 일부 구현예에서 핵산 작제물은 선형이다.

[0874] 일부 구현예에서, 유전 요소는, 예를 들어, 본 명세서에 기재된 바와 같이, 예를 들어, 숙주 세포에서, 핵산 작제물로부터 생성될 수 있다. 일부 구현예에서, 유전 요소는 Rep 분자(예를 들어, 비-아넬로바이러스 Rep 분자, 예를 들어, AAV Rep 분자, 예를 들어, AAV Rep 단백질, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 폴리펩티드)의 존재 하에 핵산 작제물로부터 생성될 수 있다. 일부 구현예에서, 유전 요소는 아넬로바이러스 Rep 단백질(예를 들어, 본 명세서에 기재된 바와 같은 ORF2 분자)에 의해 핵산 작제물로부터 생성될 수 없다.

[0875] 일부 구현예에서, 핵산 작제물은 표현형, 바이러스 수준, 유전자 발현, 다른 바이러스와의 경쟁, 질환 상태 등 중 하나 이상을 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상 조절하기에 충분한 양으로 존재한다.

[0876] **IV. 조성물**

[0877] 본 명세서에 기재된 아넬로벡터는 또한 예를 들어, 본 명세서에 기재된 바와 같이, 약제학적 부형제와 함께 약제학적 조성물에 포함될 수 있다. 일부 구현예에서, 약제학적 조성물은 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 아넬로벡터를 포함한다. 일부 구현예에서, 약제학적 조성물은 약 10^5 내지

10^{15}, 10^{5} 내지 10^{10}, 또는 10^{10} 내지 10^{15}개의 아넬로벡터를 포함한다. 일부 구현예에서, 약제학적 조성물은 약 10^{8}개(예를 들어, 약 10^{5}, 10^{6}, 10^{7}, 10^{8}, 10^{9}, 또는 10^{10}개) 게놈 당량/㎖의 아넬로벡터를 포함한다. 일부 구현예에서, 약제학적 조성물은, 예를 들어, PCT/US19/65995의 실시예 18의 방법에 따라 결정시, 10^{5} 내지 10^{10}, 10^{6} 내지 10^{10}, 10^{7} 내지 10^{10}, 10^{8} 내지 10^{10}, 10^{9} 내지 10^{10}, 10^{5} 내지 10^{6}, 10^{5} 내지 10^{7}, 10^{5} 내지 10^{8}, 10^{5} 내지 10^{9}, 10^{5} 내지 10^{11}, 10^{5} 내지 10^{12}, 10^{5} 내지 10^{13}, 10^{5} 내지 10^{14}, 10^{5} 내지 10^{15}, 또는 10^{10} 내지 10^{15}개 게놈 당량/㎖의 아넬로벡터를 포함한다. 일부 구현예에서, 약제학적 조성물은 세포당 아넬로벡터에 포함되는 적어도 1, 2, 5 또는 10, 100, 500, 1000, 2000, 5000, 8,000, 1 x 10^{4}, 1 x 10^{5}, 1 x 10^{6}, 1 x 10^{7}개 이상의 카피의 유전 요소를 진핵 세포의 집단에 운반하기에 충분한 아넬로벡터를 포함한다. 일부 구현예에서, 약제학적 조성물은 세포당 아넬로벡터에 포함되는 적어도 약 1 x 10^{4}, 1 x 10^{5}, 1 x 10^{6}, 1 x 또는 10^{7}, 또는 약 1 x 10^{4} 내지 1 x 10^{5}, 1 x 10^{4} 내지 1 x 10^{6}, 1 x 10^{4} 내지 1 x 10^{7}, 1 x 10^{5} 내지 1 x 10^{6}, 1 x 10^{5} 내지 1 x 10^{7}, 또는 1 x 10^{6} 내지 1 x 10^{7}개 카피의 유전 요소를 진핵 세포의 집단에 운반하기에 충분한 아넬로벡터를 포함한다.

[0878] 일부 구현예에서, 약제학적 조성물은 하기의 특성 중 하나 이상을 갖는다: 약제학적 조성물이 약제학적 또는 우수 제조 기준(GMP) 표준을 충족하거나; 약제학적 조성물을 우수 제조 기준(GMP)에 따라 제조하거나; 약제학적 조성물이 사전결정된 기준 값 미만의 병원체 수준을 갖거나, 예를 들어, 실질적으로 병원체가 없거나; 약제학적 조성물이 사전결정된 기준 값 미만의 오염물질 수준을 갖거나, 예를 들어, 실질적으로 오염물질이 없거나; 또는 약제학적 조성물이 낮은 면역원성을 갖거나, 예를 들어, 본 명세서에 기재된 바와 같이, 실질적으로 비-면역원성임.

[0879] 일부 구현예에서, 약제학적 조성물은 임계값 미만의 양의 하나 이상의 오염물질을 포함한다. 약제학적 조성물에서 바람직하게는 배제되거나 최소화되는 예시적인 오염물질은 제한 없이, 숙주 세포 핵산(예를 들어, 숙주 세포 DNA 및/또는 숙주 세포 RNA), 동물-유래 구성성분(예를 들어, 혈청 알부민 또는 트립신), 복제-적격 바이러스, 비-감염성 입자, 유리 바이러스 캡시드 단백질, 외래성 물질 및 응집물을 포함한다. 구현예에서, 오염물질은 숙주 세포 DNA이다. 구현예에서, 조성물은 용량당 약 10 ng 미만의 숙주 세포 DNA를 포함한다. 구현예에서, 조성물 중 숙주 세포 DNA의 수준은 여과 및/또는 숙주 세포 DNA의 효소적 분해에 의해 감소된다. 구현예에서, 약제학적 조성물은 중량 기준 10% 미만(예를 들어, 약 10%, 5%, 4%, 3%, 2%, 1%, 0.5% 또는 0.1% 미만)의 오염물질로 이루어진다.

[0880] 일 양태에서, 본 명세서에 기재된 본 발명은 하기를 포함하는 약제학적 조성물을 포함한다:

[0881] a) (i) 비-병원성 외부 단백질을 인코딩하는 서열, (ii) 유전 요소를 비-병원성 외부 단백질에 결합시키는 외부 단백질 결합 서열, 및 (iii) 조절 핵산을 인코딩하는 서열을 포함하는 유전 요소; 및 유전 요소와 회합된, 예를 들어, 이를 봉입하거나 둘러싸는 단백질성 외부를 포함하는 유전 요소를 포함하는 아넬로벡터; 및

[0882] b) 약제학적 부형제.

[0883] *소낭*

[0884] 일부 구현예에서, 조성물은 운반체 구성성분, 예를 들어, 마이크로입자, 리포좀, 소낭 또는 엑소좀을 추가로 포함한다. 일부 구현예에서, 리포좀은 내부 수성 구획을 둘러싸는 단일- 또는 다중라멜라 지질 이중층 및 상대적으로 불투과성인 외부 친유성 인지질 이중층으로 구성된 구형의 소낭 구조를 포함한다. 리포좀은 음이온성, 중성 또는 양이온성일 수 있다. 리포좀은 일반적으로 생체적합성이고, 비독성이며, 친수성 및 친유성 약물 분자들 다를 운반하고, 혈장 효소에 의한 분해로부터 이들의 카고(cargo)를 보호하고, 생물학적 막을 가로질러 이들의 로드를 수송할 수 있다(예를 들어, 검토를 위하여 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0885] 소낭은 몇몇 상이한 유형의 지질로부터 제조될 수 있지만; 약물 담체로서 리포좀을 생성하는 데 인지질이 가장 흔히 사용된다. 소낭은 제한 없이, DOTMA, DOTAP, DOTIM, DDAB를 단독으로 또는 콜레스테롤과 함께 포함하여, DOTMA와 콜레스테롤, DOTAP와 콜레스테롤, DOTIM과 콜레스테롤, 및 DDAB와 콜레스테롤을 제공할 수 있다. 다층 소낭 지질을 제조하기 위한 방법이 당업계에 알려져 있다(예를 들어, 미국 특허 6,693,086을 참고하며, 다층 소낭 지질 제제에 관한 이의 교시내용은 본 명세서에 참고로 포함됨). 지질 필름이 수용액과 혼합되는 경우 소낭 형성은 자발적일 수 있지만, 이는 또한 균질화기, 초음파처리기, 또는 압출 장치를 사용함으로써 진탕 형태로

힘을 가함으로써 더 신속히 처리될 수 있다(예를 들어, 검토를 위해 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조). 압출된 지질은 압출된 지질 제제에 관한 교시가 본 명세서에 참조로 포함되는 문헌[Templeton et al., Nature Biotech, 15:647-652, 1997]에 기재된 바와 같이 감소하는 크기의 필터를 통해 압출시킴으로써 제조될 수 있다.

[0886] 본 명세서에 기재된 바와 같이, 첨가제를 소낭에 첨가하여, 이들의 구조 및/또는 특성을 변형시킬 수 있다. 예를 들어, 콜레스테롤 또는 스핑고미엘린을 혼합물에 첨가하여, 구조의 안정화 및 내부 카고의 누출의 방지를 도울 수 있다. 추가로, 소낭은 수소화된 난(egg) 포스파티딜콜린 또는 난 포스파티딜콜린, 콜레스테롤 및 디세틸 포스페이트로부터 제조될 수 있다(예를 들어, 검토를 위하여 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조). 또한, 소낭은 합성 동안 또는 그 후에 수용자 세포 상의 반응성 기에 상보성인 반응성 기를 포함하도록 표면 변형될 수 있다. 이러한 반응성 기는 제한 없이, 말레이미드 기를 포함한다. 일 예로서, 소낭은 말레이미드 컨쥬게이트된 인지질, 예컨대 제한 없이 DSPE-MaL-PEG2000을 포함하도록 합성될 수 있다.

[0887] 소낭 형성은 주로 천연 인지질 및 지질, 예컨대 1,2-디스테아로릴-sn-글리세로-3-포스파티딜 콜린(DSPC), 스핑고미엘린, 난 포스파티딜콜린 및 모노시알로강글리오시드로 구성될 수 있다. 인지질만으로 제조된 제형은 혈장에서 덜 안정하다. 그러나, 콜레스테롤을 이용한 지질 막의 조작은 캡슐화된 카고의 신속한 방출을 감소시키거나, 1,2-디올레오일-sn-글리세로-3-포스포에탄올아민(DOPE)은 안정성을 증가시킨다(예를 들어, 검토를 위하여 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0888] 구현예에서, 지질은 지질 마이크로입자를 형성하기 위해 사용될 수 있다. 지질은 DLin-KC2-DMA4, C12-200 및 공지질 디스테로일포스파티딜 콜린을 포함하지만 이로 제한되지 않으며, 콜레스테롤 및 PEG-DMG는 자발적 소낭 형성 절차를 사용하여 제형화될 수 있다(예를 들어, 문헌[Novobrantseva, Molecular Therapy-Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3] 참조). 성분 몰비는 약 50/10/38.5/1.5(DLin-KC2-DMA 또는 C12-200/디스테로일포스파티딜 콜린/콜레스테롤/PEG-DMG)일 수 있다. 테크미라(Tekmira)는 모두가 본 발명에 사용될 수 있고/있거나 조정될 수 있는 지질 마이크로입자 및 지질 마이크로입자 제형의 다양한 양태에 관한, 미국 및 해외에서의 대략 95개의 특허 패밀리의 포트폴리오를 갖는다(예를 들어, 미국 특허 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 및 7,838,658 및 유럽 특허 1766035; 1519714; 1781593 및 1664316 참조).

[0889] 일부 구현예에서, 마이크로입자는 무작위 방식으로 배열된 하나 이상의 고형화된 중합체(들)를 포함한다. 마이크로입자는 생분해성일 수 있다. 생분해성 마이크로입자는, 예를 들어, 제한 없이, 용매 증발, 핫 멜트 마이크로캡슐화, 용매 제거, 및 분무 건조를 포함하는 당업계에 알려져 있는 방법을 사용하여 합성될 수 있다. 마이크로입자를 합성하기 위한 예시적인 방법은 문헌[Bershteyn et al., Soft Matter 4:1787-1787, 2008] 및 US 2008/0014144 A1에 기재되어 있으며, 마이크로입자 합성에 관한 이의 구체적인 교시는 본 명세서에 참조로 포함된다.

[0890] 생분해성 마이크로입자를 형성하기 위해 사용될 수 있는 예시적인 합성 중합체는 제한 없이, 지방족 폴리에스테르, 폴리 (락트산)(PLA), 폴리 (글리콜산)(PGA), 락트산 및 글리콜산의 공중합체(PLGA), 폴리카프로락톤(PCL), 다가무수물(polyanhydride), 폴리(오르토)에스테르, 폴리우레탄, 폴리(부티르산), 폴리(발레르산), 및 폴리(락티드-코-카프로락톤) 및 천연 중합체, 예컨대 알부민, 알긴산염 및 덱스트란 및 셀룰로스를 포함하는 기타 다당류, 콜라겐, 예컨대, 예를 들어, 알킬, 알킬렌과 같은 화학 기의 치환, 부가, 하이드록실화, 산화 및 당업자에 의해 일상적으로 이루어지는 기타 변형을 포함하는 이의 화학적 유도체, 알부민 및 기타 친수성 단백질, 제인 및 기타 프롤라민 및 소수성 단백질, 이의 공중합체 및 혼합물을 포함한다. 일반적으로, 이들 물질은 효소적 가수분해 또는 물에의 노출에 의해, 표면 또는 벌크 침식에 의해 분해된다.

[0891] 마이크로입자의 직경은 0.1 내지 1000 마이크로미터(㎛) 범위이다. 일부 구현예에서, 이들의 직경은 크기가 1 내지 750 ㎛, 또는 50 내지 500 ㎛, 또는 100 내지 250 ㎛의 범위이다. 일부 구현예에서, 이들의 직경은 크기가 50 내지 1000 ㎛, 50 내지 750 ㎛, 50 내지 500 ㎛, 또는 50 내지 250 ㎛의 범위이다. 일부 구현예에서, 이들의 직경은 크기가 .05 내지 1000 ㎛, 10 내지 1000 ㎛, 100 내지 1000 ㎛, 또는 500 내지 1000 ㎛의 범위이다. 일부 구현예에서, 이들의 직경은 약 0.5 ㎛, 약 10 ㎛, 약 50 ㎛, 약 100 ㎛, 약 200 ㎛, 약 300 ㎛, 약 350 ㎛, 약 400 ㎛, 약 450 ㎛, 약 500 ㎛, 약 550 ㎛, 약 600 ㎛, 약 650 ㎛, 약 700 ㎛, 약 750 ㎛, 약 800 ㎛, 약 850 ㎛, 약 900 ㎛, 약 950 ㎛ 또는 약 1000 ㎛이다. 마이크로입자 직경의 맥락에서 사용되는 바와 같이, 용

어 "약"은 언급된 절대값의 +/-5%를 의미한다.

[0892] 일부 구현예에서, 리간드는 입자의 표면 상에 존재하고, 부착될 리간드 상에 존재하는 작용성 화학기(카복실산, 알데히드, 아민, 술프하이드릴 및 하이드록실)를 통해 마이크로입자의 표면에 컨쥬게이트된다. 기능성은, 예를 들어, 마이크로입자의 에멀젼 제조 동안 작용성 화학 기를 갖는 안정화제의 혼입에 의해 마이크로입자 내로 도입될 수 있다.

[0893] 작용성 기를 마이크로입자에 도입하는 또 다른 예는 입자 제조 후 동안, 입자 및 리간드를 동종- 또는 이종 이작용성 가교결합제와 직접 가교결합시킴으로써 이루어진다. 이러한 절차는 적합한 화학물질 및 소정의 부류의 가교결합제(하기에 더 상세히 논의된 바와 같은 CDI, EDAC, 글루타르알데히드 등) 또는 제조 후에 입자 표면의 화학적 변형을 통해 리간드를 입자 표면에 커플링시키는 임의의 다른 가교결합제를 사용할 수 있다. 이는 또한 양친매성 분자, 예컨대 지방산, 지질 또는 기능적 안정화제가 입자 표면에 수동으로 흡수되고 부착되어, 이에 의해 리간드로의 테더링을 위한 작용성 말단 기를 도입할 수 있는 과정을 포함한다.

[0894] 일부 구현예에서, 마이크로입자는 이들의 외부 표면 상에 하나 이상의 표적화 기를 포함하도록 합성되어, 특정 세포 또는 조직 유형(예를 들어, 심근세포)을 표적화할 수 있다. 이들 표적화 기는 제한 없이, 수용체, 리간드, 항체 등을 포함한다. 이들 표적화 기는 세포의 표면 상의 이들의 파트너와 결합한다. 일부 구현예에서, 마이크로입자는 세포 표면을 포함하는 지질 이중층 내로 통합될 것이며, 미토콘드리아는 세포로 운반된다.

[0895] 마이크로입자는 또한, 그들의 최외각 표면 상에 지질 이중층을 포함할 수 있다. 이러한 이중층은 동일하거나 상이한 유형의 하나 이상의 지질로 구성될 수 있다. 예는 제한 없이, 인지질, 예컨대 포스포콜린 및 포스포이노시톨을 포함한다. 구체적인 예는 제한 없이, DMPC, DOPC, DSPC 및 다양한 다른 지질, 예컨대 리포좀에 대하여 본 명세서에 기재된 것들을 포함한다.

[0896] 일부 구현예에서, 담체는 예를 들어, 본 명세서에 기재된 바와 같은 나노입자를 포함한다.

[0897] 일부 구현예에서, 본 명세서에 기재된 소낭 또는 마이크로입자는 진단제로 기능화된다. 진단제의 예는 양전자 방사 단층촬영(PET), 컴퓨터 보조 단층촬영(CAT), 단일 광자 방출 컴퓨터 단층촬영, x-선, 형광투시법 및 자기 공명 영상화(MRI)에 사용되는 구매할 수 있는 영상화제; 및 조영제를 포함하지만, 이에 제한되지 않는다. MRI에서 조영제로서 사용하기에 적합한 물질의 예는 가돌리늄 킬레이트, 및 철, 마그네슘, 망간, 구리 및 크롬을 포함한다.

[0898] *담체*

[0899] 본 명세서에 기재된 조성물(예를 들어, 약제학적 조성물)은 담체를 포함하고/하거나, 이와 함께 제형화되고/되거나, 이 내에서 운반될 수 있다. 일 양태에서, 본 발명은 본 명세서에 기재된 조성물(예를 들어, 본 명세서에 기재된 아넬로벡터, 아넬로바이러스, 또는 유전 요소)을 포함하는(예를 들어, 캡슐화하는) 담체(예를 들어, 소낭, 리포좀, 지질 나노입자, 엑소좀, 적혈구, 엑소좀(예를 들어, 포유동물 또는 식물 엑소좀), 푸소좀)를 포함하는 조성물, 예를 들어, 약제학적 조성물을 포함한다.

[0900] 일부 구현예에서, 본 명세서에 기재된 조성물 및 시스템은 리포좀 또는 다른 유사한 소낭 내에 제형화될 수 있다. 일반적으로, 리포좀은 내부 수성 구획을 둘러싸는 단일- 또는 다중라멜라 지질 이중층 및 상대적으로 불투과성인 외부 친유성 인지질 이중층으로 구성된 구형의 소낭 구조이다. 리포좀은 음이온성, 중성 또는 양이온성일 수 있다. 리포좀은 일반적으로 하기의 특성 중 하나 이상(예를 들어, 전부)을 갖는다: 생체적합성, 비독성, 친수성 및 친유성 약물 분자 둘 다를 운반할 수 있음, 혈장 효소에 의한 분해로부터 그들의 카고를 보호할 수 있음, 및 생물학적 막 및 혈액 뇌 장벽(BBB)을 가로질러 이들의 로드를 수송할 수 있음(예를 들어, 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679]; 및 문헌[Zylberberg & Matosevic. 2016. Drug Delivery, 23:9, 3319-3329, doi: 10.1080/10717544.2016.1177136] 참조).

[0901] 소낭은 몇몇 상이한 유형의 지질로부터 제조될 수 있지만; 약물 담체로서 리포좀을 생성하는 데 인지질이 가장 흔히 사용된다. 다중라멜라 소낭 지질의 제조 방법은 당업계에 알려져 있다(예를 들어, 다중라멜라 소낭 지질 제제에 관한 교시가 본 명세서에 참조로 포함되는 미국 특허 6,693,086 참조). 지질 필름이 수용액과 혼합되는 경우 소낭 형성은 자발적일 수 있지만, 이는 또한 균질화기, 초음파처리기, 또는 압출 장치를 사용함으로써 진탕 형태로 힘을 가함으로써 더 신속히 처리될 수 있다(예를 들어, 검토를 위해 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조). 압출된 지질은, 예를 들어, 문헌[Templeton et al., Nature Biotech, 15:647-652, 1997]에 기재된 바와

같이 감소하는 크기의 필터를 통해 압출시킴으로써 제조될 수 있다.

[0902] 지질 나노입자(LNP)는 본 명세서에 기재된 약제학적 조성물을 위한 생체적합성 및 생분해성 운반 시스템을 제공하는 담체의 또 다른 예이다. 예를 들어, 문헌[Gordillo-Galeano et al. European Journal of Pharmaceutics and Biopharmaceutics. Volume 133, December 2018, Pages 285-308]을 참조한다. 나노구조화된 지질 담체(NLC)는 고체 지질 나노입자(SLN)의 특성을 보유하고, 약물 안정성 및 로딩 능력을 개선시키고, 약물 누출을 방지하는 변형된 고체 지질 나노입자(SLN)이다. 중합체 나노입자(PNP)는 약물 전달의 중요한 구성성분이다. 이들 나노입자는 특정 표적에 대한 약물 전달을 효과적으로 유도하고, 약물 안정성 및 약물의 제어 방출을 개선시킬 수 있다. 리포좀과 중합체를 조합한 새로운 유형의 담체인 지질-중합체 나노입자(PLN)가 또한 사용될 수 있다. 이들 나노입자는 PNP 및 리포좀의 상보적 이점을 가진다. PLN은 코어-셸 구조로 구성되며; 중합체 코어는 안정한 구조를 제공하고, 인지질 셸은 양호한 생체 적합성을 제공한다. 이와 같이, 2가지 구성성분은 약물 캡슐화 효율을 증가시키고, 표면 변형을 용이하게 하고, 수용성 약물의 누출을 방지한다. 검토를 위하여, 예를 들어, 문헌[Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122]을 참조한다.

[0903] 엑소좀도 또한, 본 명세서에 기재된 조성물 및 시스템을 위한 약물 운반 비히클로서 사용될 수 있다. 검토를 위하여, 문헌[Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; doi.org/10.1016/j.apsb.2016.02.001]을 참조한다.

[0904] 생체 외 분화된 적혈구도 또한, 본 명세서에 기재된 조성물을 위한 담체로서 사용될 수 있다. 예를 들어, WO2015073587; WO2017123646; WO2017123644; WO2018102740; WO2016183482; WO2015153102; WO2018151829; WO2018009838; 문헌[Shi et al. 2014. Proc Natl Acad Sci USA. 111(28): 10131-10136]; 미국 특허 9,644,180; 문헌[Huang et al. 2017. Nature Communications 8: 423]; 문헌[Shi et al. 2014. Proc Natl Acad Sci USA. 111(28): 10131-10136]을 참조한다.

[0905] 예를 들어, WO2018208728에 기재된 바와 같은 푸소좀 조성물도 또한 본 명세서에 기재된 조성물을 운반하기 위한 담체로서 사용될 수 있다.

[0906] *막 투과성 폴리펩티드*

[0907] 일부 구현예에서, 조성물은 세포 내로 또는 막, 예를 들어, 세포 또는 핵 막을 가로질러 성분을 운반하도록 막 투과성 폴리펩티드(MPP)를 추가로 포함한다. 막을 가로질러 물질의 수송을 용이하게 할 수 있는 막 투과성 폴리펩티드는 세포-투과 펩티드(CPP)(예를 들어, 미국 특허 제8,603,966호 참조), 식물 세포내 운반을 위한 융합 펩티드(예를 들어, 문헌[Ng et al., PLoS One, 2016, 11:e0154081] 참조), 단백질 변환 도메인, 트로얀(Trojan) 펩티드, 및 막 전위 신호(MTS)(예를 들어, 문헌[Tung et al., Advanced Drug Delivery Reviews 55:281-294 (2003)] 참조)를 포함하지만, 이에 제한되지 않는다. 일부 MPP는 양으로 하전된 측쇄를 갖는 아미노산, 예컨대 아르기닌이 풍부하다.

[0908] 막 투과성 폴리펩티드는 전신 투여 시에 생체 내에서 다중의 조직의 세포 내에서, 구성성분의 막 투과를 유도하고, 거대분자 전위를 가능하게 하는 능력을 갖는다. 막 투과성 폴리펩티드는 또한, 적절한 조건 하에서 세포와 접촉되는 경우, 외부 환경으로부터 세포질, 세포소기관, 예컨대 미토콘드리아 또는 세포의 핵을 포함하는 세포 내 환경에, 수동 확산으로 달성될 것보다 유의미하게 더 큰 양으로 통과하는 펩티드를 지칭할 수 있다.

[0909] 막을 가로질러 수송되는 구성성분은 막 투과성 폴리펩티드와 가역적으로 또는 비가역적으로 연결될 수 있다. 링커는 화학 결합, 예를 들어, 하나 이상의 공유 결합 또는 비공유 결합일 수 있다. 일부 구현예에서, 링커는 펩티드 링커이다. 이러한 링커는 2 내지 30개 아미노산 이상일 수 있다. 링커는 가요성, 강성 또는 절단 가능한 링커를 포함한다.

[0910] *조합*

[0911] 일 양태에서, 본 명세서에 기재된 아넬로벡터 또는 아넬로벡터를 포함하는 조성물은 또한, 하나 이상의 이종성 모이어티를 포함할 수 있다. 일 양태에서, 본 명세서에 기재된 아넬로벡터 또는 아넬로벡터를 포함하는 조성물은 또한, 융합물 내에 하나 이상의 이종성 모이어티를 포함할 수 있다. 일부 구현예에서, 이종성 모이어티는 유전 요소와 연결될 수 있다. 일부 구현예에서, 이종성 모이어티는 아넬로벡터의 일부로서 단백질성 외부에 봉입될 수 있다. 일부 구현예에서, 이종성 모이어티는 아넬로벡터와 함께 투여될 수 있다.

[0912] 일 양태에서, 본 발명은 본 명세서에 기재된 아넬로벡터 및 이종성 모이어티 중 어느 하나를 포함하는 세포 또는 조직을 포함한다.

[0913] 또 다른 양태에서, 본 발명은 본 명세서에 기재된 아넬로벡터 및 이종성 모이어티를 포함하는 약제학적 조성물을 포함한다.

[0914] 일부 구현예에서, 이종성 모이어티는 바이러스(예를 들어, 이펙터(예를 들어, 약물, 소분자)), 표적화제(예를 들어, DNA 표적화제, 항체, 수용체 리간드), 태그(예를 들어, 형광단, 감광성 작용제, 예컨대 킬러레드(KillerRed)), 또는 본 명세서에 기재된 편집 또는 표적화 모이어티일 수 있다. 일부 구현예에서, 본 명세서에 기재된 막 전위 폴리펩티드는 하나 이상의 이종성 모이어티에 연결된다. 일 구현예에서, 이종성 모이어티는 소분자(예를 들어, 펩티드 모방체 또는 2000 달톤 미만의 분자량을 갖는 작은 유기 분자), 펩티드 또는 폴리펩티드(예를 들어, 항체 또는 이의 항원-결합 단편), 나노입자, 앱타머 또는 약리학적 작용제(pharmacoagent)이다.

[0915] 표적화 모이어티

[0916] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 표적화 모이어티, 예를 들어, 표적 세포 상에 존재하는 관심 분자에 특이적으로 결합하는 표적화 모이어티를 추가로 포함할 수 있다. 표적화 모이어티는 관심 분자 또는 세포의 특정 기능을 조절하거나, 특정 분자(예를 들어, 효소, 단백질 또는 핵산), 예를 들어, 경로 내의 관심 분자의 하류의 특정 분자를 조절하거나, 표적에 특이적으로 결합하여, 아넬로벡터 또는 유전 요소를 국재화시킬 수 있다. 예를 들어, 표적화 모이어티는 특정 관심 분자와 상호작용하여 이의 기능을 증가시키거나, 감소시키거나, 또는 다르게 조절하는 치료제를 포함할 수 있다.

[0917] 태그화 또는 모니터링 모이어티

[0918] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 태그를 추가로 포함하여, 본 명세서에 기재된 아넬로벡터 또는 유전 요소를 표지하거나 모니터링할 수 있다. 태그화 또는 모니터링 모이어티는 화학적 작용제 또는 효소적 절단, 예컨대 단백질 분해 또는 인테인 스플라이싱에 의해 제거할 수 있다. 친화성 태그는 친화성 기법을 사용하여 태그화된 폴리펩티드를 정제하는 데 유용할 수 있다. 일부 예에는 키틴 결합 단백질(CBP), 말토스 결합 단백질(MBP), 글루타티온-S-트랜스퍼라제(GST) 및 폴리(His) 태그가 포함된다. 가용화 태그는 샤페론-결함 종, 예컨대 이. 콜라이(*E. coli*)에서 발현되는 재조합 단백질이 단백질에서 적절하게 폴딩되는 것을 지원하는 것을 돕고, 이들이 침전하지 못하게 하는 데 유용할 수 있다. 일부 예는 티오레독신(TRX) 및 폴리(NANP)를 포함한다. 태그화 또는 모니터링 모이어티는 감광성 태그, 예를 들어, 형광을 포함할 수 있다. 형광 태그는 시각화에 유용하다. GFP 및 이의 변이체는 형광 태그로서 흔히 사용되는 몇몇의 예이다. 단백질 태그는 특정 효소적 변형(예컨대, 비오틴 리가제에 의한 비오티닐화) 또는 화학적 변형(예컨대, 형광 영상화를 위한 FlAsH-EDT2와의 반응)이 일어나게 할 수 있다. 종종 태그화 또는 모니터링 모이어티를 조합하여, 단백질을 다수의 다른 구성성분과 연결시킨다. 태그화 또는 모니터링 모이어티는 또한 특정 단백질분해 또는 효소적 절단에 의해(예를 들어, TEV 프로테아제, 트롬빈, 인자 Xa 또는 엔테로펩티다제에 의해) 제거될 수 있다.

[0919] 나노입자

[0920] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 나노입자를 추가로 포함할 수 있다. 나노입자는 약 1 내지 약 1000 나노미터, 약 1 내지 약 500 나노미터 크기, 약 1 내지 약 100 nm, 약 50 nm 내지 약 300 nm, 약 75 nm 내지 약 200 nm, 약 100 nm 내지 약 200 nm, 및 그 사이의 임의의 범위의 크기를 갖는 무기 물질을 포함한다. 나노입자는 일반적으로, 나노규모 치수들의 복합 구조를 갖는다. 일부 구현예에서, 나노입자는 통상적으로 구형이지만, 나노입자 조성에 따라 상이한 형태가 가능하다. 나노입자의 외부 환경과 접촉하는 나노입자의 일부분은 일반적으로 나노입자의 표면으로서 확인된다. 본 명세서에 기재된 나노입자에서, 크기 제한은 2차원으로 제한될 수 있으며, 이에 따라, 나노입자는 약 1 내지 약 1000 nm의 직경을 갖는 복합 구조를 포함하게 되며, 여기서 특정 직경은 나노입자 조성물 및 실험 설계에 따라 의도된 나노입자의 용도에 좌우된다. 예를 들어, 치료적 응용에 사용되는 나노입자는 통상적으로 약 200 nm 이하의 크기를 갖는다.

[0921] 나노입자의 바람직한 추가의 특성, 예컨대 표면 전하 및 입체적 안정화는 또한, 특정 관심 응용을 고려하여 달라질 수 있다. 임상적 응용, 예컨대 암 치료에 바람직할 수 있는 예시적인 특성은 이의 전문이 각각 본 명세서에 참조로 포함되는 문헌[Davis et al, Nature 2008 vol. 7, pages 771-782]; 문헌[Duncan, Nature 2006 vol. 6, pages 688-701]; 및 문헌[Allen, Nature 2002 vol. 2 pages 750-763]에 기재되어 있다. 추가의 특성은 본 개시내용의 판독 시에 당업자에 의해 확인 가능하다. 나노입자 치수 및 특성은 해당 분야에 알려진 기법에 의해 검출될 수 있다. 입자 치수를 검출하기 위한 예시적인 기법은 동적 광 산란(DLS) 및 다양한 현미경관찰, 예컨대 투과 전자 현미경관찰(TEM) 및 원자력 현미경관찰(AFM)을 포함하지만, 이에 제한되지 않는다. 입자 형태를 검출하기 위한 예시적인 기법은 TEM 및 AFM을 포함하지만 이에 제한되지 않는다. 나노입자의 표면 전하를 검출하기

위한 예시적인 기법은 제타 전위 방법을 포함하지만, 이에 제한되지 않는다. 다른 화학적 특성을 검출하는 데 적합한 추가의 기법은 ^{1}H, ^{11}B, 및 ^{13}C 및 ^{19}F NMR, UV/Vis 및 적외선/라만 분광학 및 형광 분광학(나노입자가 형광 표지와 조합하여 사용되는 경우) 및 당업자에 의해 확인 가능한 추가의 기법을 포함한다.

[0922] 소분자

[0923] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 소분자를 추가로 포함할 수 있다. 소분자 모이어티는 작은 펩티드, 펩티드 모방체(예를 들어, 펩토이드), 아미노산, 아미노산 유사체, 합성 폴리뉴클레오티드, 폴리뉴클레오티드 유사체, 뉴클레오티드, 뉴클레오티드 유사체, 일반적으로 몰당 약 5,000 그램 미만의 분자량을 갖는 유기 및 무기 화합물(이종유기 및 유기금속 화합물 포함), 예를 들어, 몰당 약 2,000 그램 미만의 분자량을 갖는 유기 또는 무기 화합물, 예를 들어, 몰당 약 1,000 그램 미만의 분자량을 갖는 유기 또는 무기 화합물, 예를 들어, 몰당 약 500 그램 미만의 분자량을 갖는 유기 또는 무기 화합물, 및 이러한 화합물의 염, 에스테르 및 다른 약제학적으로 허용 가능한 형태를 포함하지만, 이에 제한되지 않는다. 소분자는 신경전달물질, 호르몬, 약물, 독소, 바이러스 또는 미생물 입자, 합성 분자, 및 효능제 또는 길항제를 포함할 수 있지만, 이들에 제한되지 않는다.

[0924] 적합한 소분자의 예는 모두 본 명세서에 참조로 포함되는 문헌["The Pharmacological Basis of Therapeutics," Goodman and Gilman, McGraw-Hill, New York, N.Y., (1996), Ninth edition]의 다음과 같은 섹션에 기재된 것들을 포함한다: 시냅스 및 신경이펙터 연결 부위에서 작용하는 약물; 중추신경계에서 작용하는 약물; 오타코이드: 염증의 약물 요법; 물, 염 및 이온; 신장 기능 및 전해질 대사에 영향을 미치는 약물; 심혈관 약물; 위장 기능에 영향을 미치는 약물; 자궁 운동에 영향을 미치는 약물; 기생충 감염의 화학요법; 미생물 질환의 화학요법; 신생물 질환의 화학요법; 면역억제를 위해 사용되는 약물; 혈액-형성 기관에 작용하는 약물; 호르몬 및 호르몬 길항제; 비타민, 피부과학 및 독성학. 소분자의 일부 예는 프리온 약물, 예컨대 타크롤리무스, 유비퀴틴 리가제 또는 HECT 리가제 저해제, 예컨대 헤클린(heclin), 히스톤 변형 약물, 예컨대 부티르산나트륨, 효소적 저해제, 예컨대 5-아자-시티딘, 안트라사이클린, 예컨대 독소루비신, 베타-락탐, 예컨대 페니실린, 항-박테리아제, 화학요법제, 항-바이러스제, 다른 유기체로부터의 조절제, 예컨대 VP64, 및 불충분한 생체 이용 가능성을 갖는 약물, 예컨대 결함이 있는 약동학을 갖는 화학치료제를 포함하지만, 이에 제한되지 않는다.

[0925] 일부 구현예에서, 소분자는 후생적 변형제, 예를 들어, 문헌[de Groote et al. Nuc. Acids Res. (2012):1-18]에 기재된 것들이다. 예시적인 소분자 후생적 변형제는 예를 들어, 본 명세서에 참조로 포함되는 문헌[Lu et al. J. Biomolecular Screening 17.5(2012):555-71], 예를 들어, 표 1 또는 2에 기재되어 있다. 일부 구현예에서, 후생적 변형제는 보리노스타트(vorinostat) 또는 로미뎁신(romidepsin)을 포함한다. 일부 구현예에서, 후생적 변형제는 부류 I, II, III 및/또는 IV 히스톤 데아세틸라제(HDAC)의 저해제를 포함한다. 일부 구현예에서, 후생적 변형제는 SirTI의 활성화제를 포함한다. 일부 구현예에서, 후생적 변형제는 가르시놀(Garcinol), Lys-CoA, C646, (+)-JQI, I-BET, BICI, MS120, DZNep, UNC0321, EPZ004777, AZ505, AMI-I, 피라졸 아미드 7b, 벤조[d]이미다졸 17b, 아실화된 답손 유도체(예를 들어, PRMTI), 메틸스타트(methylstat), 4,4'-디카복시-2,2'-비피리딘, SID 85736331, 하이드록사메이트 유사체 8, 타닐사이프로미(tanylcypromie), 비스구아니딘 및 비구아니드 폴리아민 유사체, UNC669, 비다자(Vidaza), 데시타빈(decitabine), 소듐 페닐 부티레이트(SDB), 리포산(LA), 케르세틴(quercetin), 발프로산, 하이드랄라진(hydralazine), 박트림(bactrim), 녹차 추출물(예를 들어, 에피갈로카테킨 갈레이트(EGCG)), 커큐민, 술포르판(sulforphane) 및/또는 알리신(allicin)/디알릴 디술피드를 포함한다. 일부 구현예에서, 후생적 변형제는 DNA 메틸화를 저해하며, 예를 들어, DNA 메틸트랜스퍼라제의 저해제이다(예를 들어, 5-아자시티딘이고/이거나 데시타빈이다). 일부 구현예에서, 후생적 변형제는 히스톤 변형, 예를 들어, 히스톤 아세틸화, 히스톤 메틸화, 히스톤 수모화(sumoylation) 및/또는 히스톤 인산화를 변형시킨다. 일부 구현예에서, 후생적 변형제는 히스톤 데아세틸라제의 저해제이다(예를 들어, 보리노스타트 및/또는 트리코스타틴 A이다).

[0926] 일부 구현예에서, 소분자는 약제학적 활성 작용제이다. 일 구현예에서, 소분자는 대사 활성 또는 구성성분의 저해제이다. 유용한 부류의 약제학적 활성 작용제는 항생제, 항-염증 약물, 혈관신생 또는 혈관작용제, 성장 인자 및 화학치료제(항-신생물)제(예를 들어, 종양 억제제)를 포함하지만, 이에 제한되지 않는다. 본 명세서에 기재된 범주 및 예로부터의 또는 문헌[(Orme-Johnson 2007, Methods Cell Biol. 2007;80:813-26)]으로부터의 분자 중 하나 또는 조합이 사용될 수 있다. 일 구현예에서, 본 발명은 항생제, 항-염증 약물, 혈관신생 또는 혈관작용제, 성장 인자 또는 화학치료제를 포함하는 조성물을 포함한다.

[0927] 펩티드 또는 단백질

[0928] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 펩티드 또는 단백질을 추가로 포함할 수 있다. 펩티드 모이어티는 펩티드 리간드 또는 항체 단편(예를 들어, 수용체, 예컨대 세포외 수용체에 결합하는 항체 단편), 신경펩티드, 호르몬 펩티드, 펩티드 약물, 독성 펩티드, 바이러스 또는 미생물 펩티드, 합성 펩티드 및 효능제 또는 길항제 펩티드를 포함할 수 있지만, 이들에 제한되지 않는다.

[0929] 펩티드 모이어티는 선형 또는 분지형일 수 있다. 펩티드는 약 5 내지 약 200개의 아미노산, 약 15 내지 약 150개의 아미노산, 약 20 내지 약 125개의 아미노산, 약 25 내지 약 100개의 아미노산 또는 그 사이의 임의의 범위의 길이를 갖는다.

[0930] 펩티드의 일부 예는 형광 태그 또는 마커, 항원, 항체, 항체 단편, 예컨대 단일 도메인 항체, 리간드 및 수용체, 예컨대 글루카곤-유사 펩티드-1(GLP-1), GLP-2 수용체 2, 콜레시스토키닌(cholecystokinin) B(CCKB) 및 소마토스타틴 수용체, 펩티드 치료제, 예컨대 특정 세포 표면 수용체, 예컨대 G 단백질-커플링된 수용체(GPCR) 또는 이온 채널에 결합되는 것들, 천연적으로 생물활성인 펩티드로부터의 합성 또는 유사체 펩티드, 항-미생물 펩티드, 포어-형성 펩티드, 종양 표적화 또는 세포독성 펩티드, 및 분해 또는 자가-파괴 펩티드, 예컨대 아폽토시스-유도 펩티드 신호 또는 감광제 펩티드를 포함하지만 이들에 제한되지 않는다.

[0931] 본 명세서에 기재된 본 발명에 유용한 펩티드는 또한, 작은항원-결합 펩티드, 예를 들어, 항원 결합 항체 또는 항체-유사 단편, 예컨대 단일 쇄 항체, 나노바디를 포함한다(예를 들어, 문헌[Steeland et al. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. Drug Discov Today: 21(7):1076-113] 참조). 이러한 작은 항원 결합 펩티드는 사이토졸 항원, 핵 항원, 세포소기관-내 항원에 결합할 수 있다.

[0932] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 특정 위치, 조직, 또는 세포를 표적화할 수 있는 리간드에 연결된 폴리펩티드를 포함한다.

[0933] 올리고뉴클레오티드 앱타머

[0934] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 올리고뉴클레오티드 앱타머를 추가로 포함할 수 있다. 앱타머 모이어티는 올리고뉴클레오티드 또는 펩티드 앱타머이다. 올리고뉴클레오티드 앱타머는 단백질 및 펩티드를 포함하는 사전 선택된 표적에 높은 친화도 및 특이성으로 결합할 수 있는 단일-가닥 DNA 또는 RNA(ssDNA 또는 ssRNA) 분자이다.

[0935] 올리고뉴클레오티드 앱타머는 다양한 분자 표적, 예컨대 소분자, 단백질, 핵산, 및 심지어 세포, 조직 및 유기체에 결합하기 위해 반복된 라운드의 시험관 내 선택 또는 동등하게 SELEX(systematic evolution of ligands by exponential enrichment; 지수적 농축에 의한 리간드의 계통적 진화)를 통해 조작될 수 있는 핵산 종이다. 앱타머는 차별적인 분자 인식을 제공하며, 화학적 합성에 의해 생성될 수 있다. 추가적으로, 앱타머는 바람직한 저장 특성을 가지며, 치료적 응용에서 면역원성을 거의 또는 전혀 유발하지 않을 수 있다.

[0936] DNA 및 RNA 앱타머는 둘 다 다양한 표적에 대하여 강력한 결합 친화성을 나타낼 수 있다. 예를 들어, DNA 및 RNA 앱타머는 t 라이소자임, 트롬빈, 인간 면역결핍 바이러스 트랜스-작용 반응성 요소(HIV TAR)(en.wikipedia.org/wiki/Aptamer - cite_note-10 참조), 헤민(hemin), 인터페론 γ 혈관 내피 성장 인자(VEGF), 전립선 특이적 항원(PSA), 도파민 및 비-정규 종양유전자, 열 충격 인자 1(HSF1)을 위하여 선택된다.

[0937] 펩티드 앱타머

[0938] 일부 구현예에서, 본 명세서에 기재된 조성물 또는 아넬로벡터는 펩티드 앱타머를 추가로 포함할 수 있다. 펩티드 앱타머는 12 내지 14 kDa인 저분자량을 갖는 펩티드를 포함하여, 하나(또는 그 이상)의 짧은 가변 펩티드 도메인을 갖는다. 펩티드 앱타머는 세포에 특이적으로 결합하고 세포 내부의 단백질-단백질 상호작용을 간섭하도록 설계될 수 있다.

[0939] 펩티드 앱타머는 특정 표적 분자에 결합하도록 선택되거나 조작된 인공 단백질이다. 이들 단백질은 가변 서열의 하나 이상의 펩티드 루프를 포함한다. 이들 단백질은 통상적으로 조합 라이브러리로부터 단리되며, 종종 유도된 돌연변이 또는 가변 영역 돌연변이유발 및 선택 라운드에 의해 후속적으로 개선된다. 생체 내에서, 펩티드 앱타머는 세포 단백질 표적에 결합할 수 있고, 이들의 표적화된 분자와 다른 단백질의 정상적인 단백질 상호작용을 방해하는 것을 포함하여, 생물학적 효과를 발휘할 수 있다. 구체적으로, 전사 인자 결합 도메인에 부착된 가변 펩티드 앱타머 루프는 전사 인자 활성화 도메인에 부착된 표적 단백질에 대해 스크리닝된다. 생체 내에서 이러한 선택 전략을 통해 표적에 대한 펩티드 앱타머의 결합은 하류 효모 마커 유전자의 발현으로 검출된다. 이러한 실험은 표현형을 유발하기 위해 앱타머에 의해 결합되는 특정 단백질 및 앱타머가 파괴하는 단백질 상호작용을

확인한다. 추가적으로, 적절한 기능성 모이어티로 유도체화된 펩티드 앱타머는 표적 단백질의 특정 번역후 변형을 유발하거나 표적의 세포내 국재화를 변경할 수 있다.

[0940] 펩티드 앱타머는 또한 시험관 내에서 표적을 인식할 수 있다. 펩티드 앱타머는 바이오센서에서 항체 대신에 사용되는 것으로 확인되었으며, 비활성 및 활성 단백질 형태 둘 다를 포함하는 집단에서 단백질의 활성 이소형(isoform)을 검출하는 데 사용되었다. 펩티드 앱타머 "헤드"가 독특한 서열의 이중-가닥 DNA "테일(tail)"에 공유적으로 연결된 태드폴(tadpole)로서 알려져 있는 유도체는 (예를 들어, 정량적 실시간 중합효소 연쇄 반응을 사용하여) 이들의 DNA 테일의 PCR에 의해 혼합물 중 드문 표적 분자의 정량화를 가능하게 한다.

[0941] 펩티드 앱타머 선택은 상이한 시스템을 사용하여 이루어질 수 있지만, 현재 효모 2-하이브리드 시스템이 가장 많이 사용된다. 펩티드 앱타머는 또한 파지 디스플레이 및 다른 표면 디스플레이 기술, 예컨대 mRNA 디스플레이, 리보솜 디스플레이, 박테리아 디스플레이 및 효모 디스플레이에 의해 구축된 조합 펩티드 라이브러리로부터 선택될 수 있다. 이들 실험 절차는 또한 바이오패닝(biopanning)으로서 알려져 있다. 바이오패닝으로부터 수득되는 펩티드 중에, 미모토프(mimotope)는 펩티드 앱타머의 한 종류로서 간주될 수 있다. MimoDB 조합 펩티드 라이브러리로부터 패닝된 모든 펩티드는 명칭이 MimoDB인 특수 데이터에 저장되어 있다.

[0942] **VI. 사용 방법**

[0943] 본 명세서에 기재된 아넬로벡터 및 아넬로벡터를 포함하는 조성물은, 예를 들어, 질환, 장애 또는 병태의 치료를 필요로 하는 대상체(예를 들어, 포유동물 대상체, 예를 들어, 인간 대상체)에서의 질환, 장애 또는 병태의 치료 방법에 사용될 수 있다. 본 명세서에 기재된 약제학적 조성물의 투여는, 예를 들어, 비경구(정맥내, 종양내, 복강내, 근육내, 강내 및 피하를 포함함) 투여로 이루어질 수 있다. 아넬로벡터는 단독으로 투여되거나 약제학적 조성물로 제형화될 수 있다. 일부 구현예에서, 아넬로벡터는 단일 용량, 예를 들어, 제1 복수로 투여될 수 있다. 일부 구현예에서, 아넬로벡터는 적어도 2개의 용량, 예를 들어, 제1 복수에 이어서 제2 복수로 투여될 수 있다. 일부 구현예에서, 아넬로벡터는 다중 용량, 예를 들어, 제1 복수, 제2 복수, 제3 복수, 선택적으로 제4 복수, 선택적으로 제5 복수, 및/또는 선택적으로 추가 복수의 다중 용량으로 투여될 수 있다.

[0944] 아넬로벡터는 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.

[0945] 일부 구현예에서, 예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터 또는 이를 포함하는 조성물의 투여는 아넬로벡터에 의해 포함되는 유전 요소가, 예를 들어, 대상체 내의 표적 세포로 운반되게 할 수 있다.

[0946] 예를 들어, 이펙터(예를 들어, 내인성 또는 외인성 이펙터)를 포함하는 본 명세서에 기재된 아넬로벡터 또는 이의 조성물을 사용하여 이펙터를 세포, 조직 또는 대상체에 운반할 수 있다. 일부 구현예에서, 아넬로벡터 또는 이의 조성물을 사용하여, 이펙터를 골수, 혈액, 심장, GI 또는 피부에 운반한다. 본 명세서에 기재된 아넬로벡터 조성물의 투여에 의한 이펙터의 운반은 세포, 조직, 또는 대상체에서 비코딩 RNA 또는 폴리펩티드의 발현 수준을 조절할 수 있다(예를 들어, 증가시키거나 감소시킬 수 있다). 이러한 방식으로의 발현 수준의 조절은 이펙터가 운반되는 세포에서 기능적 활성의 변경을 초래할 수 있다. 일부 구현예에서, 조절된 기능적 활성은 효소적, 구조적 또는 조절적 성질을 가질 수 있다.

[0947] 일부 구현예에서, 아넬로벡터 또는 이의 카피는 세포로의 운반 후 24시간(예를 들어, 1일, 2일, 3일, 4일, 5일, 6일, 1주, 2주, 3주, 4주, 30일 또는 1개월)에 세포에서 검출 가능하다. 구현예에서, 아넬로벡터 또는 이의 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, 아넬로벡터 또는 이의 조성물이 외인성 단백질을 인코딩하는 유전 요소를 포함하는) 일부 구현예에서, 효과는 1, 2, 3, 4, 5, 6 또는 7일, 2, 3 또는 4주 또는 1, 2, 3, 6 또는 12개월 미만 동안 지속된다.

[0948] 본 명세서에 기재된 아넬로벡터, 또는 아넬로벡터를 포함하는 조성물로 치료될 수 있는 질환, 장애, 및 병태의 예는 제한 없이, 면역 장애, 인터페론병증(예를 들어, I형 인터페론병증), 감염성 질환, 염증성 장애, 자가면역 병태, 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 구현예에서, 아넬로벡터는 아넬로벡터가 접촉되는 세포 내에서 활성 또는 기능을 조절한다(예를 들어, 증가시키거나 감소시킨다). 일부 구현예에서, 아넬로벡터는 아넬로벡터가 접촉되는 세포에서 분자(예를 들어, 핵산 또는 단백질)의 수준 또는 활성을 조절한다(예를 들어, 증가시키거나 감소시킨다). 일부 구현예에서, 아넬로벡터는 아넬로벡터가 접촉되는 세포, 예를

들어, 암 세포의 생존력을, 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 또는 그 이상 감소시킨다. 일부 구현예에서, 아넬로벡터는 아넬로벡터가 접촉되는 세포, 예를 들어, 암 세포의 생존력을, 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 또는 그 이상 감소시키는 이펙터, 예를 들어, miRNA, 예를 들어, miR-625를 포함한다. 일부 구현예에서, 아넬로벡터는, 예를 들어, 카스파제-3 활성을 증가시킴으로써, 아넬로벡터가 접촉되는 세포, 예를 들어, 암 세포의 아폽토시스를, 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 또는 그 이상 증가시킨다. 일부 구현예에서, 아넬로벡터는, 예를 들어, 카스파제-3 활성을 증가시킴으로써, 아넬로벡터가 접촉되는 세포, 예를 들어, 암 세포의 아폽토시스를, 예를 들어, 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 또는 그 이상 증가시키는 이펙터, 예를 들어, miRNA, 예를 들어, miR-625를 포함한다.

[0949] **VII. 투여/운반**

[0950] 조성물(예를 들어, 본 명세서에 기재된 바와 같은 아넬로벡터를 포함하는 약제학적 조성물)은 약제학적으로 허용 가능한 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-무함유일 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](본 명세서에 참조로 포함됨)에서 찾을 수 있다.

[0951] 본 명세서에 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/설계하거나 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 래트; 및/또는 조류, 예컨대 상업적으로 관련된 조류, 예컨대 가금류, 닭, 오리, 거위 및/또는 칠면조를 포함하지만, 이에 제한되지 않는다.

[0952] 일부 구현예에서, 약제학적 조성물의 투여가 고려되는 대상체는 인간이다. 일부 구현예에서, 대상체는, 예를 들어, 0 내지 4주령의 신생아이다. 일부 구현예에서, 대상체는, 예를 들어, 4주령 내지 1세의 영아이다. 일부 구현예에서, 대상체는, 예를 들어, 1세 내지 12세의 어린이이다. 일부 구현예에서, 대상체는 18세 미만이다. 일부 구현예에서, 대상체는, 예를 들어, 12세 내지 18세의 청소년이다. 일부 구현예에서, 대상체는 18세 초과이다. 일부 구현예에서, 대상체는, 예를 들어, 18세 내지 25세의 청년이다. 일부 구현예에서, 대상체는, 예를 들어, 25세 내지 50세의 성인이다. 일부 구현예에서, 대상체는 노인, 예를 들어, 적어도 50세 이상의 성인이다.

[0953] 본 명세서에 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.

[0954] 일 양태에서, 본 발명은 대상체로 아넬로벡터를 운반하는 방법을 특징으로 한다. 방법은 본 명세서에 기재된 바와 같은 아넬로벡터를 포함하는 약제학적 조성물을 대상체에게 투여하는 단계를 포함한다. 일부 구현예에서, 투여된 아넬로벡터는 대상체에서 복제한다(예를 들어, 대상체의 바이러스체의 일부가 됨).

[0955] 약제학적 조성물은 야생형 또는 고유 바이러스 요소 및/또는 변형된 바이러스 요소를 포함할 수 있다. 아넬로벡터는 하나 이상의 아넬로바이러스 서열(예를 들어, 핵산 서열 또는 이의 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 CAV 서열을 포함할 수 있다. 아넬로벡터는 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. 아넬로벡터는 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. 아넬로벡터는 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.

[0956] 일부 구현예에서, 아넬로벡터는 내인성 유전자 및 단백질 발현을 기준, 예를 들어, 건강한 대조군과 비교하여, 예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상 증가(자극)시키기에 충분하다. 특정 구현예에서, 아넬로벡터는 내인성 유전자 및 단백질 발현을 기준, 예를 들어, 건강한 대조군과 비교하여, 예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상 감소(저해)시키기에 충분하다.

[0957] 일부 구현예에서, 아넬로벡터는 숙주 또는 숙주 세포에서 하나 이상의 바이러스 특성, 예를 들어, 향성, 감염성, 면역억제/활성화를 기준, 예를 들어, 건강한 대조군과 비교하여, 예를 들어, 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상 저해/향상시킨다.

[0958] 일 양태에서, 본 발명은 이전에 아넬로벡터, 예를 들어, 제1 복수의 아넬로벡터를 투여받은 대상체, 예를 들어, 인간 대상체에게 이펙터를 운반하는 방법을 특징으로 하며, 방법은 제2 복수의 아넬로벡터의 투여를 포함한다. 또 다른 양태에서, 본 발명은 대상체, 예를 들어, 인간 대상체에게 이펙터를 운반하는 방법을 특징으로 하며, 방법은 대상체에게 제1 복수의 아넬로벡터를 투여하는 단계 및 대상체에게 제2 복수의 아넬로벡터를 후속적으로 투여하는 단계를 포함한다. 일부 구현예에서, 본 명세서에 기재된 방법은 제3, 제4, 제5, 및/또는 추가 복수의 아넬로벡터의 투여를 추가로 포함한다. 일부 구현예에서, 제1 및 제2 복수는 동일한 투여 경로, 예를 들어, 정맥내 투여를 통해 투여된다. 일부 구현예에서, 제1 및 제2 복수는 상이한 투여 경로를 통해 투여된다. 일부 구현예에서, 제1 복수의 아넬로벡터는 제1 약제학적 조성물의 일부로서 대상체에게 투여된다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제2 약제학적 조성물의 일부로서 대상체에게 투여된다.

[0959] 일부 구현예에서, 제1 및 제2 복수는 거의 동일한 투약량의 아넬로벡터를 포함하며, 예를 들어, 제1 복수 및 제2 복수의 아넬로벡터는 거의 동일한 양 및/또는 농도의 아넬로벡터를 포함한다. 일부 구현예에서, 제2 복수는 90 내지 110%, 예를 들어, 95 내지 105%의 수의 제1 복수의 아넬로벡터를 포함한다. 일부 구현예에서, 제1 복수는 제2 복수보다 더 많은 투약량의 아넬로벡터를 포함하며, 예를 들어, 제1 복수는 제2 복수에 비해 더 많은 양 및/또는 더 큰 농도의 아넬로벡터를 포함한다. 일부 구현예에서, 제1 복수는 제2 복수보다 저 적은 투약량의 아넬로벡터를 포함하며, 예를 들어, 제1 복수는 제2 복수에 비해 더 많은 양 및/또는 더 작은 농도의 아넬로벡터를 포함한다. 일부 구현예에서, 대상체는 반복 용량의 아넬로벡터를 받으며, 반복 용량은 적어도 1, 2, 3, 4, 또는 5년 동안 투여된다. 일부 구현예에서, 반복 용량은 약 1, 2, 3, 또는 4주마다, 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 또는 12개월마다 투여된다.

[0960] 일부 구현예에서, 대상체에게 투여된 제1 복수의 아넬로벡터에 포함된 유전 요소는 이의 투여 후 적어도 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 또는 150일 후에, 예를 들어, 실시예 1에 기재된 바와 같이, 예를 들어, 고해상도 용융(HRM) 분석에 의해, 대상체에서 검출 가능하다. 일부 구현예에서, 대상체에게 투여된 제2 복수의 아넬로벡터에 포함된 유전 요소는 이의 투여 후 적어도 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 또는 150일 후에, 예를 들어, 실시예 1에 기재된 바와 같이, 예를 들어, 고해상도 용융(HRM) 분석에 의해, 대상체에서 검출 가능하다.

[0961] 일부 구현예에서, 대상체에게 투여되는 제1 및/또는 제2 복수의 아넬로벡터는 이펙터를 포함한다. 일부 구현예에서, 제1 및/또는 제2 복수는 외인성 이펙터를 포함한다. 일부 구현예에서, 제1 및/또는 제2 복수는 내인성 이펙터를 포함한다. 일부 구현예에서, 제2 복수의 아넬로벡터의 이펙터는 제1 복수의 아넬로벡터의 이펙터와 동일한 이펙터이다. 일부 구현예에서, 제2 복수의 아넬로벡터의 이펙터는 제1 복수의 아넬로벡터의 이펙터와 상이하다. 일부 구현예에서, 제2의 복수의 아넬로벡터는 제1의 복수의 아넬로벡터에 의해 운반되는 이펙터의 수와 거의 동일한 수의 이펙터 카피를 대상체에게 운반한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터에 의해 대상체에게 운반되는 이펙터 카피의 적어도 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 수준으로 이펙터를 대상체에게 전달한다(예를 들어, 제1 복수에 의해 운반되는 이펙터는 제2 복수에 의해 운반되는 이펙터와 동일하거나 상이할 수 있음). 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터보다 더 많은 카피(예를 들어, 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 500, 또는 1000배만큼 많은 카피)의 이펙터를 대상체에게 운반한다. 일부 구현예에서, 제2 복수의 아넬로벡터는 제1 복수의 아넬로벡터 투여의 생물학적 효과 이상인 대상체에 대한 생물학적 효과(예를 들어, 표적 유전자의 녹다운, 또는 바이오마커의 상향조절)를 갖는다.

[0962] 일부 구현예에서, 복수의 아넬로벡터를 수용한 것에 기초하여 대상체를 확인하거나 선택하는 것은 대상체 유래의 샘플에 대한 분석을 수행하는 것을 포함한다. 일부 구현예에서, 복수의 아넬로벡터를 수용한 것에 기초하여 대상체를 확인하거나 선택하는 것은 제3자(예를 들어, 실험실)로부터 정보를 수득하는 것을 포함하며, 여기서

제3자는 대상체 유래의 샘플에 대하여 분석을 수행하였다. 일부 구현예에서, 복수의 아넬로벡터를 수용한 것에 기초하여 대상체를 확인하거나 선택하는 것은 대상체의 병력을 검토하는 것을 포함한다.

[0963] 일부 구현예에서, 바이러스 유전 정보에서 표현되지 않는 하나 이상의 바이러스 주를 추가로 포함하는 약제학적 조성물이 대상체에게 투여된다.

[0964] 일부 구현예에서, 본 명세서에 기재된 아넬로벡터를 포함하는 약제학적 조성물은 바이러스 감염을 조절하기에 충분한 용량 및 시간으로 투여된다. 바이러스 감염의 일부 비-제한적인 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), 코사바이러스(Cosavirus) A, 우두 바이러스, 콕사키바이러스(Coxsackievirus), 크리민-콩고 출혈성 열 바이러스(Crimean-Congo hemorrhagic fever virus), 뎅기(Dengue) 바이러스, 도리(Dhori) 바이러스, 더그비(Dugbe) 바이러스, 두벤하게(Duvenhage) 바이러스, 동부 말 뇌염 바이러스, 에볼라바이러스, 에코바이러스(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 아데노바이러스, 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, 오뇽-뇽(O'nyong-nyong) 바이러스, 올프(Orf) 바이러스, 오로퓨스(Oropouche) 바이러스, 피친데(Pichinde) 바이러스, 폴리오바이러스, 푼타 토로 플레보바이러스(Punta toro phlebovirus), 푸말라(Puumala) 바이러스, 광견병 바이러스, 리프트계곡열 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, 셈리키 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, 베네주엘란(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 구현예에서, 아넬로벡터는 대상체에 이미 존재하는 바이러스를, 예를 들어, 기준 대상과 비교하여 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상 능가하고/능가하거나 대체하기에 충분하다. 특정 구현예에서, 아넬로벡터는 만성 또는 급성 바이러스 감염과 경쟁하기에 충분하다. 특정 구현예에서, 아넬로벡터는 바이러스 감염으로부터 보호하기 위하여 예방적으로 투여될 수 있다(예를 들어, 바이러스전(provirotic)). 일부 구현예에서, 아넬로벡터는 (예를 들어, 표현형, 바이러스 수준, 유전자 발현, 다른 바이러스와의 경쟁, 질환 상태 등을 적어도 약 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 또는 그 이상) 조절하기에 충분한 양으로 존재한다. 일부 구현예에서, "치료", "치료하는" 및 이의 동족어는 (예를 들어, 아넬로벡터, 예를 들어, 본 명세서에 기재된 바와 같이 제조된 아넬로벡터를 투여함으로써), 예를 들어, 질환, 병리학적 병태, 또는 장애를 개선하거나, 호전시키거나, 안정화시키거나,

예방하거나 치유하기 위한 의도를 갖는 대상체의 의학적 관리를 포함한다. 일부 구현예에서, 치료는 적극 치료(질환, 병리학적 병태, 또는 장애를 개선하고자 하는 치료), 원인 치료(연관 질환, 병리학적 병태, 또는 장애의 원인을 대상으로 하는 치료), 완화 치료(증상의 경감을 위해 설계된 치료), 예방적 치료(연관 질환, 병리학적 병태, 또는 장애의 예방, 최소화, 또는 이의 발생을 부분적으로 또는 완전히 억제하고자 하는 치료), 및/또는 지지 치료(또 다른 치료법을 보충하기 위해 이용되는 치료)를 포함한다.

[0965] 본 명세서에 언급된 모든 참고문헌 및 간행물은 본 명세서에 참고로 포함된다.

[0966] 하기의 실시예는 본 발명의 일부 구현예를 추가로 예시하기 위해 제공되지만, 본 발명의 범주를 제한하고자 하지 않으며; 당업자에게 알려져 있는 다른 절차, 방법, 또는 기법이 대안적으로 사용될 수 있다는 것이 실시예의 예시적인 성질에 의해 이해될 것이다.

[0967] **실시예**

[0968] **목차**

[0969] 실시예 1: 포유동물 세포에서 전장 아넬로바이러스 ORF1 단백질 패널의 발현

[0970] 실시예 2: AAV 캡시드의 부재 하에 AAV Rep에 의한 AAV ITR-측접 DNA의 복제

[0971] 실시예 3: AAV 변이체 이식유전자 리포터 작제물을 이용한 크로스-패킹을 통한 아넬로벡터의 생성

[0972] 실시예 4: 상이한 기원의 포유동물 및 비-인간 영장류 세포에서 아넬로벡터 형질도입을 통한 리포터 작제물의 운반

[0973] 실시예 5: 아넬로-AAV 벡터의 생성 및 MOLT4 세포에서의 성공적인 형질도입

[0974] 실시예 6: 조작된 Ring2 아넬로바이러스 DNA는 AAV Rep 단백질을 통해 복제한다

[0975] 실시예 7: 인간 성장 호르몬을 인코딩하는 상이한 아넬로벡터에 의한 특정 세포주의 효과적인 형질도입

[0976] 실시예 8: 벡터 형질도입의 신속한 평가를 위한 Ring 2 아넬로벡터의 정제

[0977] **실시예 1: 포유동물 세포에서 전장 아넬로바이러스 ORF1 단백질 패널의 발현**

[0978] 본 실시예에서, 아넬로바이러스 게놈 패널 유래의 ORF1 단백질을 Expi-293 세포에서 발현시켰다. 8가지 상이한 아넬로바이러스에 대한 ORF1 서열을 확인하였다: 3가지 알파토르크바이러스(Ring1, Ring5, 및 Ring20), 3가지 베타토르크바이러스(Ring2, Ring9, 및 Ring10), 및 2가지 감마토르크바이러스(Ring3 및 Ring 4). 또한 IDT의 코돈 최적화를 사용하여 인간 세포에서의 발현을 위해 각각의 뉴클레오티드 서열을 코돈 최적화하였다. 코돈 최적화 서열을 IDT로부터 유전자 단편으로 주문하고, 서브클로닝한 다음, hEF1a 프로모터를 이용하고 N-말단 3xFlag 태그를 갖는 발현 플라스미드로 클로닝하였다.

[0979] hEF1a-구동 3xFlag-ORF1 유전자를 보유하는 각각의 플라스미드를 Expi-293 세포 내로 형질감염시켰다. 간략하게, 2.5 μg의 플라스미드 DNA를 100 μL의 무혈청 배지에서 2.5 μL의 PEI와 혼합하였다. 복합화를 위한 20분의 인큐베이션 후, PEI-DNA 혼합물을 $1x10^6$개 Expi-293 세포에 적가하였다. 그 다음 세포를 225 rpm에서 2일 동안 진탕하면서 37℃, 8% CO2에서 인큐베이션하였다.

[0980] 형질감염된 세포 용해물을 웨스턴 블롯 상에서 실행하였다. 간략하게, 100 μL의 배지에서 $5x10^5$개의 세포를 수집하고 25 μL의 4x LDS 샘플 버퍼 및 12.5 μL의 20% BME와 혼합하였다. 샘플을 실행 전 95℃에서 5분 동안 끓였다. 20 μL의 각각의 샘플을 190 V에서 1x MES SDS 실행 버퍼 중 NuPAGE 4 내지 12% 비스-트리스 겔(Invitrogen)에서 실행하였다. 그 다음 단백질을 90 V에서 1시간 동안 습식 전달을 통해 니트로셀룰로스 막으로 옮겼다. 20 mM 트리스, 0.5 M NaCl, 0.1% Brij58 pH 7.5에서 1시간 동안 블롯을 차단하였다. 마우스 항-Flag 항체의 1:2000 희석액을 블롯에 첨가하고 실온에서 밤새 인큐베이션하였다. 블롯을 세척하고 AP-토끼 항-마우스 2차 항체의 1:5000 희석액에 2시간 동안 담갔다. 그 다음 블롯을 세척하고 밴드가 나타날 때까지 블롯 현상액에 담갔다.

[0981] N-말단 3xFlag-태깅된 아넬로바이러스 ORF1 단백질에 대한 발현이 관찰되었다(도 1). 각각은 3xFlag-태깅된 ORF1에 대한 예상된 크기로 실행되었으며: 알파토르크 Ring1 ORF1은 91 kda에서, 베타토르크 Ring2 ORF1은 79 kda에서, 감마토르크 Ring4는 82 kda에서 실행되었다. 다른 아넬로바이러스 균주 유래의 다수의 ORF1 단백질에 대해서도 발현이 또한 관찰되었다(데이터는 나타내지 않음).

[0982] **실시예 2: AAV 캡시드의 부재 하에 AAV Rep에 의한 AAV ITR-측접 DNA의 복제**

[0983] 본 실시예에서, AAV 캡시드 단백질을 생성하지 않는 AAV-Rep 발현 플라스미드에 의해 플라스미드로부터 ITR-측접 리포터 유전자 작제물을 복제하였다. 고유 AAV P5 프로모터의 제어 하에 Rep78, Rep68, Rep52, 및 Rep40을 생성하는 완전한 AAV2 Rep 유전자를 갖는 발현 벡터를 구축하였다. 추가로, 유도성 TRE-tight 프로모터의 제어 하에 완전한 AAV2 Rep 유전자를 갖는 발현 벡터를 구축하였다. 복제에 대한 양성 대조군으로서, AAV2 RepCap 발현 플라스미드를 사용하였다(Cell BioLabs #VPK-422). 복제 표적으로서, CMV 프로모터에 의해 구동되고 AAV ITR이 측접한 hrGFP 리포터를 보유하는 플라스미드를 사용하였다(Cell BioLabs #AAV-400). 각각의 조건은 AAV pHelper 플라스미드(Cell BioLabs #340202), 및 Ring2 ORF1과 ORF2 단백질을 발현하는 플라스미드를 포함하였다. PEI를 사용하여 Expi293 세포로 플라스미드를 형질감염시켰다. 형질감염 4일 후, 세포 펠렛을 수집하였다.

[0984] 그 다음 각각의 샘플로부터의 총 DNA를 서던 블롯에서 실행하였다. 간략하게, 총 DNA를 세포 펠렛으로부터 단리하고, 제한 엔도뉴클레아제로 분해한 다음, 아가로스 겔 상에서 실행하고, 나일론-막으로 옮겼다. 3개의 형질감염되지 않은 DNA 대조군, 즉 pITR-hrGFP 플라스미드, 제한 효소를 통해 플라스미드로부터 ITR-hrGFP DNA를 추출함으로써 생성된 ITR-hrGFP 게놈 DNA, 및 pHelper 플라스미드 DNA를 서던 블롯에 포함하였다. 비오티닐화 DNA 단편을 사용하여 hrGFP 및 pHelper DNA 서열에 대해 블롯을 조사하였고, LiCor Odyssey 이미저에서 스트렙타비딘-연결 IRDye800으로 검출하였다(도 2). 상대 복제 효율을 결정하기 위해, 서던 상의 ITR-hrGFP 게놈 밴드와 pHelper 밴드의 밀도를 ImageJ를 사용하여 정량화하였다. 복제된 ITR-hrGFP의 양을 pHelper 플라스미드 형질감염 입력 양에 대해 정규화한 다음, pRepCap 복제 수준과 관련하여 분석하였다.

[0985] 서던 블롯 분석으로 CAP-무함유 AAV Rep 작제물이 플라스미드로부터 ITR-hrGFP 게놈을 성공적으로 복제하였음을 입증하였다(도 2). 밴드 강도를 정량화하고 형질감염 입력에 대해 정규화한 후, P5-구동 Rep 작제물은 RepCap의 ITR-hrGFP 게놈의 60%를 복제한 반면 TRE-tight-구동 Rep는 RepCap과 거의 동일하게 수행하였다. 이러한 결과는 ITR-함유 DNA 작제물이 Cap-무함유 AAV-Rep 발현 벡터를 이용하여 효율적으로 복제될 수 있음을 입증하였다. 또한, TRE-tight-프로모터 Rep 작제물은 AAV Cap 단백질을 생성하지 않으면서, 표준 pRepCap 플라스미드와 동일한 수준으로 DNA를 복제하였다.

[0986] **실시예 3: AAV 변이체 이식유전자 리포터 작제물을 이용한 크로스-패킹을 통한 아넬로벡터의 생성**

[0987] 본 실시예에서, 아넬로벡터는 AAV2 ITR 사이에 측접된 아넬로바이러스 비-코딩 서열과 마찬가지로, 전통적인 AAV 생성 구성성분(AAV rep 발현 플라스미드 및 pHelper 플라스미드) 및 리포터 나노루시퍼라제(nLuc)를 포함하는 이식유전자 플라스미드와 함께, 아넬로 ORF 단백질(ORF1, ORF2)의 공동-발현을 통해 생성되는 것으로 나타났다. 이식유전자는 상응하는 아넬로바이러스 게놈과 유사한 크기를 가졌다(± 0.3 kb). 다른 변형에서, 일부 실험에서 벡터 DNA가 아넬로바이러스 서열을 포함할 때 더 효율적으로 패키징하는 것으로 밝혀졌기 때문에, 비-코딩 아넬로바이러스 서열을 포함하였다. 이러한 아넬로벡터는 AAV2 ITR을 함유하는 리포터 작제물을 캡슐화하는 아넬로바이러스 단백질 외부로서 생성되었다. 본 실시예에서, 이식유전자의 복제 및 증폭은 AAV Rep-매개 활성을 통해 발생하는 반면, 복제된 이식유전자의 캡슐화에 필요한 구성성분은 아넬로바이러스 ORF1 및 ORF2 단백질의 트랜스-발현을 통해 발생하였다.

[0988] 간략하게, 상기 열거된 플라스미드를 PEI-Pro를 사용하여 Expi-293F 세포에 플라스미드 대 플라스미드 비율 1:1 및 DNA 대 PEI 몰 비율 1:1로 동시-형질감염시켰다. 형질감염 4일 후(dpt), 세포를 수확하고 원심분리에 의해 조건화 배지(CM)로부터 펠렛화하였다. 그 다음 세포를 화학적 또는 기계적 수단으로 용해하고, 프로테아제 저해제의 존재 하에 DNase로 처리한 다음, 지질 제거를 위해 세제로 처리하였다. 그 다음 아넬로벡터 입자를 2개의 초원심분리 단계를 통해 세포 파편과 숙주 단백질로부터 단리하였다. 첫 번째 스핀은 밀도 1.25 g/ml와 1.4 g/ml 사이의 물질이 추출되는 2-단계 CsCl 밀도 구배로 이루어졌다. 밤새 투석한 후, 이 물질을 선형 CsCl 구배에 적용하였다. 그 다음 분획을 1 ml 분취량으로 추출하고 굴절률을 취한 다음, 정량적 실시간 PCR(qPCR)을 사용하여 정량화를 위해 물질을 탈염하여 DNase 보호 이식유전자 특정 게놈을 검출하였다. 밀도 범위 1.27 내지 1.35 내의 분획을 함께 풀링한 다음 0.001% PS-80을 함유하는 버퍼에서 50 kDa MWCO를 사용하여 밤새 투석하였다. 그 다음 MWCO가 100 kDa인 원심 분리막 농축기를 사용하여 물질을 농축하였다. 그 다음 정량적 실시간 PCR(qPCR)을 사용하여 최종 물질을 정량화하여 아넬로바이러스 핵산을 검출하였다.

[0989] 도 3a는 선형 구배 분획에서 나노루시퍼라제 이식유전자에서 앰플리콘의 qPCR에 의해 수득한 벡터 게놈 카피 수를 나타낸다. 1.31 g/mL의 분획 밀도에서 벡터 카피의 명확한 피크가 관찰되었다. 이에 반해, 도 3b에 나타낸 바와 같이, 형질감염에서 ORF1 아넬로바이러스 유전자가 생략된 경우에는 이러한 피크가 전혀 관찰되지 않았다. 이들 데이터는 벡터 신호가 발현되는 ORF1에 의존적임을 나타낸다. 이와 함께, 이러한 데이터는 생성되는 아넬

로벡터와 일치한다.

[0990] **실시예 4: 상이한 기원의 포유동물 및 비-인간 영장류 세포에서 아넬로벡터 형질도입을 통한 리포터 작제물의 운반**

[0991] 본 실시예에서, 아넬로벡터는 AAV2-ITR 사이에 측접된 아넬로바이러스 비-코딩 서열과 마찬가지로, 전통적인 AAV 생성 구성성분(AAV rep 발현 플라스미드 및 pHelper 플라스미드) 및 리포터를 포함하는 이식유전자 플라스미드와 함께, 아넬로바이러스 ORF 단백질(ORF1, ORF2)의 공동-발현을 통해 생성되었다. 이 경우, 아넬로벡터는 루시퍼라제 리포터(nLuc) 또는 형광 리포터(mCherry, GFP)를 발현하는 이식유전자로 제조되었다. 본 실시예에서, 인간(Vero) 및 비-인간 영장류(Vero) 세포주의 성공적인 형질도입은 nLuc, mCherry 또는 GFP를 발현하는 ITR-측접 이식유전자를 포함하는 R2-아넬로벡터를 사용하여 입증되었다.

[0992] 선형 밀도 구배로 벡터를 정제한 다음 50 kDa MWCO 막을 사용하여 투석하여 이식유전자 단백질 캐리오버(carry-over)를 감소시켰다. Vero 및 IGR-OV1 세포에서 벡터 물질을 37℃에서 3시간 동안 인큐베이션하여 형질도입을 수행하였으며, 이 때 조건은 세포에서 바이러스의 결합 및 내재화를 허용한다. 제0일(D0) 샘플을 이 인큐베이션(nLuc 형질도입을 위함) 직후에 수확하였고 나머지 샘플은 분석 전 2일 동안 인큐베이션하였다. R2-nLuc 벡터의 경우, 발광 기반 판독을 통해 nLuc 단백질의 양을 측정하는 루시퍼라제 분석을 수행하였다. 도 4 내지 5에 나타낸 바와 같이, 아넬로벡터를 이용한 형질도입은 D0에서 D2로 1.5 log 증가한 반면, 아넬로바이러스 ORF1 및 ORF2 단백질을 발현하지 않는 물질을 이용한 형질도입은 D0에서 D2로 감소하였다. IGR-OV1 세포에서 3 log 증가가 관찰되었다(도 5). 두 세포주 모두에서, 동일한 MOI를 사용하였다(0.4). 이러한 결과는 0.2의 MOI에서 추가의 리포터(즉, GFP 및 mCherry)를 운반하는 아넬로벡터를 이용한 Vero 및 IGR-OV1 세포의 형질도입에 의해 더 강조되었다. 현미경 검사는 이러한 아넬로벡터에 의한 Vero 및 IGR-OV1 세포 둘 다의 성공적인 형질도입과 각각의 형광 리포터의 발현을 나타내었다. 아넬로바이러스 ORF1 및 ORF2 단백질을 발현하지 않는 물질로 형질도입된 대조군 세포는 어느 리포터에 의해서도 상당한 형광을 나타내지 않았다.

[0993] **실시예 5: 아넬로-AAV 벡터의 생성 및 MOLT4 세포에서의 성공적인 형질도입**

[0994] 본 실시예에서는, 아넬로바이러스 캡시드 단백질(ORF1)이 세포에서 비-동족 복제 ssDNA를 패키징할 수 있는지 여부를 테스트하였다. ORF1 단백질에 의해 패키징된 적색 형광 "mKate" 리포터 유전자를 인코딩하는 ssDNA를 생성할 수 있는 여러 AAV 구성성분(AAV Rep, 리포터 이식유전자 및 pHelper 플라스미드 구성성분을 인코딩하는 플라스미드)을 사용하였다. PEI를 사용하여 293F 세포에서 다음과 같은 형질감염을 수행하였다:

[0995] (1) AAV 캡시드 플라스미드(mKate 플라스미드, AAV Rep 및 pHelper 플라스미드)를 뺀 AAV 입자 생성의 주요 구성성분,

[0996] (2) AAV 시스템의 주요 구성성분 및 Ring2의 ORF1과 ORF2, 또는

[0997] (3) Ring2 ORF2만 있는 AAV 시스템의 주요 구성성분.

[0998] 4일 후, 세포를 용해시킨 다음, CsCl 단계 구배에 걸쳐 처리하였다(도 6). 밀도 범위 1.2 내지 1.4 g/ml 내의 분획을 수집하고 투석한 다음 MOLT4 세포(인간 T 림프모세포 세포주)를 세포당 1개 벡터의 MOI로 감염시키는 데 사용하였다. 양성 형질도입 사건은 감염 3일 후(dpi)에 유세포 측정법을 사용하여 mKate 발현 세포의 정량화를 통해 측정되었다. AAV 복제 기구 및 mKate 이식유전자만 함유하는 조건 1은 mKate를 발현하는 세포의 양성 집단을 제공하지 못한 반면, AAV 복제 기구와 함께 ORF1 및 2를 함유하는 조건 2는 mKate를 발현하는 세포의 35%를 초래하였다.

[0999] 이것이 진정한 형질도입 사건인지 여부를 추가로 확인하기 위해, 아넬로바이러스(ORF1)의 캡시드 단백질이 제외된 조건 3을 도입하였다. 이것은 조건 2의 설정에서 본 발명자들이 MOLT4 세포를 형질도입할 수 있었고 이 형질도입이 ORF1-의존적임을 시사하는 어떠한 검출가능한 형질도입 사건도 초래하지 않았다. 추가 작업으로 이러한 형질도입을 추가 세포 유형 및 Ring 4.0 아넬로-AAV 벡터로 확장하였다. 흥미롭게도, 형질도입이 수행되었을 때, Ring2 벡터에 대한 Raji 세포와 Ring4에 대한 293T 세포의 형질도입 효율이 더 높은 것으로 나타났다.

[1000] **실시예 6: 조작된 Ring2 아넬로바이러스 DNA는 AAV Rep 단백질을 통해 복제한다**

[1001] Ring2 아넬로바이러스 게놈은, 예를 들어, 본 명세서에 기재된 바와 같이, MOLT-4 세포에서 자연적으로 복제할 수 있는 것으로 나타났지만, 이제까지 HEK293 세포에서는 제대로 복제하지 않았다. 다루기 쉬운 HEK293 세포주에서 더 강력한 게놈 복제를 구동하기 위해, Ring2 버전은 AAV 복제를 위해 알려진 시스 요소를 보유하도록 조작되었다. 야생형 AAV에서, AAV Rep 단백질은 AAV ITR 내의 DNA 서열(시스 요소)에 결합하고 DNA 복제를 구동한

다. 이러한 활성에 필요한 최소 서열은 본 명세서에서 "Rep 결합 모티프"(RBM) 및 "말단 분해 부위"(TRS)로 확인되었다. 본 실시예에서, 이들 부위를 함유하는 AAV ITR 서열의 62 bp를 Ring2 게놈의 3' 비-코딩 영역(NCR)에 통합하였다(도 7a).

[1002] AAV Rep 단백질이 Ring2+RBM/TRS DNA의 복제를 구동하는지 여부를 테스트하기 위해, AAV ITR 요소(RBM 및 TRS)를 포함하는 조작된 아넬로바이러스 게놈을 보유하는 플라스미드를 트랜스-발현된 AAV Rep가 있거나 없는 Expi-293 세포에 동시-형질감염시켰다. 형질감염 4일 후 전체 DNA를 수확하고, 분해하여 플라스미드를 선형화하고 복제되지 않은 DNA를 DpnI로 분해한 다음, Ring2 게놈을 조사하는 서던 블롯에서 실행하였다(도 7b). AAV-RBM/TRS가 없는 야생형 Ring2 게놈의 경우, 선형화된 입력 플라스미드 DNA가 관찰되었지만(레인 1 및 3), DpnI(레인 2 및 4)의 존재 하에서 분해되었으며, 이는 DNA가 세포에서 복제하지 않았음을 나타낸다. 그러나, 3' NCR에 RBM/TRS가 있는 Ring2는 DpnI-내성 밴드(레인 8, 녹색 화살표)로 표시된 바와 같이, AAV Rep의 존재 하에서 성공적으로 복제하였다. Rep 없이 선형화된 플라스미드(레인 5)는 DpnI(레인 6)에 의해 분해되었으며, 이에 의해 복제가 Rep-의존적임을 확인하였다.

[1003] 이 데이터는 Expi-HEK293 세포에서 아넬로바이러스 DNA 복제를 위한 시스템의 성공적인 조작을 입증하였다. 이론에 구속되고자 하지 않지만, Ring2-3'NCR-RBM/TRS로부터 플라스미드 백본을 제거하기 위해 시험관 내 환화가 사용될 수 있고, 생성된 작제물이 AAV-Rep를 이용하여 복제되고/되거나 트랜스-발현된 Ring2 ORF1 단백질을 사용하여 패키징될 수 있음이 고려된다.

[1004] **실시예 7: 인간 성장 호르몬을 인코딩하는 상이한 아넬로벡터에 의한 특정 세포주의 효과적인 형질도입**

[1005] 상기 실시예는 시험관 내에서 세포주를 형질도입할 수 있는 형광 및 발광 페이로드를 인코딩하는 아넬로벡터를 포함하여, Expi293 세포에서 AAV 복제 기구를 이용함으로써 아넬로벡터의 생성을 입증하였다. 본 실시예에서, 생체 내 실험에 적합할 수 있는 생물학적 활성 페이로드인 인간 성장 호르몬(hGH)을 인코딩하는 아넬로벡터를 제조하였다. 간략하게, Expi293 세포를 바이러스 벡터(페이로드, AAV Rep 및 pHelper) 및 AAV2 캡시드(양성 대조군), RING2 캡시드, RING9 캡시드, 또는 캡시드 없음(음성 대조군)을 생성하는 데 필요한 플라스미드로 형질감염시켰다. 형질감염 4일 후, 세포를 수확하고 0.5% Triton X-100 함유 버퍼에서 2회 동결-해동하여 용해시켰다. 그 다음 용해물을 벤조나제로 처리한 후, 염화세슘 단계 구배를 사용하여 부분 벡터 정제를 수행하였다. 단계 구배 물질을 밤새 투석하여 염화세슘을 제거한 다음 인간 난소암 세포주 IGR-OV1 또는 원숭이 신장 세포주 Vero와 함께 3시간 동안 인큐베이션하였다. 이 처리 후, 세포를 PBS로 3회 세척하여 벡터 생성 단계로부터의 캐리오버 hGH를 포함하여, 임의의 오염 DNA 또는 단백질을 제거하였다. 새로운 배지를 형질도입된 세포에 첨가하고 37℃ 및 5% CO_2에서 인큐베이션하였다. 30분(제0일 시점), 48시간(제2일 시점), 및 72시간(제3일 시점) 후에 배양 배지를 수확하여 형질도입된 세포에 의해 분비되는 hGH의 양을 ELISA로 정량화하였다.

[1006] 도 8a 내지 8b에 나타낸 바와 같이, RING2 또는 RING9 벡터가 형질도입된 IGR-OV1 세포(도 8a) 및 Vero 세포(도 8b)의 배양 배지에서 hGH 분비량이 증가하였다. hGH를 운반하는 AAV2(양성 대조군)는 또한 비록 더 낮은 수준이지만 제2일 및 제3일에 hGH의 분비를 나타내었다. 음성 대조군으로 처리된 샘플은 분비된 hGH의 양의 유사한 증가를 나타내지 않았다. 이 데이터는 각각 생물학적 활성 페이로드를 인코딩하는 상이한 캡시드를 갖는 2개의 형질도입-적격 아넬로벡터의 성공적인 생성을 입증하였다.

[1007] **실시예 8: 벡터 형질도입의 신속한 평가를 위한 Ring 2 아넬로벡터의 정제**

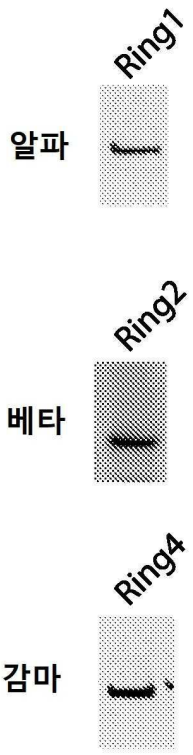
[1008] 벡터를 부분적으로 정제하지 않으면서 바이러스 형질도입을 평가하는 것은 미정제 용해물로 유발된 높은 세포 사멸로 인해 역사적으로 어려웠다. 본 실시예에는, 벡터 정제의 현재 2일 과정을 우회하여 용해물의 직접 분석을 가능하게 하고, 벡터 생성 또는 설계의 개선과 관련하여 더 빠르게 결정을 내릴 수 있게 하는 신속한 방법이 기재되어 있다. 293F 세포 유래의 용해물을 필요한 모든 구성성분의 존재(+ AAV Rep) 또는 부재(- AAV Rep) 하에서 생성된 Ring2-ITR-나노루시퍼라제(nLuc) 벡터로 형질감염시켰다. 샘플을 정화한 다음 버퍼에서 1:1로 희석하여 pH 9로 조정하고 전도도를 15 mS/cm로 낮추었다. 그 다음 조정된 용해물을 MustangQ 컬럼에 로딩하고 결합되지 않은 물질을 수집하였다. pH가 중성인 많은 염을 함유하는 버퍼를 사용하여 결합된 물질을 용출하였다. 그 다음 샘플을 qPCR 및 형질도입 검정으로 벡터 회수에 대해 평가하였다. 100 ul(대략 1/20)의 총 용출 샘플을 IGR 세포에 첨가하고 제0일과 제2일에 nLuc 활성을 측정하여 형질도입 검정을 수행하였다. D0에서 D2까지의 발광 증가로 형질도입을 측정하였다.

[1009] 도 9에 나타낸 바와 같이, 필요한 모든 플라스미드가 동시-형질감염된 샘플만이 양성 형질도입 신호를 나타내었다. 또한, 미정제 세포 용해물은 24시간 후 높은 세포 사멸을 초래하였다. 이러한 결과는 형질도입 효율을 측정

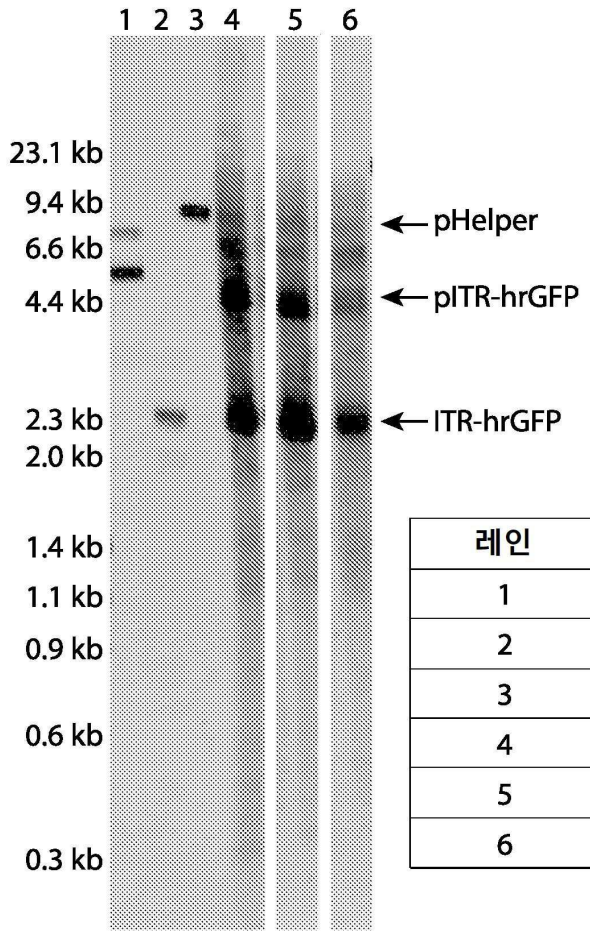
하기 위해 미정제 세포 용해물에서 아넬로벡터를 농축하고 부분적으로 정제할 수 있는 빠른 절차(실습 시간 30분)를 입증하였다. 이러한 접근법은 생성 및 설계 최적화의 처리량을 개선시키기 위한 스크리닝 방법으로서 사용될 수 있다.

도면

도면1

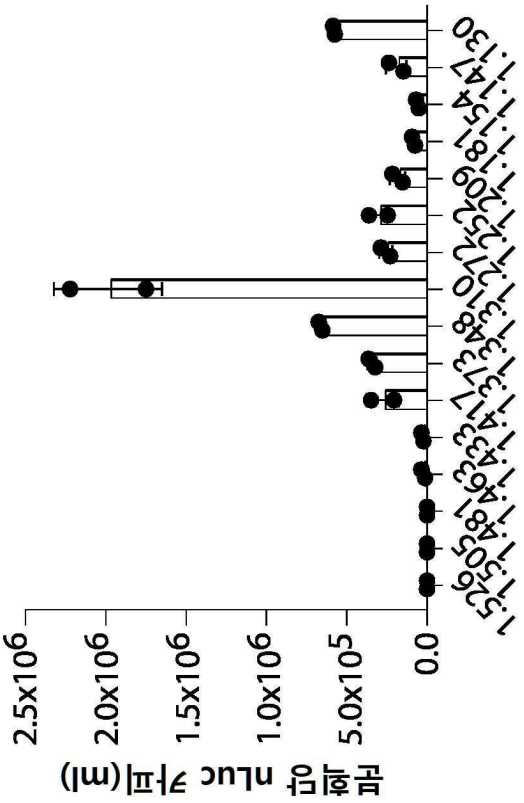


도면2

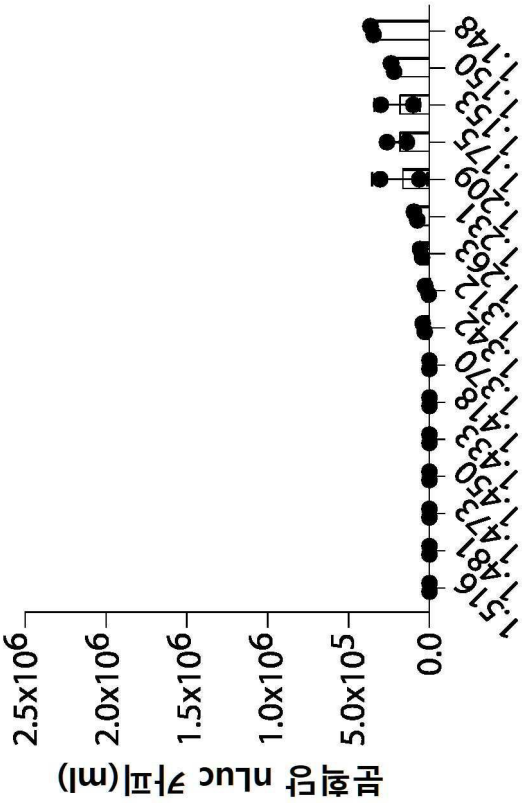


레인	Rep
1	pITR-hrGFP 플라스미드 대조군
2	ITR-hrGFP 게놈 대조군
3	pHelper 플라스미드 대조군
4	P5_Rep
5	TRE-t_Rep
6	RepCap

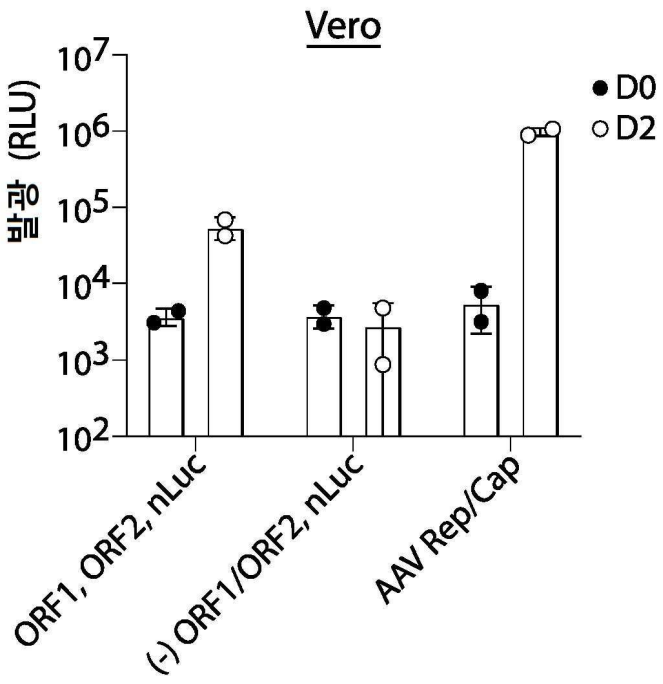
도면3a


분획당 nLuc 카피(ml)
2.5x10^6
2.0x10^6
1.5x10^6
1.0x10^6
5.0x10^5
0.0
1:526
1:505
1:481
1:463
1:433
1:417
1:373
1:348
1:310
1:272
1:252
1:209
1:181
1:154
1:147
1:130

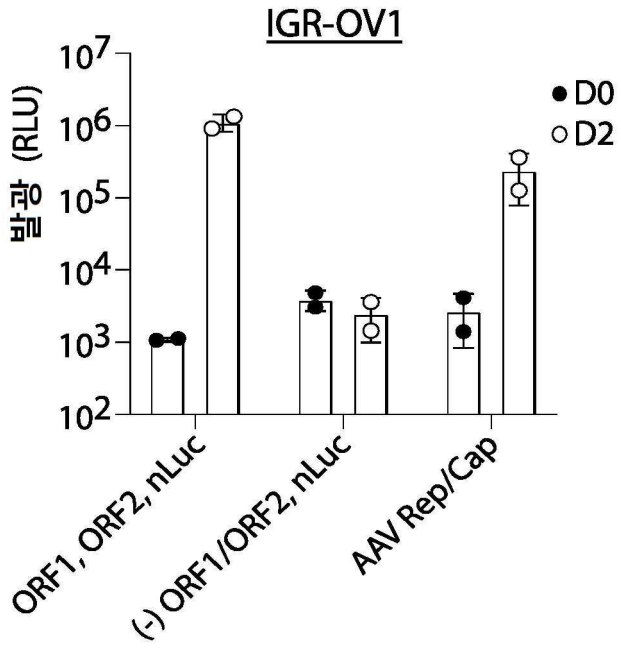
도면3b


분획당 nLuc 카피(ml)
2.5x10^6
2.0x10^6
1.5x10^6
1.0x10^6
5.0x10^5
0.0
1:516
1:481
1:473
1:450
1:433
1:418
1:397
1:372
1:346
1:323
1:309
1:275
1:253
1:233
1:209
1:175
1:153
1:150
1:148

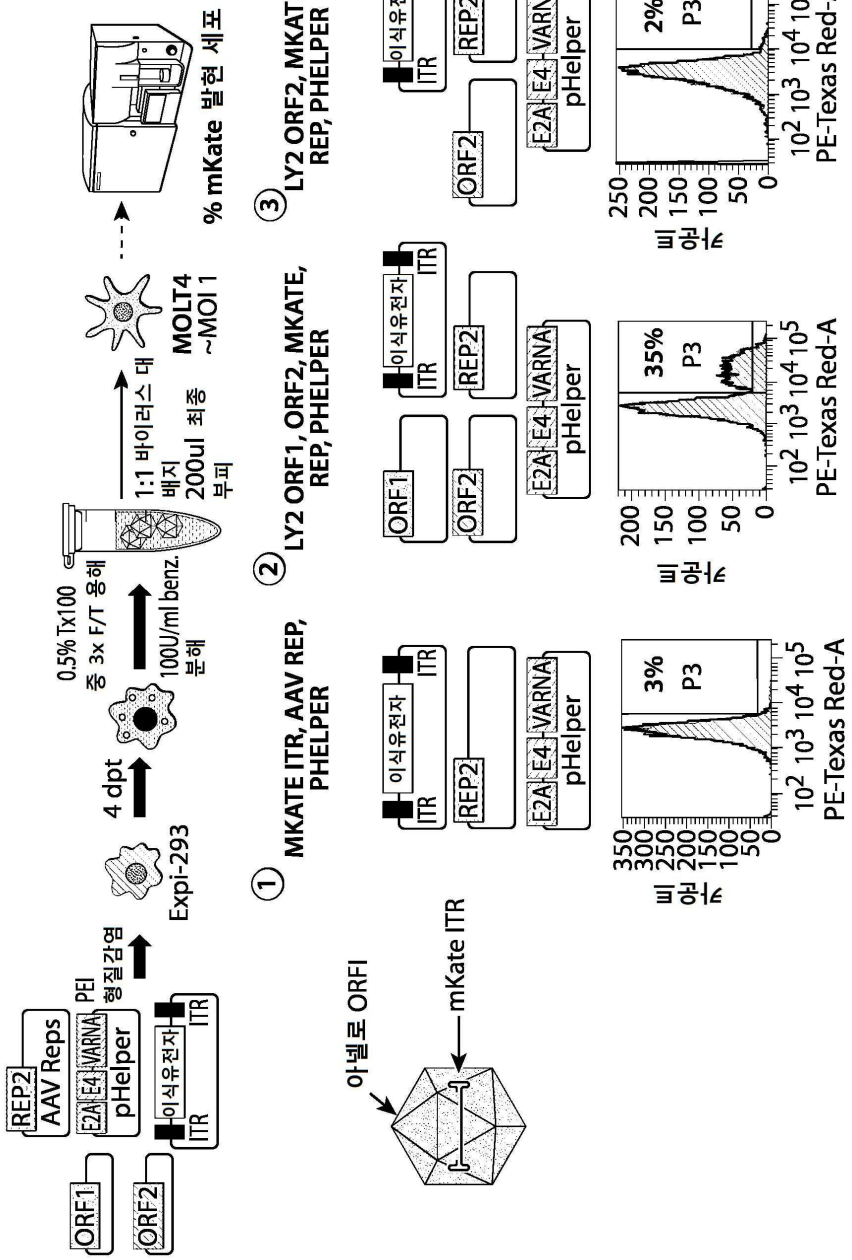
도면4


Vero
발광 (RLU)
10^7
10^6
10^5
10^4
10^3
10^2
D0
D2
ORF1, ORF2, nLuc
(-) ORF1/ORF2, nLuc
AAV Rep/Cap

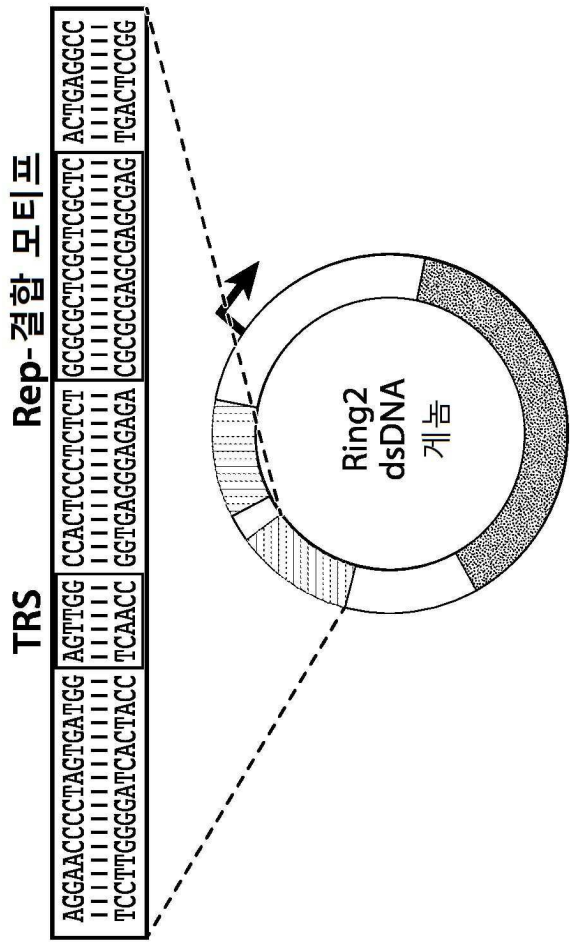
도면5


IGR-OV1
발광 (RLU)
10^7
10^6
10^5
10^4
10^3
10^2
D0
D2
ORF1, ORF2, nLuc
(-) ORF1/ORF2, nLuc
AAV Rep/Cap

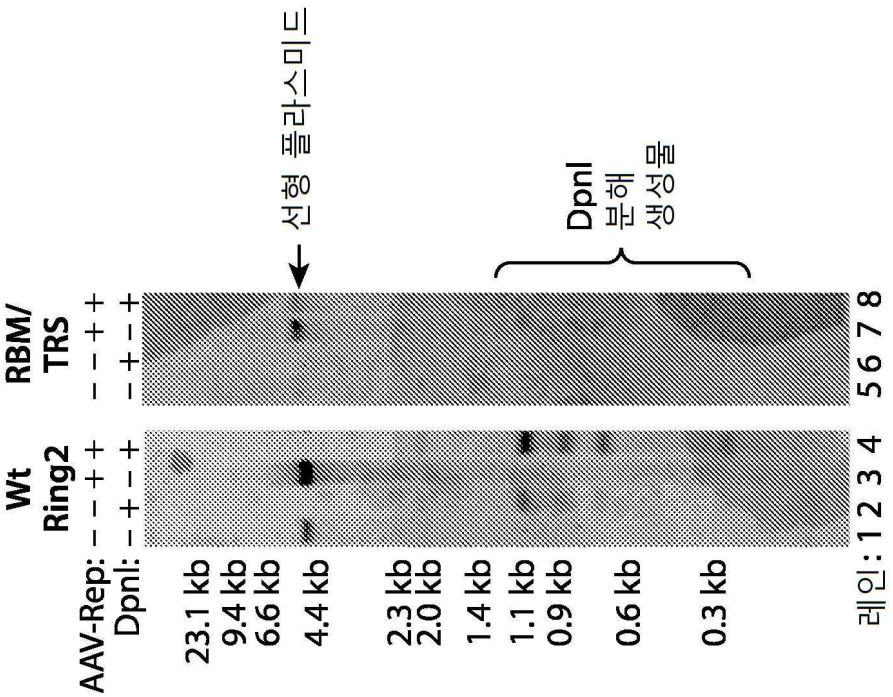
도면6


REP2
AAV Reps
E2A E4 VARNA
pHelper
ITR
이식유전자
ITR
PEI
형질감염
Expi-293
4 dpt
0.5% Tx100
중 3x F/T 용해
100U/ml benz.
분해
1:1 바이러스 대
배지
200ul 최종
부피
MOLT4
~MOI 1
% mKate 발현 세포
① MKATE ITR, AAV REP, PHELPER
② LY2 ORF1, ORF2, MKATE, REP, PHELPER
③ LY2 ORF2, MKATE, REP, PHELPER
아넬로 ORFI
mKate ITR
ORF1
ORF2
REP2
3%
P3
35%
2%
세포수
PE-Texas Red-A

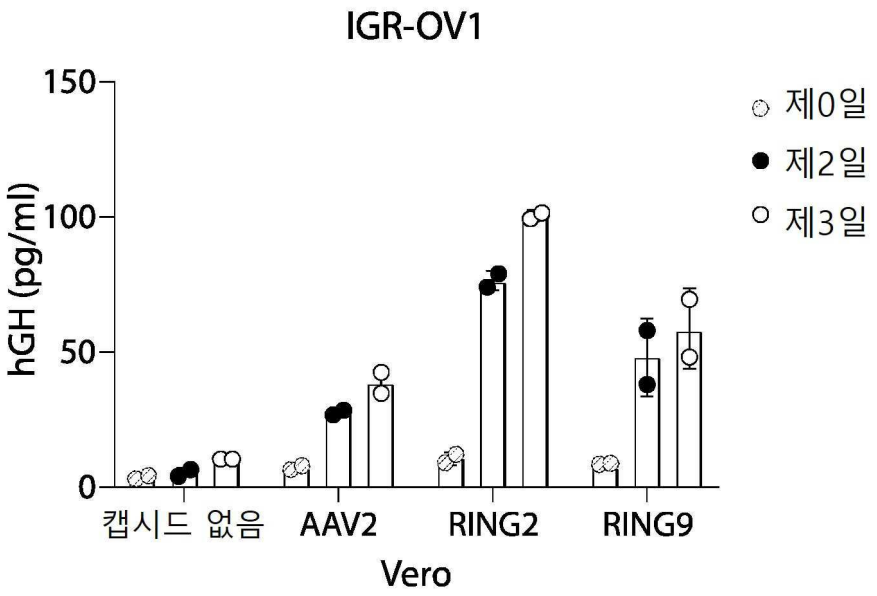
도면7a


TRS
Rep-결합 모티프
AGGAACCCCTAGTGATGG
TCCTTGGGGATCACTACC
AGTTGG
TCAACC
CCACTCCCTCTCT
GGTGAGGGAGAGA
GCGCGCTCGCTCGCTC
CGCGCGAGCGAGCGAG
ACTGAGGCC
TGACTCCGG
Ring2
dsDNA
게놈

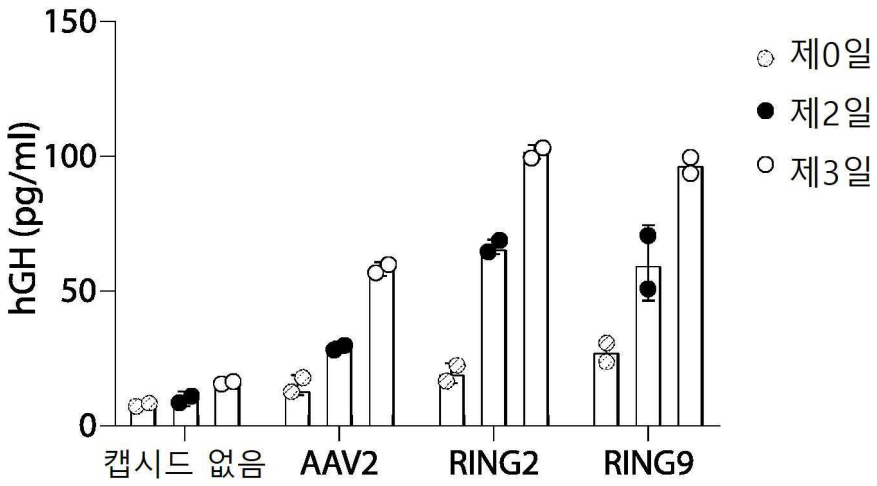
도면7b


Wt
Ring2
RBM/
TRS
AAV-Rep: – – + +
DpnI: – + – +
– – + +
– + – +
선형 플라스미드
DpnI
분해
생성물
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
1.4 kb
1.1 kb
0.9 kb
0.6 kb
0.3 kb
레인: 1 2 3 4 5 6 7 8

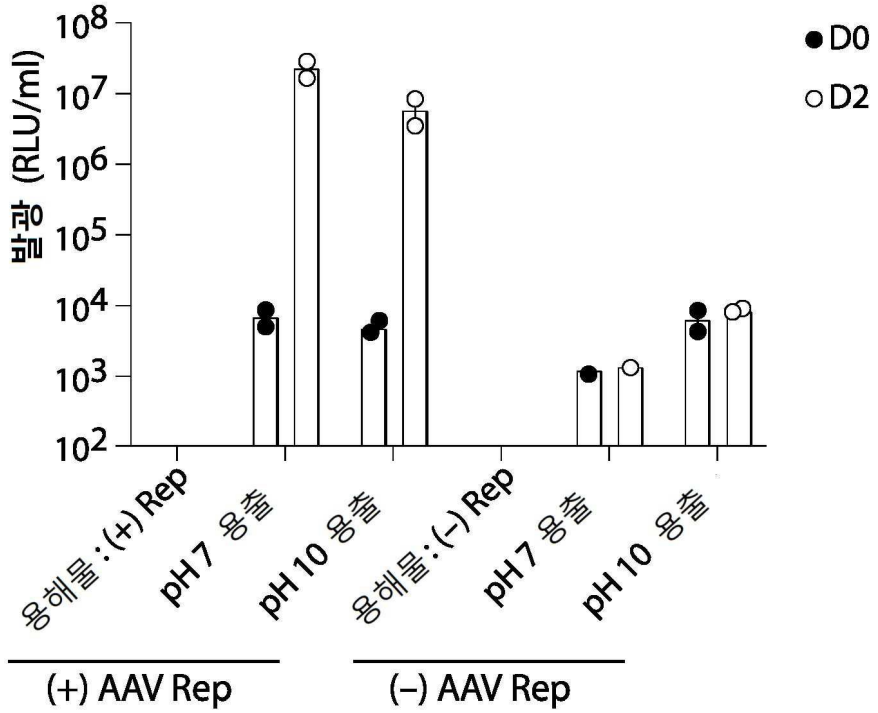
도면8a


IGR-OV1
150
100
50
0
hGH (pg/ml)
제0일
제2일
제3일
캡시드 없음
AAV2
RING2
RING9
Vero

도면8b


150
100
50
0
hGH (pg/ml)
제0일
제2일
제3일
캡시드 없음
AAV2
RING2
RING9

도면9



도면10a



도면10b



도면10c



도면10d



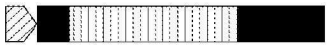
도면10e



도면10f



도면10g



도면10h



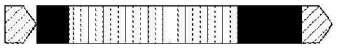
도면10i



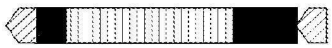
도면10j



도면10k



도면10l



서 열 목 록

SEQUENCE LISTING

<110> FLAGSHIP PIONEERING INNOVATIONS V, INC.

<120> HYBRID AAV-ANELLOVECTORS

<130> V2057-7015WO

<140><141><150> 63/147,102

<151> 2021-02-08

<160> 1067

<170> PatentIn version 3.5

<210> 1

<400> 1

000

<210> 2

<400> 2

000

<210> 3

<400> 3

000

<210> 4

<400> 4

000

<210> 5

<400> 5

000

<210> 6

<400> 6

000

<210> 7

<400> 7

000

<210> 8

<400> 8

000

<210> 9

<400> 9

000

<210> 10

<400> 10

000

<210> 11

<400> 11

000

<210> 12

<400> 12

000

<210> 13

<400> 13

000

<210> 14

<400> 14

000

<210> 15

<400> 15

000

<210> 16

<211> 3753

<212> DNA

<213> Alphatorquevirus sp.

<400> 16

tgctacgtca ctaacccacg tgtcctctac aggccaatcg cagtctatgt cgtgcacttc 60

ctgggcatgg tctacataat tatataaatg cttgcacttc cgaatggctg agtttttgct 120

gcccgtccgc ggagaggagc cacggcaggg gatccgaacg tcctgagggc gggtgccgga 180

ggtgagttta cacaccgaag tcaaggggca attcgggctc aggactggcc gggctttggg 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 gagtagaccc caacccccac atccgtagag 540

ccaggcctgc cccggccgct ccggagccct cacaggttga ttcgagacca gccctgacat 600

ggcatgggga tggtggaagc gacggaggcg ctggtggttc cggaagcggt ggacccgtgg 660

cagacttcgc agacgatggc ctcgatcagc tcgtcgccgc cctagacgac gaagagtaag 720

gaggcgcaga cggtggagga gggggagacg aaaaacaagg acttacagac gcaggagacg 780

ctttagacgc aggggacgaa aagcaaaact tataataaaa ctgtggcaac ctgcagtaat 840

taaaagatgc agaataaagg gatacatacc actgattata agtgggaacg gtacctttgc 900

cacaaacttt accagtcaca taaatgacag aataatgaaa ggccccttcg ggggaggaca 960

cagcactatg aggttcagcc tctacatttt gtttgaggag cacctcagac acatgaactt 1020

ctggaccaga agcaacgata acctagagct aaccagatac ttgggggctt cagtaaaaat 1080

atacaggcac ccagaccaag actttatagt aatatacaac agaagaaccc ctctaggagg 1140

caacatctac acagcaccct ctctacaccc aggcaatgcc attttagcaa aacacaaaat 1200

attagtacca agtttacaga caagaccaaa gggtagaaaa gcaattagac taagaatagc 1260

acccccccaca ctctttacag acaagtggta ctttcaaaag gacatagccg acctcaccct 1320

tttcaacatc atggcagttg aggctgactt gcggtttccg ttctgctcac cacaaactga 1380

caacacttgc atcagcttcc aggtccttag ttccgtttac aacaactacc tcagtattaa 1440

tacctttaat aatgacaact cagactcaaa gttaaaagaa tttttaaata aagcatttcc 1500

aacaacaggc acaaaaggaa caagtttaaa tgcactaaat acatttagaa cagaaggatg 1560

cataagtcac ccacaactaa aaaaaccaaa cccacaaata aacaaaccat tagagtcaca 1620
atactttgca cctttagatg ccctctgggg agaccccata tactataatg atctaaatga 1680
aaacaaaagt ttgaacgata tcattgagaa aatactaata aaaaacatga ttacatacca 1740
tgcaaaacta agagaatttc caaattcata ccaaggaaac aaggcctttt gccacctaac 1800
aggcatatac agcccaccat acctaaacca aggcagaata tctccagaaa tatttggact 1860
gtacacagaa ataatttaca acccttacac agacaaagga actggaaaca aagtatggat 1920
ggacccacta actaaagaga acaacatata taaagaagga cagagcaaat gcctactgac 1980

tgacatgccc ctatggactt tacttttgg atatacagac tggtgtaaaa aggacactaa 2040
taactgggac ttaccactaa actacagact agtactaata tgcccttata cctttccaaa 2100
attgtacaat gaaaaagtaa aagactatgg gtacatcccg tactcctaca aattcggagc 2160
gggtcagatg ccagacggca gcaactacat acccttcag tttagagcaa agtggtaccc 2220
cacagtacta caccagcaac aggtaatgga ggacataagc aggagcgggc cctttgcacc 2280
taaggtagaa aaaccaagca ctcagctggt aatgaagtac tgttttaact ttaactgggg 2340
cggtaaccct atcattgaac agattgttaa agaccccagc ttccagccca cctatgaaat 2400

acccggtacc ggtaacatcc ctagaagaat acaagtcatc gacccgcggg tcctgggacc 2460
gcactactcg ttccggtcat gggacatgcg cagacacaca tttagcagag caagtattaa 2520
gagagtgtca gaacaacaag aaacttctga ccttgtattc tcaggcccaa aaaagcctcg 2580
ggtcgacatc ccaaaacaag aaacccaaga agaaagctca cattcactcc aaagagaatc 2640
gagaccgtgg gagaccgagg aagaaagcga gacagaagcc ctctcgcaag agagccaaga 2700
ggtccccttc caacagcagt tgcagcagca gtaccaagag cagctcaagc tcagacaggg 2760
aatcaaagtc ctcttcgagc agctcataag gacccaacaa ggggtccatg taaacccatg 2820

cctacggtag gtcccaggca gtggctgttt ccagagagaa agccagcccc agctcctagc 2880
agtggagact gggccatgga gtttctcgca gcaaaaatat ttgataggcc agttagaagc 2940
aaccttaaag ataccccta ctacccatat gttaaaaacc aatacaatgt ctactttgac 3000
cttaaatttg aataaacagc agcttcaaac ttgcaaggcc gtgggagttt cactggtcgg 3060
tgtctacctc taaaggtcac taagcactcc gagcgtaagc gaggagtgcg accctccccc 3120
ctggaacaac ttcttcggag tccggcgcta cgccttcggc tgcgccggac acctcagacc 3180
ccccctccac ccgaaacgct tgcgcgtttc ggaccttcgg cgtcgggggg gtcgggagct 3240

ttattaaacg gactccgaag tgctcttgga cactgagggg gtgaacagca acgaaagtga 3300

gtggggccag acttcgccat aaggccttta tcttcttgcc atttgtcagt gtccggggtc 3360

gccataggct tcgggctcgt ttttaggcct tccggactac aaaaatcgcc attttggtga 3420

cgtcacggcc gccatcttaa gtagttgagg cggacggtgg cgtgagttca aaggtcacca 3480

tcagccacac ctactcaaaa tggtggacaa tttcttccgg gtcaaaggtt acagccgcca 3540

tgttaaaaca cgtgacgtat gacgtcacgg ccgccatttt gtgacacaag atggccgact 3600

tccttcctct ttttcaaaaa aaagcggaag tgccgccgcg gcggcggggg gcggcgcgct 3660

gcgcgcgccg cccagtaggg ggagccatgc gccccccccc 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
165 170 175
Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
180 185 190
Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr

195 200 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
210 215 220
Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225 230 235 240
Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
245 250 255
Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser

260 265 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
275 280 285
Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
290 295 300
Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305 310 315 320
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> 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 ccgtcactta 60

gttccccttt ttgcaacaac ttctgctttt ttccaactgc cggaaaacca cataatttgc 120

atggctaacc acaaactgat atgctaatta acttccacaa aacaacttcc ccttttaaaa 180

ccacacctac aaattaatta ttaaacacag tcacatcctg ggaggtacta ccacactata 240

ataccaagtg cacttccgaa tggctgagtt tatgccgcta gacggagaac gcatcagtta 300

ctgactgcgg actgaacttg ggcgggtgcc gaaggtgagt gaaaccaccg aagtcaaggg 360

gcaattcggg ctagttcagt ctagcggaac gggcaagaaa cttaaaatta ttttatttt 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 agacgcccct ttcgcagaag atttcgaaga aaaagaaggg 720

taagacctac ttatactact attcctctaa agcaatggca accgccatat aaaagaacat 780

gctatataaa aggacaagac tgtttaatat actatagcaa cttaagactg ggaatgaata 840

gtacaatgta tgaaaaaagt attgtacctg tacattggcc gggaggggggt tctttttctg 900

taagcatgtt aactttagat gccttgtatg atatacataa actttgtaga aactggtgga 960

catccacaaa ccaagactta ccactagtaa gatataaagg atgcaaaata acattttatc 1020

aaagcacatt tacagactac atagtaagaa tacatacaga actaccagct aacagtaaca 1080

aactaacata cccaaacaca catccactaa tgatgatgat gtctaagtac aaacacatta 1140

tacctagtag acaaacaaga agaaaaaaga aaccatacac aaaaatattt gtaaaaccac 1200

ctccgcaatt tgaaaacaaa tggtactttg ctacagacct ctacaaaatt ccattactac 1260

aaatacactg cacagcatgc aacttacaaa acccatttgt aaaaccagac aaattatcaa 1320

acaatgttac attatggtca ctaaacacca taagcataca aaatagaaac atgtcagtgg 1380

atcaaggaca atcatggcca tttaaaatac taggaacaca aagcttttat ttttactttt 1440

acaccggagc aaacctacca ggtgacacaa cacaaatacc agtagcagac ctattaccac 1500

taacaaaccc aagaataaac agaccaggac aatcactaaa tgaggcaaaa attacagacc 1560

atattacttt cacagaatac aaaaacaaat ttacaaatta ttggggtaac ccatttaata 1620

aacacattca agaacaccta gatatgatac tatactcact aaaaagtcca gaagcaataa 1680

aaaacgaatg gacaacagaa aacatgaaat ggaaccaatt aaacaatgca ggaacaatgg 1740

cattaacacc atttaacgag ccaatattca cacaaataca atataaccca gatagagaca 1800

caggagaaga cactcaatta tacctactct ctaacgctac aggaacagga tgggacccac 1860

caggaattcc agaattaata ctagaaggat ttccactatg gttaatatat tggggatttg 1920

cagactttca aaaaaaccta aaaaaagtaa caaacataga cacaaattac atgttagtag 1980

caaaaacaaa atttacacaa aaacctggca cattctactt agtaatacta aatgacacct 2040
ttgtagaagg caatagccca tatgaaaaac aacctttacc tgaagacaac attaaatggt 2100
acccacaagt acaataccaa ttagaagcac aaaacaaact actacaaact gggccattta 2160
caccaaacat acaaggacaa ctatcagaca atatatcaat gttttataaa ttttacttta 2220
aatggggagg aagcccacca aaagcaatta atgttgaaaa tcctgcccac cagattcaat 2280
atcccatacc ccgtaacgag catgaaacaa 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 aggtgaaatt 2640
gtttgatgat aaaccaaaaa accgtagatt tacacctgag gaatttgaaa ctgagttaca 2700
aatagcaaaa tggttaaaga gacccccaag atcctttgta aatgatcctc ccttttaccc 2760
atggttacca cctgaacctg ttgtaaactt taagcttaat tttactgaat aaaggccagc 2820

attaattcac ttaaggagtc tgtttattta agttaaacct taataaacgg tcaccgcctc 2880
cctaatacgc aggcgcagaa agggggctcc gccccttta acccccaggg ggctccgccc 2940
cctgaaaccc ccaaggggggc tacgcccccct 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 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
50 55 60
Tyr Ser Ser Ser Ser Ala Thr Ser Asp Ser Ser Ser Ser Cys Thr Glu
65 70 75 80

Ser Glu

<210> 61

<400> 61

000

<210> 62

<400> 62

000

<210> 63

<400> 63

000

<210> 64

<400> 64

000

<210> 65

<400> 65

000

<210> 66

<400> 66

000

<210> 67

<400> 67

000

<210> 68

<400> 68

000

<210> 69

<400> 69

000

<210> 70

<400> 70

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 dcgcgcgcsg cccrsyrggg grdssmmwgc skcsccccccc 60

cscgcgcatg cgcrcgggkc ccccccccyv sgggggggctc cgcccccccg 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

gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcatg cgcgggkccc ccccccnncg gggggctccg cccccggcc 120

cccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 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

gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcat 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

ccccccggcc ccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 59

<210> 125

<211> 156

<212> DNA

<213> Artificial Sequence

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

<400> 125

gcggcggggg ggcggccgcg ttcgcgcgcc gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc 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 gcccaccagg gggtg 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 ccccccc 17

<210> 133

<211> 72

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 133

gggggggctc cgccccccg gccccccccc gtgctaaacc caccgcgcat gcgcgaccac 60

gcccccgccg 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 cgggtccccc cccccacggg gggctccgcc 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 gcgcgcg 17

<210> 137

<211> 19

<212> DNA

<213> Artificial Sequence

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

<400> 137

cgccgggggg gcgccagcg 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 ccccccccac gggggctcc g 31

<210> 140

<211> 17

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 140

ccccccggcc cccccc 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 agcctccccc 60

ccccgcgcgc atgcgcgcgg gtcccccccc ctccgggggg ctccgccccc 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 cgcatgcgcg cgggtccccc cccctccggg 60

gggctccgcc ccccggcccc cccc 84

<210> 144

<211> 104

<212> DNA

<213> Artificial Sequence

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

<400> 144

cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc ccccccgcg gggggctccg ccccccggcc 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 cgcat 15

<210> 150

<211> 17

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 150

gcgcggggcc ccccccc 17

<210> 151

<211> 13

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 151

gcgggggct ccg 13

<210> 152

<211> 14

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 152

ccccccggcc cccc 14

<210> 153

<211> 122

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 153

gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

ccccccccg cgcatgcgcg gggccccccc ccgcgggggg ctccgccccc cggccccccc 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 cgcat 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

gcgggggct ccg 13

<210> 159

<211> 17

<212> DNA

<213> Artificial Sequence

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

<400> 159

ccccccggcc ccccccg 17

<210> 160

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 160

cgcgctgcgc gcgccgccca gtaggggggag 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 attttgtgac acaagatggc cgacttcctt cc 172

<210> 163

<211> 36

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 163

cgcgctgcgc gcgccgccca gtaggggggag ccatgc 36

<210> 164

<211> 36

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 164

gcgctdcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 165

<211> 36

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 165

gcgcttcgcg cgccgcccac taggggcgt tgcgcg 36

<210> 166

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 166

gcgctgcgcg cgccgcccag taggggcgc aatgcg 36

<210> 167

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 167

gcgctgcgcg cgcggccccc gggggaggca ttgcct 36

<210> 168

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 168

gcgctgcgcg cgcgcgccgg gggggcgcca gcgccc 36

<210> 169

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 169

gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 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
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
165 170 175
Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
180 185 190
Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr
195 200 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
210 215 220
Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225 230 235 240
Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
245 250 255

Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
260 265 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
275 280 285
Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
290 295 300
Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305 310 315 320
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 Glu Xaa Xaa Xaa Xaa Xaa Xaa
130 135 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 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 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
225 230 235 240

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 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 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 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 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 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 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 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

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 300

gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 301

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 301

gcguacguca caagucacgu ggaggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 302

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 302

gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 303

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 303

gcacacguca uaagucacgu gguggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 304

<211> 66

<212> RNA

<213> Alphatorquevirus sp.

<400> 304

cuuccggguc auaggucaca ccuacgucac aagucacgug gggaggguug gcguauagcc 60

cggaag 66

<210> 305

<211> 68

<212> RNA

<213> Alphatorquevirus sp.

<400> 305

gccgggggggc ugccgccccc cccggggaaa gggggggggcc ccccccgggg gggggguuugc 60

ccccggc 68

<210> 306

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 306

auacgucauc agucacgugg gggaaggcgu gccuaaaccc ggaagcaucc ucguccacgu 60

gacugugacg uguguggc 78

<210> 307

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 307

cauuuuaagu aaggcggaag cagcucggcg uacacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ugg 73

<210> 308

<211> 71

<212> RNA

<213> Alphatorquevirus sp.

<400> 308

gucacaaguc acguggggag gguuggcguu uaacccggaa gccaauccuc uuacguggcc 60

ugucacguga c 71

<210> 309

<211> 70

<212> RNA

<213> Alphatorquevirus sp.

<400> 309

cgaccgcguc ccgaaggcgg guacccgagg ugaguuuaca caccgagguu aagggccaau 60

ucgggcuugg 70

<210> 310

<211> 59

<212> RNA

<213> Alphatorquevirus sp.

<400> 310

cgcgguaucg uagccgacgc ggaccccguu uucggggccc ccgcggggcu cucggcgcg 59

<210> 311

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 311

cgccauuuug ugauacgcgc guccccuccc ggcuuccgua caacgucagg cggggcgugg 60

ccguaucaga aaauggcg 78

<210> 312

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 312

gcuacgucau aagucacgug acugggcagg uacuaaaccc ggaaguaucc ucggucacgu 60

ggccugucac guaguug 77

<210> 313

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 313

ggcusugacg ucaaagucac gugggraggg uggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacagcc 80

<210> 314

<211> 66

<212> RNA

<213> Alphatorquevirus sp.

<400> 314

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

gaaggu 66

<210> 315

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 315

gguugugacg ucaaagucac guggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 316

<211> 67

<212> RNA

<213> Alphatorquevirus sp.

<400> 316

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 317

<211> 67

<212> RNA

<213> Alphatorquevirus sp.

<400> 317

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 318

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 318

ggcugugacg ucaaagucac guggggaggg cggcguuaaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 319

<211> 79

<212> RNA

<213> Alphatorquevirus sp.

<400> 319

agaccacgug guaagucacg ugggggcagc ugcuguaaac ccggaaguag cugacccgcg 60

ugacugguca cgugaccug 79

<210> 320

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 320

cgccauuuua uaauacgcgc guccccuccc ggcuuccgua cuacgucagg cggggcgugg 60

ccguauuaga aaauggug 78

<210> 321

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 321

uaaguaaggc ggaaccaggc ugucacccug ugucaaaggu caagggacag ccuuccggcu 60

ugcacaaaau gg 72

<210> 322

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 322

ugccuacguc auaagucacg uggggacggc ugcuguaaac acggaaguag cugacccgcg 60

ugacuuguca cgugagca 78

<210> 323

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 323

uuguguaagg cggaacaggc ugacaccccg ugucaaaggu caggggucag ccuccgcuuu 60

gcaccaaaug gu 72

<210> 324

<211> 79

<212> RNA

<213> Alphatorquevirus sp.

<400> 324

uaccuacguc auaagucacg ugggaagagc ugcugugaac cuggaaguag cugacccgcg 60

uggcuuguca cgugagugc 79

<210> 325

<211> 75

<212> RNA

<213> Alphatorquevirus sp.

<400> 325

uuuuccuggc ccguccgcgg cgagagcgcg agcgaagcga gcgaucgggc gucccgaggg 60

cgggugccgg aggug 75

<210> 326

<211> 68

<212> RNA

<213> Alphatorquevirus sp.

<400> 326

aaagugagug gggccagacu ucgccauagg gccuuuaacu uccgggugcg ucuggggccc 60

gccauuuu 68

<210> 327

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 327

gugacguuac ucucacguga ugggggcgug cucuaacccg gaagcauccu cgaccacgug 60

acugugacgu cac 73

<210> 328

<211> 75

<212> RNA

<213> Alphatorquevirus sp.

<400> 328

agcgucuacu acguacacuu ccuggggugu guccugccac uguauauaaa ccagaggggu 60

gacgaauggu agagu 75

<210> 329

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 329

gugacgucaa agucacgugg ugacggccau uuuaacccgg aaguggcugu ugucacguga 60

cuugacguca cgg 73

<210> 330

<211> 62

<212> RNA

<213> Alphatorquevirus sp.

<400> 330

gcuuuagacg ccauuuuagg cccucgcggg cacccguagg cgcguuuuaa ugacgucacg 60

gc 62

<210> 331

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 331

cacccguagg cgcguuuuaa ugacgucacg gcagccauuu ugucgugacg uuugagacac 60

gugaugggggg cgu 73

<210> 332

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 332

gucgugacgu uugagacacg ugaugggggc gugccuaaac ccggaagcau cccuggucac 60

gugacucuga cgucacggcg 80

<210> 333

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 333

cgaaagugag uggggccaga cuucgccaua aggccuuuaa cuuccgggug cgugugggggg 60

ccgccauuuu agcuucg 77

<210> 334

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 334

cugugacguc aaagucacgu ggggagggcg gcguguaacc cggaagucau ccucgucacg 60

ugaccugacg ucacgg 76

<210> 335

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 335

cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgua 60

gcggcggggg ggc 73

<210> 336

<211> 67

<212> RNA

<213> Alphatorquevirus sp.

<400> 336

gguagaguuu uuuccgcccg uccgcagcga ggacgcgagc gcagcgagcg gccgagcgac 60

ccguggg 67

<210> 337

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 337

gcugugacgu uucagucacg uggggaggga acgccuaaac ccggaagcgu cccuggucac 60

gugauuguga cgucacggcc 80

<210> 338

<211> 63

<212> RNA

<213> Alphatorquevirus sp.

<400> 338

ccgccauuuu gugacuuccu uccgcuuuuu caaaaaaaaa gaggaagugu gacguagcgg 60

cgg 63

<210> 339

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 339

gacugugacg ucaaagucac guggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 340

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 340

cuguccgcca ucuugugacu uccuuccgcu uuuucaaaaa aaaagaggaa guaugacgug 60

gcggcggggg ggc 73

<210> 341

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 341

gguugugacg ucaaagucac guggggaggg cggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacggcc 80

<210> 342

<211> 65

<212> RNA

<213> Alphatorquevirus sp.

<400> 342

cccgccaucu ugugacuucc uuccgcuuuu ucaaaaaaaa agaggaagug ugacguagcg 60

gcggg 65

<210> 343

<211> 67

<212> RNA

<213> Alphatorquevirus sp.

<400> 343

gcccguccgc ggcgagagcg cgagcgaagc gagcgaucga gcgucccgug ggcgggugcc 60

guaggug 67

<210> 344

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 344

gacugugacg ucaaagucac guggggagga gggcguguaa cccggaaguc auccucguca 60

cgugaccuga cgucacgg 78

<210> 345

<211> 62

<212> RNA

<213> Alphatorquevirus sp.

<400> 345

ucgcgucuua gugacgucac ggcagccauc uugguccuga cgucacuguc acguggggag 60

gg 62

<210> 346

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 346

ugacgucacu gucacguggg gagggaacac gugaacccgg aaguguccu ggucacguga 60

caugacguca cggccg 76

<210> 347

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 347

cgccauuuua aguaagcaug gcgggcggug augucaaaug uuaaagguca cagccgguca 60

ugcuugcaca aaauggcg 78

<210> 348

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 348

cgccauuuua aguaagcaug gcgggcggug acgugcaaug ucaaagguca cagccuguca 60

ugcuugcaca aaauggcg 78

<210> 349

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 349

ccaucuuaag uaguugaggc ggacgguggc gucgguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 350

<211> 67

<212> RNA

<213> Alphatorquevirus sp.

<400> 350

gccugucaug cuugcacaaa auggcggacu uccgcuuccg ggucgccgcc auauuugguc 60

acgugac 67

<210> 351

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 351

gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 352

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 352

gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 353

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 353

gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 354

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 354

gccauuuuaa guagcugacg ucaaggauug acgugaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 355

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 355

gcacacguca uaagucacgu gguggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugauuuguca cgugugua 78

<210> 356

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 356

gccauuuuaa gucagcucug gggaggcgug acuuccaguu caaaggucau ccucaccaua 60

acuggcacaa aauggc 76

<210> 357

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 357

gccauuuuaa guagcugacg ucaaggauug acguaaaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggu 76

<210> 358

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 358

gcauacguca caagucacgu ggaggggaca cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 359

<211> 79

<212> RNA

<213> Alphatorquevirus sp.

<400> 359

gcgccauguu aaguggcugu cgccgaggau ugacgucaca guucaaaggu cauccucgac 60

gguaaccgca aacauggcg 79

<210> 360

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 360

caugcgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 361

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 361

uggcagcacu uccgaauggc ugaguuuucc acgcccgucc gcggagaggg agccacggag 60

gugaucccga acg 73

<210> 362

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 362

gccauuuuaa gucagcgcug gggaggcaug acuguaaguu caaaggucau ccucaccgga 60

acugacacaa aauggccg 78

<210> 363

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 363

gccaucuuaa guggcugucg ccgaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 364

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 364

auacgucaua agucacaugu cuaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 365

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 365

ccauuuuaag uaaggcggaa gcagcugucc cuguaacaaa auggcggcga cagccuuccg 60

cuuugcacaa aauggag 77

<210> 366

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 366

gccaucuuaa guggcugucg cugaggauug acgucacagu ucaaagguca uccucggcgg 60

uaaccgcaaa gauggcgguc 80

<210> 367

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 367

cauacgucau aagucacaug acaggagucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 368

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 368

cgccaucuua aguggcuguc gccgaggauu ggcgucacag uucaaagguc auccucggcg 60

guaaccgcaa agauggcggu 80

<210> 369

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 369

cauacgucau aagucacaug acaggggucc acuuaaacac ggaaguaggc cccgacaugu 60

gacucgucac gugugu 76

<210> 370

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 370

gcauacguca caagucacgu gggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cguggugu 78

<210> 371

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 371

ccgccauuuu aggcuguugc cgggcguuug acuuccgugu uaaaggucaa acacccagcg 60

acaccaaaaa auggccg 77

<210> 372

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 372

cuacgucaua agucacguga cagggagggg cgacaaaccc ggaagucauc cucgcccacg 60

ugacuuacca cguggug 77

<210> 373

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 373

gccauuuuaa guaggugacg uccaggacug acguaaaguu caaaggucau ccucggcgga 60

accuauacaa aauggcg 77

<210> 374

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 374

cuacgucaua agucacgugg ggacggcugu acuuaaacac ggaaguaggc cccgucacgu 60

gauuuaccac guggug 76

<210> 375

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 375

gccauuuuaa guaaggcgga agagcucuag cuauacaaaa uggcggcgga gcacuuccgc 60

uuugcccaaa aug 73

<210> 376

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 376

gccauuuuaa guagcugacg ucaaggauug acguagaggu uaaaggucau ccucggcgga 60

agcuacacaa aauggug 77

<210> 377

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 377

gcauacguca caagucacgu ggggggggacc cgcuguaacc cggaaguagg ccccgucacg 60

ugacuuacca cgugugua 78

<210> 378

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 378

ggcgccauuu uaaguaagca uggcgggcgg cgacgucaca ugucaaaggu caccgcacuu 60

ccgugcuugc acaaaauggc 80

<210> 379

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 379

ugcuacguca ucgagacacg uggugccagc agcuguaaac ccggaagucg cugacacacg 60

ugucuuguca cgu 73

<210> 380

<211> 78

<212> RNA

<213> Alphatorquevirus sp.

<400> 380

gccauuuuaa guaagcaccg ccuagggaug acguauaagu ucaaagguca uccucagccg 60

gaacuuacac aaaauggu 78

<210> 381

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 381

acgucauaug ucacgugggg aggcccugcu gcgcaaacgc ggaaguaggc cccgucacgu 60

gucauaccac gu 72

<210> 382

<211> 77

<212> RNA

<213> Alphatorquevirus sp.

<400> 382

ccauuuuaag uaaggcggaa gcagcuccac uuucucacaa aauggcggcg gggcacuucc 60

ggcuugccca aaauggc 77

<210> 383

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 383

ccauuuuaag uaaggcggaa guuucuccac uauacaaaau ggcggcggag cacuuccggc 60

uugcccaaaa ug 72

<210> 384

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 384

ccaucuuaag uaguugaggc ggacgguggc gugaguucaa aggucaccau cagccacacc 60

uacucaaaau gg 72

<210> 385

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 385

cgccaucuua aguaguugag gcggacggug gcgugaguuc aaaggucacc aucagccaca 60

ccuacucaaa auggug 76

<210> 386

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 386

uuucggaccu ucggcgucgg gggggucggg ggcuuuacua aacagacucc gagaugccau 60

uggacacuga ggg 73

<210> 387

<211> 76

<212> RNA

<213> Alphatorquevirus sp.

<400> 387

ccauuuuaag uaggugccgu ccagcacugc uguuccgggu uaaagggcau ccucggcgga 60

accuauacaa aauggc 76

<210> 388

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 388

cuacgucauc gaugacgugg ggaggcguac uaugaaacgc ggaaguaggc cccgcuacgu 60

caucaucacg ugg 73

<210> 389

<211> 73

<212> RNA

<213> Alphatorquevirus sp.

<400> 389

ccauuuuaag uaaggcggaa gagcugcucu auauacaaaa uggcggagga gcacuuccgg 60

cuugcccaaa aug 73

<210> 390

<211> 75

<212> RNA

<213> Alphatorquevirus sp.

<400> 390

ugccuacgua acaagucacg uggggagggu uggcguauaa cccggaaguc aauccuccca 60

cguggccugu cacgu 75

<210> 391

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 391

uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 392

<211> 72

<212> RNA

<213> Alphatorquevirus sp.

<400> 392

uaaguaaggc ggaaccaggc ugucaccccg ugucaaaggu caggggucag ccuuccgcuu 60

uacacaaaau gg 72

<210> 393

<211> 80

<212> RNA

<213> Alphatorquevirus sp.

<400> 393

gcagccauuu uaagucagcu ucggggaggg ucacgcaaag uucaaagguc auccucaccg 60

gaacugguac aaaauggccg 80

<210> 394

<211> 74

<212> RNA

<213> Alphatorquevirus sp.

<400> 394

ugcuacguca uaagugacgu agcugguguc ugcuguaaac acggaaguag gccccgccac 60

gucacuuguc acgu 74

<210> 395

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 395

aguagcugac gucaaggauu gac 23

<210> 396

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 396

caagucacgu ggaggggacc cg 22

<210> 397

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 397

aaguagcuga cgucaaggau ugacg 25

<210> 398

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 398

auaagucacg ugguggggac ccg 23

<210> 399

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 399

uggggagggu uggcguauag cccgga 26

<210> 400

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 400

ccccccccgg gggggguuu gccc 24

<210> 401

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 401

aucagucacg ugggggaagg cgugc 25

<210> 402

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 402

aaguaaggcg gaagcagcuc gg 22

<210> 403

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 403

agucacgugg ggaggguugg c 21

<210> 404

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 404

cccgaaggcg gguacccgag gu 22

<210> 405

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 405

uaucguagcc gacgcggacc ccg 23

<210> 406

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 406

auuuugugau acgcgcgucc ccuccc 26

<210> 407

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 407

aagucacgug acugggcagg u 21

<210> 408

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 408

ugacgucaaa gucacguggg ragggu 26

<210> 409

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 409

gaucgagcgu cccgugggcg ggu 23

<210> 410

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 410

ugacgucaaa gucacguggg gagggcgg 28

<210> 411

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 411

gaucgagcgu cccgugggcg ggu 23

<210> 412

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 412

gaucgagcgu cccgugggcg ggu 23

<210> 413

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 413

ugacgucaaa gucacguggg gagggcgg 28

<210> 414

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 414

acgugguaag ucacgugggg gcagcu 26

<210> 415

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 415

auuuuauaau acgcgcgucc ccucc 25

<210> 416

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 416

aagggacagc cuuccggcuu gc 22

<210> 417

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 417

cauaagucac gugggacgg cugcu 25

<210> 418

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 418

uaaggcggaa caggcugaca cccc 24

<210> 419

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 419

uacgucauaa gucacguggg aagagcug 28

<210> 420

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 420

ucgggcgucc cgagggcggg ug 22

<210> 421

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 421

aaagugagug gggccagacu ucgcc 25

<210> 422

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 422

cucucacgug auggggcgu gc 22

<210> 423

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 423

ucuacuacgu acacuuccug gggugugu 28

<210> 424

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 424

uggcuguugu cacgugacuu ga 22

<210> 425

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 425

agacgccauu uuaggcccuc gcgg 24

<210> 426

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 426

ugucgugacg uuugagacac gugau 25

<210> 427

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 427

ugacguuuga gacacgugau gggggcgugc 30

<210> 428

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 428

agugaguggg gccagacuuc gc 22

<210> 429

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 429

ugugacguca aagucacgug gggagggcgg 30

<210> 430

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 430

aaaagaggaa guaugacgua gcggcgg 27

<210> 431

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 431

agcgagcggc cgagcgaccc g 21

<210> 432

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 432

uucagucacg uggggaggga acgc 24

<210> 433

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 433

aaaagaggaa gugugacgua gcgg 24

<210> 434

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 434

ugugacguca aagucacgug gggagggcgg 30

<210> 435

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 435

aaaagaggaa guaugacgug gcgg 24

<210> 436

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 436

ugacgucaaa gucacguggg gagggcgg 28

<210> 437

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 437

aaaaaagagg aagugugacg uagcggcgg 29

<210> 438

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 438

gaucgagcgu cccgugggcg ggu 23

<210> 439

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 439

ugugacguca aagucacgug gggaggaggg 30

<210> 440

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 440

uugguccuga cgucacuguc a 21

<210> 441

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 441

cgucacuguc acguggggag ggaacac 27

<210> 442

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 442

uaaguaagca uggcgggcgg ugau 24

<210> 443

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 443

aaguaagcau ggcgggcggu ga 22

<210> 444

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 444

uaaguaguug aggcggacgg uggc 24

<210> 445

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 445

ucaugcuugc acaaaauggc ggacuuccg 29

<210> 446

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 446

aguagcugac gucaaggauu gac 23

<210> 447

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 447

aguagcugac gucaaggauu gac 23

<210> 448

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 448

acaagucacg ugggggggac ccg 23

<210> 449

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 449

aaguagcuga cgucaaggau ugacg 25

<210> 450

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 450

auaagucacg ugguggggac ccg 23

<210> 451

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 451

aagucagcuc uggggaggcg ugacuu 26

<210> 452

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 452

aguagcugac gucaaggauu gac 23

<210> 453

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 453

caagucacgu ggaggggaca cg 22

<210> 454

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 454

uguuaagugg cugucgccga ggauuga 27

<210> 455

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 455

uaagucacau gacagggguc ca 22

<210> 456

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 456

cggagaggga gccacggagg ug 22

<210> 457

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 457

aagucagcgc uggggaggca uga 23

<210> 458

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 458

ucuuaagugg cugucgccga ggauugac 28

<210> 459

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 459

aagucacaug ucuagggguc cacu 24

<210> 460

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 460

aaguaaggcg gaagcagcug ucc 23

<210> 461

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 461

aucuuaagug gcugucgcug aggauugac 29

<210> 462

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 462

uaagucacau gacaggaguc cacu 24

<210> 463

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 463

aaguggcugu cgccgaggau ug 22

<210> 464

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 464

uaagucacau gacagggguc ca 22

<210> 465

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 465

acaagucacg ugggggggac ccg 23

<210> 466

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 466

auuuuaggcu guugccgggc guuugacu 28

<210> 467

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 467

auaagucacg ugacagggag ggg 23

<210> 468

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 468

aaguagguga cguccaggac u 21

<210> 469

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 469

cauaagucac guggggacgg cugu 24

<210> 470

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 470

uaaguaaggc ggaagagcuc uagcua 26

<210> 471

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 471

aguagcugac gucaaggauu gac 23

<210> 472

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 472

acaagucacg ugggggggac ccg 23

<210> 473

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 473

uaaguaagca uggcgggcgg cgac 24

<210> 474

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 474

aucgagacac guggugccag cagcu 25

<210> 475

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 475

ucauccucag ccggaacuua cacaaaaugg 30

<210> 476

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 476

auaugucacg uggggaggcc cugcug 26

<210> 477

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 477

aaguaaggcg gaagcagcuc cacuuu 26

<210> 478

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 478

aaguaaggcg gaaguuucuc cacu 24

<210> 479

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 479

uaaguaguug aggcggacgg uggc 24

<210> 480

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 480

uaaguaguug aggcggacgg ugg 23

<210> 481

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 481

gaccuucggc gucgggggggg ucggggg 27

<210> 482

<211> 19

<212> RNA

<213> Alphatorquevirus sp.

<400> 482

auccucggcg gaaccuaua 19

<210> 483

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 483

aucgaugacg uggggaggcg uacuau 26

<210> 484

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 484

uggcggagga gcacuuccgg cuug 24

<210> 485

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 485

aacaagucac guggggaggg uuggc 25

<210> 486

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 486

aggggucagc cuuccgcuuu a 21

<210> 487

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 487

aggggucagc cuuccgcuuu a 21

<210> 488

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 488

uaagucagcu ucggggaggg ucac 24

<210> 489

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 489

ucauaaguga cguagcuggu gucugcu 27

<210> 490

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 490

cauccucggc ggaagcuaca caa 23

<210> 491

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 491

ggccccguca cgugacuuac cac 23

<210> 492

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 492

ucauccucgg cggaagcuac acaa 24

<210> 493

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 493

ggccccguca cgugauuugu cac 23

<210> 494

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 494

ccgggucaua ggucacaccu acgucac 27

<210> 495

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 495

ggcugccgcc ccccccgggg aaaggggg 28

<210> 496

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 496

auccucgucc acgugacugu ga 22

<210> 497

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 497

gagcacuucc ggcuugccca a 21

<210> 498

<211> 20

<212> RNA

<213> Alphatorquevirus sp.

<400> 498

caauccucuu acguggccug 20

<210> 499

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 499

cgagguuaag ggccaauucg ggcu 24

<210> 500

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 500

gggccccgc ggggcucucg gcg 23

<210> 501

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 501

gcggggcgug gccguaucag aaaaugg 27

<210> 502

<211> 19

<212> RNA

<213> Alphatorquevirus sp.

<400> 502

ccucggucac guggccugu 19

<210> 503

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 503

ccucgucacg ugaccugacg ucacag 26

<210> 504

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 504

ccguccgcgg cgagagcgcg agcga 25

<210> 505

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 505

auccucguca cgugaccuga cgucacg 27

<210> 506

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 506

ccguccgcgg cgagagcgcg agcga 25

<210> 507

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 507

ccguccgcgg cgagagcgcg agcga 25

<210> 508

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 508

auccucguca cgugaccuga cgucacg 27

<210> 509

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 509

cugacccgcg ugacugguca cguga 25

<210> 510

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 510

cggggcgugg ccguauuaga aaaugg 26

<210> 511

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 511

aguaaggcgg aaccaggcug ucacccugu 29

<210> 512

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 512

uagcugaccc gcgugacuug ucac 24

<210> 513

<211> 19

<212> RNA

<213> Alphatorquevirus sp.

<400> 513

ggucagccuc cgcuuugca 19

<210> 514

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 514

gcugacccgc guggcuuguc acgugagu 28

<210> 515

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 515

ggcccguccg cggcgagagc gcgag 25

<210> 516

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 516

uccgggugcg ucuggggggcc gccauuu 27

<210> 517

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 517

auccucgacc acgugacugu g 21

<210> 518

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 518

auaaaccaga ggggugacga augguagagu 30

<210> 519

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 519

caaagucacg uggugacggc cau 23

<210> 520

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 520

guaggcgcgu uuuaaugacg ucacgg 26

<210> 521

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 521

uaggcgcguu uuaaugacgu cacggcag 28

<210> 522

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 522

aucccugguc acgugacucu gacgucacg 29

<210> 523

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 523

gcgugugggg gccgccauuu uagcuu 26

<210> 524

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 524

ucauccucgu cacgugaccu gacgucacg 29

<210> 525

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 525

cgccaucuug ugacuuccuu ccgcuuuuu 29

<210> 526

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 526

uagaguuuuu uccgcccguc cg 22

<210> 527

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 527

gucccugguc acgugauugu gac 23

<210> 528

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 528

cauuuuguga cuuccuuccg cuuuuu 26

<210> 529

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 529

ucauccucgu cacgugaccu gacgucacg 29

<210> 530

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 530

ccgccaucuu gugacuuccu uccgcuuuuu 30

<210> 531

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 531

auccucguca cgugaccuga cgucacg 27

<210> 532

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 532

cgccaucuug ugacuuccuu ccgcuuuuuc 30

<210> 533

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 533

ccguccgcgg cgagagcgcg agcga 25

<210> 534

<211> 29

<212> RNA

<213> Alphatorquevirus sp.

<400> 534

ucauccucgu cacgugaccu gacgucacg 29

<210> 535

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 535

cuuagugacg ucacggcagc cau 23

<210> 536

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 536

gucccugguc acgugacaug acguc 25

<210> 537

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 537

cacagccggu caugcuugca caaa 24

<210> 538

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 538

acagccuguc augcuugcac aa 22

<210> 539

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 539

caccaucagc cacaccuacu caaa 24

<210> 540

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 540

cgggucgccg ccauauuugg ucacguga 28

<210> 541

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 541

cauccucggc ggaagcuaca caa 23

<210> 542

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 542

cauccucggc ggaagcuaca caa 23

<210> 543

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 543

ggccccguca cgugacuuac cac 23

<210> 544

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 544

ucauccucgg cggaagcuac acaa 24

<210> 545

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 545

ggccccguca cgugauuugu cac 23

<210> 546

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 546

gucauccuca ccauaacugg cacaa 25

<210> 547

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 547

cauccucggc ggaagcuaca caa 23

<210> 548

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 548

ggccccguca cgugacuuac cac 23

<210> 549

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 549

auccucgacg guaaccgcaa acaug 25

<210> 550

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 550

ggccccgaca ugugacucgu c 21

<210> 551

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 551

agcacuuccg aauggcugag uuuucca 27

<210> 552

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 552

auccucaccg gaacugacac aa 22

<210> 553

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 553

cauccucggc gguaaccgca aagaug 26

<210> 554

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 554

uaggccccga caugugacuc gu 22

<210> 555

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 555

acagccuucc gcuuugcaca a 21

<210> 556

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 556

cauccucggc gguaaccgca aagaugg 27

<210> 557

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 557

uaggccccga caugugacuc guc 23

<210> 558

<211> 20

<212> RNA

<213> Alphatorquevirus sp.

<400> 558

uccucggcgg uaaccgcaaa 20

<210> 559

<211> 21

<212> RNA

<213> Alphatorquevirus sp.

<400> 559

ggccccgaca ugugacucgu c 21

<210> 560

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 560

ggccccguca cgugacuuac cac 23

<210> 561

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 561

ucaaacaccc agcgacacca aaaaaugg 28

<210> 562

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 562

ccucgcccac gugacuuacc ac 22

<210> 563

<211> 20

<212> RNA

<213> Alphatorquevirus sp.

<400> 563

ccucggcgga accuauacaa 20

<210> 564

<211> 22

<212> RNA

<213> Alphatorquevirus sp.

<400> 564

gccccgucac gugauuuacc ac 22

<210> 565

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 565

gcggcggagc acuuccgcuu ugcccaaa 28

<210> 566

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 566

cauccucggc ggaagcuaca caa 23

<210> 567

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 567

ggccccguca cgugacuuac cac 23

<210> 568

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 568

caccgcacuu ccgugcuugc acaaa 25

<210> 569

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 569

ucgcugacac acgugucuug ucac 24

<210> 570

<211> 30

<212> RNA

<213> Alphatorquevirus sp.

<400> 570

cauuuuaagu aagcaccgcc uagggaugac 30

<210> 571

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 571

guaggccccg ucacguguca uaccac 26

<210> 572

<211> 25

<212> RNA

<213> Alphatorquevirus sp.

<400> 572

ggcggggcac uuccggcuug cccaa 25

<210> 573

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 573

cggcggagca cuuccggcuu gcccaa 26

<210> 574

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 574

caccaucagc cacaccuacu caaa 24

<210> 575

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 575

accaucagcc acaccuacuc aaa 23

<210> 576

<211> 28

<212> RNA

<213> Alphatorquevirus sp.

<400> 576

gacuccgaga ugccauugga cacugagg 28

<210> 577

<211> 18

<212> RNA

<213> Alphatorquevirus sp.

<400> 577

aguaggugcc guccagca 18

<210> 578

<211> 27

<212> RNA

<213> Alphatorquevirus sp.

<400> 578

aaguaggccc cgcuacguca ucaucac 27

<210> 579

<211> 23

<212> RNA

<213> Alphatorquevirus sp.

<400> 579

aaggcggaag agcugcucua uau 23

<210> 580

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 580

caauccuccc acguggccug ucac 24

<210> 581

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 581

aaggcggaac caggcuguca ccccgu 26

<210> 582

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 582

aaggcggaac caggcuguca ccccgu 26

<210> 583

<211> 24

<212> RNA

<213> Alphatorquevirus sp.

<400> 583

cauccucacc ggaacuggua caaa 24

<210> 584

<211> 26

<212> RNA

<213> Alphatorquevirus sp.

<400> 584

uaggccccgc cacgucacuu gucacg 26

<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 gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc gccgcc 156

<210> 802

<211> 150

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 802

ccgagcgtta gcgaggagtg cgaccctacc ccctgggccc acttcttcgg agccgcgcgc 60

tacgccttcg gctgcgcgcg gcacctcaga cccccgctcg 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 cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc 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 gggctccgcc ccccccgcg 60

catgcgcggg gccccccccc gcgggggggct ccgccccccg gtcccccccc 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 ggctgccgcc ccccccgcg 60

catgcgcgcg gggcccccccc ccgcgggggg ctccgccccc cggcccccccc ccccg 115

<210> 806

<211> 104

<212> DNA

<213> Artificial Sequence

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

<400> 806

cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc cccccccgcg gggggctccg cccccggcc cccc 104

<210> 807

<211> 108

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 807

ggcggcggcg cgcgcgctac gcgcgcgcgc cggggagctc tgccccccccc cgcgcatgcg 60

cgcgggtccc ccccccgcgg ggggctccgc cccccggtcc ccccccccg 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

<211> 3264

<212> DNA

<213> Gammatorquevirus sp.

<400> 878

taaaatggcg gcaaccaatc attttatact ttcactttcc aattacaagc cgccacgtca 60

cagaacaggg gtggagactt taaaactata taaccaagtg atgtgacgaa tggctgagtt 120

taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgcccg 180

tgggcgggag cccgaggtga gtgaaaccac cgaggtctag gggcaattcg ggctagggca 240

gtctagcgga acgggcaaga aacttaaaat atgttttgtt tcagatgcag acacctgctt 300

cacagataag ctcagacgac ttctttgtac acactccatt taatgcagta actaaacagc 360

aaatatggat gtctcaaatt gctgatggac atgacaacat ttgtcactgc caccgtcctt 420

ttgctcacct gcttgctaat atttttcctc ctggtcataa agacagggat cttaccatta 480

atcaaatact tgctagagat cttacagaaa catgccattc tggtggagac gaaggaacaa 540

gcggtggtgg ggtcgccgct tccgctaccg ccgctacaac aaatataaaa ccagaaggag 600

acgcagaata cccagaagac gaaatagaag atttactaag acacgcagga gaagaaaaag 660

aaagaaggta agaagaaaac ttaaaaaaat tactattaaa caatggcagc cagattcagt 720

gaaaaaatgt aaaattaaag gatatagtac tttagttatg ggtgcacaag gaaaacaata 780

caactgttac acaaaccaag caagtgacta tgttcagcct aaagcaccac aaggtggggg 840

ctttggctgt gaagtattta atttaaaatg gctataccaa gaatatactg cacacagaaa 900

tatttggaca aaaacaaatg aatatacaga cctttgtaga tacactggag ctcaaataat 960

tttatacagg cacccagatg ttgattttat agtcagctgg gacaatcagc cacctttttt 1020

acttaacaaa tatacatatc cagaactgca accacaaaac cttttactag ctagaaggaa 1080

aagaattatt cttagtcaaa aatcaaaccc caaaggaaaa ctaagaatta aactaagaat 1140

accaccacca aaacaaatga taacaaaatg gttttttcaa agagactttt gtgatgtgaa 1200

tctgtttaaa ctatgtgctt ctgctgcttc tttccgctac ccaggtatca gtcatggagc 1260

tcaaagtact atttttctg catatgcttt aaacactgac ttttatcaat gcagtgactg 1320

gtgccaaact aacacagaaa ctggctacct aaacattaaa acacaacaaa tgccactatg 1380

gtttcattac agagagggtg gcaaagagaa atggtataaa tacaccaaca aagaacacag 1440

accatataca aatacatatc ttaaaagtat tagctataat gatggattgt tttctcctaa 1500

agccatgttt gcatttgaag taaaagcggg gggtgaagga acaacagaac caccacaagg 1560

cgcccaatta attgctaacc ttccactcat tgcactaaga tataatccac atgaagacac 1620

aggccatggc aatgaaattt accttacatc aacttttaaa ggtacatatg acaaacctaa 1680

agttactgat gctctatact ttaacaatgt acccctgtgg atgggatttt atggctactg 1740

ggactttata ttacaagaaa caaaaaacaa aggtgtcttt gatcaacata tgtttgttgt 1800

taaatgtcct gccttaaggc ccatatcaca agtcacaaaa caagtatact acccacttgt 1860

agacatggac ttttgttcag ggagactgcc atttgatgaa tatttatcca aagacattaa 1920

aagtcattgg tatcccactg cagaaagaca aacagttaca ataaataatt ttgttacagc 1980

aggtccatac atgcctaaat ttgaacccac agacaaagac agtacatggc aattaaacta 2040

tcactataaa tttttttta agtggggtgg tccacaagtc acagacccaa ctgttgaaga 2100

cccatgcagc agaaacaaat atcctgtccc cgatacaatg caacaaacaa tacaaattaa 2160

aaaccctgaa aagctgcacc cagcaaccct cttccatgac tgggacctta gaaggggctt 2220

cattacacaa gcagctatta aaagaatgtc agaaaacctc caaattgatt catctttcga 2280

atctgatggc acagaatcac ccaaaaaaaa gaaaagatgc accaaagaaa tcccaacaca 2340

aaaccaaaag caagaagaga tccaagaatg tctcctctca ctctgcgaag agcctacatg 2400
ccaagaagaa acagaggacc tccagctctt catccagcag cagcagcagc agcagtacaa 2460
gctcagaaaa aacctcttca aactcctcac tcacctgaaa aaaggacaga gaataagtca 2520
actacaaacg ggacttttag agtaatacca tttaaaccag gttttgaaca agaaacagaa 2580
aaagaacttg ccatagcttt ctgcagacca cctagaaaat ataaaaatga tccccctttt 2640
tatccctggt taccatggac acccccttgta cactttaacc ttaattacaa aggctaggcc 2700
aacactgttc acttagtggt gtatgtttaa taaagtttca cccccaaaaa aaaaaaaaaa 2760

aaaaaaaaaa aaaaaaaaaa taaaaaattg caaaaattcg gcgctcgcgc gcgctgcgcg 2820
cgcgcgagcg ccgtcacgcg ccggcgctcg cgcgccgcgc gtatgtgcta acacaccacg 2880
cacctagatt ggggtgcgcg cgctagcgcg cgcaccccaa tgcgccccgc cctcgttccg 2940
acccgcttgc gcgggtcgga ccacttcggg ctcgggggggg cgcgcctgcg gcgctttttt 3000
actaaacaga ctccgagccg ccatttggcc ccccctaagc tccgcccccc tcatgaatat 3060
tcataaagga aaccacataa ttagaattgc cgaccacaaa ctgccatatg ctaattagtt 3120
ccccttttac acagtaaaaa ggggaagtgg gggggcatag ccccccccaca ccccccgcgg 3180

gggggcaga gccccccccc gcacccccc cctacgtcac aatccacgcc cccgccgcca 3240
tcttgggtgc ggcagggcgg gggc 3264

<210> 879
<211> 128
<212> PRT
<213> Gammatorquevirus sp.
<400> 879

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

65 70 75 80

Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala

85 90 95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp

100 105 110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Arg

115 120 125

<210> 880

<211> 272

<212> PRT

<213> Gammatorquevirus sp.

<400> 880

Met Gln Thr Pro Ala Ser Gln Ile Ser Ser Asp Asp Phe Phe Val His

1 5 10 15

Thr Pro Phe Asn Ala Val Thr Lys Gln Gln Ile Trp Met Ser Gln Ile

20 25 30

Ala Asp Gly His Asp Asn Ile Cys His Cys His Arg Pro Phe Ala His

35 40 45

Leu Leu Ala Asn Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr

50 55 60

Ile Asn Gln Ile Leu Ala Arg Asp Leu Thr Glu Thr Cys His Ser Gly

65 70 75 80

Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala

85 90 95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp

100 105 110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Ser

115 120 125

Gly Val Val His Lys Ser Gln Thr Gln Leu Leu Lys Thr His Ala Ala

130 135 140

Glu Thr Asn Ile Leu Ser Pro Ile Gln Cys Asn Lys Gln Tyr Lys Leu

145 150 155 160

Lys Thr Leu Lys Ser Cys Thr Gln Gln Pro Ser Ser Met Thr Gly Thr
165 170 175
Leu Glu Gly Ala Ser Leu His Lys Gln Leu Leu Lys Glu Cys Gln Lys
180 185 190
Thr Ser Lys Leu Ile His Leu Ser Asn Leu Met Ala Gln Asn His Pro
195 200 205
Lys Lys Arg Lys Asp Ala Pro Lys Lys Ser Gln His Lys Thr Lys Ser
210 215 220

Lys Lys Arg Ser Lys Asn Val Ser Ser His Ser Ala Lys Ser Leu His
225 230 235 240
Ala Lys Lys Lys Gln Arg Thr Ser Ser Ser Ser Ser Ser Ser Ser Ser
245 250 255
Ser Ser Ser Thr Ser Ser Glu Lys Thr Ser Ser Asn Ser Ser Leu Thr
260 265 270

<210> 881
<211> 261
<212> PRT
<213> Gammatorquevirus sp.
<400> 881

Met Gln Thr Pro Ala Ser Gln Ile Ser Ser Asp Asp Phe Phe Val His

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

65 70 75 80
Gly Asp Glu Gly Thr Ser Gly Gly Gly Val Ala Ala Ser Ala Thr Ala

85 90 95

Ala Thr Thr Asn Ile Lys Pro Glu Gly Asp Ala Glu Tyr Pro Glu Asp

100 105 110

Glu Ile Glu Asp Leu Leu Arg His Ala Gly Glu Glu Lys Glu Arg Arg

115 120 125

Ile Thr Gln Lys Lys Glu Lys Met His Gln Arg Asn Pro Asn Thr Lys

130 135 140

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

145 150 155 160

Ala Tyr Met Pro Arg Arg Asn Arg Gly Pro Pro Ala Leu His Pro Ala

165 170 175

Ala Ala Ala Ala Ala Val Gln Ala Gln Lys Lys Pro Leu Gln Thr Pro

180 185 190

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

195 200 205

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

210 215 220

Glu Leu Ala Ile Ala Phe Cys Arg Pro Pro Arg Lys Tyr Lys Asn Asp

225 230 235 240

Pro Pro Phe Tyr Pro Trp Leu Pro Trp Thr Pro Leu Val His Phe Asn

245 250 255

Leu Asn Tyr Lys Gly

260

<210

> 882

<211> 67

<212> PRT

<213> Gammatorquevirus sp.

<400> 882

Met Asp Met Thr Thr Phe Val Thr Ala Thr Val Leu Leu Leu Thr Cys

1 5 10 15

Leu Leu Ile Phe Phe Leu Leu Val Ile Lys Thr Gly Ile Leu Pro Leu

20 25 30

Ile Lys Tyr Leu Leu Glu Ile Leu Gln Lys His Ala Ile Leu Val Glu
35 40 45
Thr Lys Glu Gln Ala Val Val Gly Ser Pro Leu Pro Leu Pro Pro Leu

50 55 60
Gln Gln Ile
65

<210> 883
<211> 677
<212> PRT
<213> Gammatorquevirus sp.
<400> 883

Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1 5 10 15
Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
20 25 30
Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
35 40 45

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

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

165 170 175

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

Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln Ser Thr Ile Phe Ser
245 250 255
Ala Tyr Ala Leu Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln
260 265 270
Thr Asn Thr Glu Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro
275 280 285
Leu Trp Phe His Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr
290 295 300

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

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

Thr Lys Asn Lys Gly Val Phe Asp Gln His Met Phe Val Val Lys Cys
420 425 430

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

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

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

Cys Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe
625 630 635 640
Ile Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe
645 650 655
Lys Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln

660 665 670

Thr Gly Leu Leu Glu

675

<210> 884

<211> 212

<212> PRT

<213> Gammatorquevirus sp.

<400>

884

Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg

1 5 10 15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg

20 25 30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg

35 40 45

Lys Arg Lys Lys Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val Glu

50 55 60

Asp Pro Cys Ser Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln Gln

65 70 75 80

Thr Ile Gln Ile Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu Phe

85 90 95

His Asp Trp Asp Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile Lys

100 105 110

Arg Met Ser Glu Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp Gly

115 120 125

Thr Glu Ser Pro Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro Thr

130 135 140

Gln Asn Gln Lys Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu Cys

145 150 155 160

Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe Ile

165 170 175

Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe Lys

180 185 190

Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln Thr
195 200 205

Gly Leu Leu Glu
210

<210> 885
<211> 119
<212> PRT
<213> Gammatorquevirus sp.
<400> 885

Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1 5 10 15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg
20 25 30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg
35 40 45

Lys Arg Lys Lys Asn His Pro Lys Lys Arg Lys Asp Ala Pro Lys Lys
50 55 60

Ser Gln His Lys Thr Lys Ser Lys Lys Arg Ser Lys Asn Val Ser Ser
65 70 75 80

His Ser Ala Lys Ser Leu His Ala Lys Lys Lys Gln Arg Thr Ser Ser
85 90 95

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Thr Ser Ser Glu Lys Thr
100 105 110

Ser Ser Asn Ser Ser Leu Thr
115

<210> 886
<211> 3176
<212> DNA
<213> Gammatorquevirus sp.
<400> 886

taaaatggcg ggagccaatc attttatact ttcactttcc aattaaaaat ggccacgtca 60

caaacaaggg gtggagccat ttaaactata taactaagtg gggtggcgaa tggctgagtt 120

taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgccca 180

tgggcgggag ccgaggtgag tgaaaccacc gaggtctagg ggcaattcgg gctagggcag 240

tctagcggaa cgggcaagaa acttaaaaca atatttgttt tacagatggt tagtatatcc 300

tcaagtgatt tttttaagaa aacgaaattt aatgaggaga cgcagaacca agtatggatg 360

tctcaaattg ctgactctca tgataatatc tgcagttgct ggcatccatt tgctcacctt 420

cttgcttcca tatttcctcc tggccacaaa gatcgtgatc ttactattaa ccaaattctt 480

ctaagagatt ataaagaaaa atgccattct ggtggagaag aaggagaaaa ttctggacca 540

acaacaggtt taattacacc aaaagaagaa gatatagaaa aagatggccc agaaggcgcc 600

gcagaagaag accatacaga cgccctgttc gccgccgccg tagaaaactt cgaaaggtaa 660

agagaaaaaa aaaatcttta attgttagac aatggcaacc agacagtata agaacttgta 720

aaattatagg acagtcagct atagttgttg gggctgaagg aaagcaaatg tactgttata 780

ctgtcaataa gttaattaat gtgcccccaa aaacaccata tgggggaggc tttggagtag 840

accaatacac actgaaatac ttatatgaag aatacagatt tgcacaaaac atttggacac 900

aatctaatgt actgaaagac ttatgcagat acataaatgt taagctaata ttctacagag 960

acaacaaaac agactttgtc ctttcctatg acagaaaccc acctttcaa ctaacaaaat 1020

ttacataccc aggagcacac ccacaacaaa tcatgcttca aaaacaccac aaattcatac 1080

tatcacaaat gacaaagcct aatggaagac taacaaaaaa actcaaaatt aaacctccta 1140

aacaaatgct ttctaaatgg ttcttttcaa aacaattctg taaataccct ttactatctc 1200

ttaaagcttc tgcactagac cttaggcact cttacctagg ctgctgtaat gaaaatccac 1260

aggtattttt ttattattta aaccatggat actacacaat aacaaactgg ggagcacaat 1320

cctcaacagc atacagacct aactccaagg tgacagacac aacatactac agatacaaaa 1380

atgacagaaa aaatattaac attaaaagcc atgaatacga aaaaagtata tcatatgaaa 1440

acggttattt tcaatctagt ttcttacaaa cacagtgcat atataccagt gagcgtggtg 1500

aagcctgtat agcagaaaaa ccactaggaa tagctattta caatccagta aaagacaatg 1560

gagatggtaa tatgatatac cttgtaagca ctctagcaaa cacttgggac cagcctccaa 1620

aagacagtgc tattttaata caaggagtac ccatatggct aggcttattt ggatatttag 1680

actactgtag acaaattaaa gctgacaaaa catggctaga cagtcatgta ctagtaattc 1740

aaagtcctgc tatttttact tacccaaatc caggagcagg caaatggtat tgtccactat 1800

cacaaagttt tataaatggc aatggtccgt ttaatcaacc acctacactg ctacaaaaag 1860

caaagtggtt tccacaaata caataccaac aagaaattat taatagcttt gtagaatcag 1920

gaccatttgt tcccaaatat gcaaatcaaa ctgaaagcaa ctgggaacta aaatataaat 1980
atgtttttac atttaagtgg ggtggaccac aattccatga accagaaatt gctgacccta 2040
gcaaacaaga gcagtatgat gtccccgata ctttctacca aacaatacaa 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 tcctctctct ctgcaaaaaa gataccttcc 2340

aagaaacaga gacacaagaa gacctccagc agctcatcaa gcagcagcag gagcagcagc 2400
tcctcctcaa gagaaacatc ctccagctca tccacaaact aaaagagaat caacaaatgc 2460
ttcagcttca cacaggcatg ttaccttaac cagatttaaa cctggatttg aagagcaaac 2520
agagagagaa ttagcaatta tatttcatag gccccctaga acctacaaag aggaccttcc 2580
attctatccc tggctaccac ctgcacccct tgtacaattt 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 ctagcgcgcg caccccaatg cgccccgccc tcgttccgac 2880
ccgcttgcgc gggtcggacc acttcgggct cgggggggcg cgcctgcggc gctttttac 2940
taaacagact ccgagccgcc atttggcccc ctaagctccg ccccccctcat gaatattcat 3000
aaaggaaacc acataattag aattgccgac cacaaactgc catatgctaa ttagttcccc 3060
ttttacaaag taaaagggga agtgaacata gccccacacc cgcaggggca aggccccgca 3120
cccctacgtc 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
35 40 45

Lys Leu Arg Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala
50 55 60

Asp Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln
65 70 75 80

Thr Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile
85 90 95

His Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys
100 105 110

Arg Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser
115 120 125

Glu Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr
130 135 140

Val Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu
145 150 155 160

Cys Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln
165 170 175

Gln Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn
180 185 190
Ile Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln
195 200 205
Leu His Thr Gly Met Leu Pro
210 215

<210> 893
<211> 129
<212> PRT
<213> Gammatorquevirus sp.
<400> 893

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 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
<211> 3696
<212> DNA
<213> Alphatorquevirus sp.

<400> 894

attttgttca gcccgccaat ttctctttca aacaggccaa tcagctacta cttcgtgcac 60

ttcctggggc gtgtcctgcc gctctatata agcagaggcg gtgacgaatg gtagagtttt 120

tcttggcccg tccgcggcga gagcgcgagc gaagcgagcg atcgagcgtc ccgagggcgg 180

gtgccggagg tgagtttaca caccgcagtc aaggggcaat tcgggctcgg gactggccgg 240

gctatgggca agattcttaa aaaattcccc cgatcccttt gccgccagga cataaaaaca 300

tgccgtggag accgccggtc catagtgtcc aggggcgaga ggatcagtgg ttcgcaagct 360

tttttcacgg ccacgattcg ttttgcggct gcggtgaccc tcttggccat attaatagca 420

ttgctcatcg ctttcctcgc gccggtccac caaggccccc tccggggcta gatcagccta 480

acccccggga gcagggcccg gccggacccg gagggccgcc cgccatcttg gccctgccgg 540

ctccgcccgc ggagcctgac gacccgcagc cacggcgtgg tggtggggac ggtggcgccg 600

ccgctggcgc cgcagacgac catacacaac gagactacga cgaagaagag ctagacgagc 660

ttttccgcgc cgccgccgaa gacgatttgt aagtaggaga tggcgccggc cttacaggcg 720

caggaggaga cgcgggcgac gcagacgcag acgcagacgc agacataagc ccaccctaat 780

actcagacag tggcaacctg actgtatcag acactgtaaa ataacaggat ggatgcccct 840

cattatctgt ggaaaggggt ccacccagtt caactacatc acccacgcgg acgatatcac 900

ccccagggga gcctcctacg gaggcaattt cacaaacatg actttctccc tggaggccat 960

atatgaacag ttcctatacc acagaaacag gtggtcggcc tctaaccacg acctagaact 1020

gtgcagatac aaggggacca ccttaaaact ctacagacac ccagaagtag actacatagt 1080

tacctacagc agaacaggac cctttgaaat cagccacatg acctacctca gcactcaccc 1140

catgctaatg ctgctaaaca agcaccacat tgtggtgccc agcttaaaga ctaagcccag 1200

aggcagaaag gccataaaag tcaggataag gcccccaaaa ctcatgaaca acaagtggta 1260

cttcaccaga gacttctgta acataggcct cttccagctc tgggccacag gcttagaact 1320

cagaaacccc tggctcagaa tgagcaccct gagcccctgc ataggcttta atgtcctcaa 1380

aaacagcatt tacacaaacc tcagcaacct gccacaatac aaaaacgaaa gactaaacat 1440

cattaacaac atacttcacc cacaagaaat tacaggtaca aacaacaaaa agtggcagta 1500

cacatacacc aaactcatgg cccctattta ctattcagca aacagggcca gcacctatga 1560

ctgggaaaat tacagcaaag aaacaaacta caataataca tatgttaaat ttacccagaa 1620

aagacaggaa aaactaacta aaattagaaa agagtggcag atgctttatc cacaacaacc 1680

cacagcactg ccagactcct atgacctcct acaagagtat ggcctctaca gtccatacta 1740

cctaaacccc acaagaataa acctagactg gatgacccca tacacacacg tcagatacaa 1800
tcccctagta gacaagggct ttggaaacag aatatacatc cagtggtgct cagaagcaga 1860
tgttagctac aacaggacaa aatccaagtg tctgctacaa gacatgcccc tgtttttcat 1920
gtgctatggc tacatagact gggcaataaa aaacactgga gtgtcatctc tagtgaagga 1980
cgccagaatc tgcatcaggt gtccctacac agagccacaa ctagttggct ccacagaaga 2040
cataggcttt gtacccatct cagaaacctt catgaggggc gacatgccgg tacttgcacc 2100
atacataccg ttaagctggt tttgcaagtg gtatcccaac atagctcacc aaaaggaagt 2160

ccttgagtca atcatttcct gcagcccctt catgccccgt gaccaagaca tgaacggttg 2220
ggatatcaca atcggttaca aaatggactt cttatggggc ggttcccctc tcccctcaca 2280
gccaatcgac gacccctgcc agcagggaac ccacccgatt cccgaccccg ataaacaccc 2340
tcgcctccta caagtctcga acccgaaact actcggaccg aggacagtgt tccacaagtg 2400
ggacatcaga cgtgggcagt ttagcaaaag aagtattaag agagtgtcag aatactcaag 2460
cgatgatgaa tctcttgcgc caggtctccc atcaaagcga aacaagctcg actcggcgtt 2520
ccgaggagaa aatcgagagc aaaaagaatg ctattctctc ctcaaagcgc tcgaggaaga 2580

agagacccca gaagaagaag aaccagcacc ccaagaaaaa gcccagaaag aggagctact 2640
ccaccagctc cagctccaga gacgccacca gcgagtcctc agacgagggc tcaagctcgt 2700
ctttacagac atcctccgac tccgccaggg agtccactgg aacccggagc tcacatagcg 2760
cccccacctt acataccaga cctgcttttt cccaatactg gtaaaaaaaa aaaattctct 2820
cccttcgatt gggagacaga ggcgcaaata gcggggtgga tgcggcggcc catgcgcttc 2880
tatccctcag acacccctca ctacccgtgg ctacccccccg agcgagatat cccgaaaata 2940
tgtaacataa acttcaaaat aaagcttcaa gagtgagtga ttcgaggccc tcctctgttc 3000

acttagcggt gtctacctct taaggtcact aagcactccg agcgtaagcg aggagtgcga 3060
ccctctacca aggggcaact tcctcggggt ccggcgctac gcgcttcgcg ctgcgccgga 3120
catctcggac ccctcgaccc gaatcgcttg cgcgattcgg acctgcggcc tcgggggggt 3180
cgggggcttt actaaacaga ctccgaggtg ccattggaca ctgtaggggg tgaacagcaa 3240
cgaaagtgag tggggccaga cttcgccata aggcctttat cttcttgcca ttggatagtg 3300
acttccgggt ccgcctgggg gccgccattt tagcttcggc cgccatttta ggccctcgcg 3360
ggcctccgta ggcgcgcttt agtgacgtca cggcagccat tttgtcgtga cgtttgagac 3420

acgtgatggg ggcgtgccta aacccggaag catccctggt cacgtgactc tgacgtcacg 3480

gcggccatct tgtgctgtcc gccatcttgt aacttccttc cgctttttca aaaaaaaaga 3540

ggaagtgtga cgtagcggcg gggggggcggc gcgcttcgcg cgccgcccac cagggggcgc 3600

tgcgcgcccc ccgcgcatgc gcaggggcct ctcgaggggc tccgcccccc ccccgtgcta 3660

aatttaccgc gcatgcgcga ccacgccccc gccgcc 3696

<210> 895

<211> 130

<212> PRT

<213> Alphatorquevirus sp.

<400> 895

Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln

1 5 10 15

Trp Phe Ala Ser Phe Phe His Gly His Asp Ser Phe Cys Gly Cys Gly

20 25 30

Asp Pro Leu Gly His Ile Asn Ser Ile Ala His Arg Phe Pro Arg Ala

35 40 45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Asp Gln Pro Asn Pro Arg Glu

50 55 60

Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro

65 70 75 80

Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly

85 90 95

Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp

100 105 110

Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp

115 120 125

Asp Leu

130

<210> 896

<211> 303

<212

> PRT

<213> Alphatorquevirus sp.

<400> 896

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

50 55 60
Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro
65 70 75 80
Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly
85 90 95
Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp
100 105 110
Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp

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

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

Lys Lys Lys Pro Arg Lys Arg Ser Tyr Ser Thr Ser Ser Ser Ser Arg

245 250 255

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

260 265 270

Thr Ser Ser Asp Ser Ala Arg Glu Ser Thr Gly Thr Arg Ser Ser His

275 280 285

Ser Ala Pro Thr Leu His Thr Arg Pro Ala Phe Ser Gln Tyr Trp

290 295 300

<210> 897

<211> 292

<212> PRT

<213> Alphatorquevirus sp.

<400> 897

Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln

1 5 10 15

Trp Phe Ala Ser Phe Phe His Gly His Asp Ser Phe Cys Gly Cys Gly

20 25 30

Asp Pro Leu Gly His Ile Asn Ser Ile Ala His Arg Phe Pro Arg Ala

35 40 45

Gly Pro Pro Arg Pro Pro Pro Gly Leu Asp Gln Pro Asn Pro Arg Glu

50 55 60

Gln Gly Pro Ala Gly Pro Gly Gly Pro Pro Ala Ile Leu Ala Leu Pro

65 70 75 80

Ala Pro Pro Ala Glu Pro Asp Asp Pro Gln Pro Arg Arg Gly Gly Gly

85 90 95

Asp Gly Gly Ala Ala Ala Gly Ala Ala Asp Asp His Thr Gln Arg Asp

100 105 110

Tyr Asp Glu Glu Glu Leu Asp Glu Leu Phe Arg Ala Ala Ala Glu Asp

115 120 125

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

130 135 140

Arg Lys Ser Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Ala Arg
145 150 155 160
Gly Arg Arg Asp Pro Arg Arg Arg Arg Thr Ser Thr Pro Arg Lys Ser
165 170 175
Pro Glu Arg Gly Ala Thr Pro Pro Ala Pro Ala Pro Glu Thr Pro Pro
180 185 190

Ala Ser Pro Gln Thr Arg Ala Gln Ala Arg Leu Tyr Arg His Pro Pro
195 200 205
Thr Pro Pro Gly Ser Pro Leu Glu Pro Gly Ala His Ile Ala Pro Pro
210 215 220
Pro Tyr Ile Pro Asp Leu Leu Phe Pro Asn Thr Gly Lys Lys Lys Lys
225 230 235 240
Phe Ser Pro Phe Asp Trp Glu Thr Glu Ala Gln Ile Ala Gly Trp Met
245 250 255

Arg Arg Pro Met Arg Phe Tyr Pro Ser Asp Thr Pro His Tyr Pro Trp
260 265 270
Leu Pro Pro Glu Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys
275 280 285
Ile Lys Leu Gln
290

<210> 898
<211> 180
<212> PRT
<213> Alphatorquevirus sp.
<400> 898
Met Pro Trp Arg Pro Pro Val His Ser Val Gln Gly Arg Glu Asp Gln
1 5 10 15

Trp Ser Pro Ile Lys Ala Lys Gln Ala Arg Leu Gly Val Pro Arg Arg
20 25 30
Lys Ser Arg Ala Lys Arg Met Leu Phe Ser Pro Gln Ser Ala Arg Gly
35 40 45
Arg Arg Asp Pro Arg Arg Arg Arg Thr Ser Thr Pro Arg Lys Ser Pro

50 55 60

Glu Arg Gly Ala Thr Pro Pro Ala Pro Ala Pro Glu Thr Pro Pro Ala
65 70 75 80

Ser Pro Gln Thr Arg Ala Gln Ala Arg Leu Tyr Arg His Pro Pro Thr
85 90 95

Pro Pro Gly Ser Pro Leu Glu Pro Gly Ala His Ile Ala Pro Pro Pro
100 105 110

Tyr Ile Pro Asp Leu Leu Phe Pro Asn Thr Gly Lys Lys Lys Lys Phe
115 120 125

Ser Pro Phe Asp Trp Glu Thr Glu Ala Gln Ile Ala Gly Trp Met Arg
130 135 140

Arg Pro Met Arg Phe Tyr Pro Ser Asp Thr Pro His Tyr Pro Trp Leu
145 150 155 160

Pro Pro Glu Arg Asp Ile Pro Lys Ile Cys Asn Ile Asn Phe Lys Ile
165 170 175

Lys Leu Gln Glu
180

<210> 899

<211> 49

<212> PRT

<213> Alphatorquevirus sp.

<400> 899

Ile Val Ser Arg Gly Glu Arg Ile Ser Gly Ser Gln Ala Phe Phe Thr
1 5 10 15

Ala Thr Ile Arg Phe Ala Ala Ala Val Thr Leu Leu Ala Ile Leu Ile

20 25 30

Ala Leu Leu Ile Ala Phe Leu Ala Pro Val His Gln Gly Pro Leu Arg
35 40 45

Gly

<210> 900

<211> 728

<212> PRT

<213> Alphatorquevirus sp.

<400> 900

Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr
35 40 45

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg
50 55 60

His Lys Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg
65 70 75 80

His Cys Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly
85 90 95

Ser Thr Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg
100 105 110

Gly Ala Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu
115 120 125

Ala Ile Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser
130 135 140

Asn His Asp Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu
145 150 155 160

Tyr Arg His Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly
165 170 175

Pro Phe Glu Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu
180 185 190

Met Leu Leu Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys
195 200 205

Pro Arg Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu
210 215 220

Met Asn Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu
225 230 235 240
Phe Gln Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg
245 250 255
Met Ser Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser
260 265 270
Ile Tyr Thr Asn Leu Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu
275 280 285

Asn Ile Ile Asn Asn Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn
290 295 300
Asn Lys Lys Trp Gln Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr
305 310 315 320
Tyr Ser Ala Asn Arg Ala Ser Thr Tyr Asp Trp Glu Asn Tyr Ser Lys
325 330 335
Glu Thr Asn Tyr Asn Asn Thr Tyr Val Lys Phe Thr Gln Lys Arg Gln
340 345 350

Glu Lys Leu Thr Lys Ile Arg Lys Glu Trp Gln Met Leu Tyr Pro Gln
355 360 365
Gln Pro Thr Ala Leu Pro Asp Ser Tyr Asp Leu Leu Gln Glu Tyr Gly
370 375 380
Leu Tyr Ser Pro Tyr Tyr Leu Asn Pro Thr Arg Ile Asn Leu Asp Trp
385 390 395 400
Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro Leu Val Asp Lys Gly
405 410 415

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

465 470 475 480

Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile Gly Phe Val Pro Ile
485 490 495
Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile
500 505 510
Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys
515 520 525
Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp
530 535 540

Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe
545 550 555 560
Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys
565 570 575
Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu
580 585 590
Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His
595 600 605

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg
610 615 620
Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro
625 630 635 640
Ser Lys Arg Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu
645 650 655
Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr
660 665 670

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu
675 680 685
Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg
690 695 700
Arg Gly Leu Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly
705 710 715 720

Val His Trp Asn Pro Glu Leu Thr
725

<210> 901

<211> 197

<212> PRT

<213> Alphatorquevirus sp.

<400> 901

Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Pro Ile Asp Asp Pro Cys Gln Gln Gly
35 40 45

Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu Leu Gln Val
50 55 60

Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His Lys Trp Asp
65 70 75 80

Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg Val Ser Glu
85 90 95

Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro Ser Lys Arg
100 105 110

Asn Lys Leu Asp Ser Ala Phe Arg Gly Glu Asn Arg Glu Gln Lys Glu
115 120 125

Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr Pro Glu Glu
130 135 140

Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu Leu Leu His
145 150 155 160

Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg Arg Gly Leu
165 170 175

Lys Leu Val Phe Thr Asp Ile Leu Arg Leu Arg Gln Gly Val His Trp
180 185 190

Asn Pro Glu Leu Thr
195

<210> 902

<211> 145

<212> PRT

<213> Alphatorquevirus sp.

<400> 902

Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30

Arg Arg Arg Arg Arg Arg Phe Val Ser His Gln Ser Glu Thr Ser Ser
35 40 45

Thr Arg Arg Ser Glu Glu Lys Ile Glu Ser Lys Lys Asn Ala Ile Leu
50 55 60

Ser Ser Lys Arg Ser Arg Lys Lys Arg Pro Gln Lys Lys Lys Asn Gln
65 70 75 80

His Pro Lys Lys Lys Pro Arg Lys Arg Ser Tyr Ser Thr Ser Ser Ser
85 90 95

Ser Arg Asp Ala Thr Ser Glu Ser Ser Asp Glu Gly Ser Ser Ser Ser
100 105 110

Leu Gln Thr Ser Ser Asp Ser Ala Arg Glu Ser Thr Gly Thr Arg Ser
115 120 125

Ser His Ser Ala Pro Thr Leu His Thr Arg Pro Ala Phe Ser Gln Tyr
130 135 140

Trp
145

<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

<211> 677

<212> PRT

<213> Gammatorquevirus sp.

<400> 919

Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg
1 5 10 15
Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg

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

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

145 150 155 160

Val Ser Trp Asp Asn Gln Pro Pro Phe Leu Leu Asn Lys Tyr Thr Tyr

165 170 175

Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu Ala Arg Arg Lys Arg Ile

180 185 190

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

195 200 205

Arg Ile Pro Pro Pro Lys Gln Met Ile Thr Lys Trp Phe Phe Gln Arg

210 215 220

Asp Phe Cys Asp Val Asn Leu Phe Lys Leu Cys Ala Ser Ala Ala Ser

225 230 235 240

Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln Ser Thr Ile Phe Ser

245 250 255

Ala Tyr Ala Leu Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln

260 265 270

Thr Asn Thr Glu Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro

275 280 285

Leu Trp Phe His Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr

290 295 300

Thr Asn Lys Glu His Arg Pro Tyr Thr Asn Thr Tyr Leu Lys Ser Ile

305 310 315 320

Ser Tyr Asn Asp Gly Leu Phe Ser Pro Lys Ala Met Phe Ala Phe Glu

325 330 335

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

340 345 350

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

355 360 365

Asp Thr Gly His Gly Asn Glu Ile Tyr Leu Thr Ser Thr Phe Lys Gly

370 375 380

Thr Tyr Asp Lys Pro Lys Val Thr Asp Ala Leu Tyr Phe Asn Asn Val

385 390 395 400

Pro Leu Trp Met Gly Phe Tyr Gly Tyr Trp Asp Phe Ile Leu Gln Glu

405 410 415

Thr Lys Asn Lys Gly Val Phe Asp Gln His Met Phe Val Val Lys Cys

420 425 430

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

435 440 445

Leu Val Asp Met Asp Phe Cys Ser Gly Arg Leu Pro Phe Asp Glu Tyr

450 455 460

Leu Ser Lys Asp Ile Lys Ser His Trp Tyr Pro Thr Ala Glu Arg Gln

465 470 475 480

Thr Val Thr Ile Asn Asn Phe Val Thr Ala Gly Pro Tyr Met Pro Lys

485 490 495

Phe Glu Pro Thr Asp Lys Asp Ser Thr Trp Gln Leu Asn Tyr His Tyr

500 505 510

Lys Phe Phe Phe Lys Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val

515 520 525

Glu Asp Pro Cys Ser Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln

530 535 540

Gln Thr Ile Gln Ile Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu

545 550 555 560

Phe His Asp Trp Asp Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile

565 570 575

Lys Arg Met Ser Glu Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp

580 585 590

Gly Thr Glu Ser Pro Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro

595 600 605

Thr Gln Asn Gln Lys Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu

610 615 620

Cys Glu Glu Pro Thr Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe

625 630 635 640

Ile Gln Gln Gln Gln Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe

645 650 655

Lys Leu Leu Thr His Leu Lys Lys Gly Gln Arg Ile Ser Gln Leu Gln

660 665 670

Thr Gly Leu Leu Glu

675

<210> 920

<211> 59

<212> PRT

<213> Gammatorquevirus sp.

<400> 920

Met Pro Phe Trp Trp Arg Arg Arg Asn Lys Arg Trp Trp Gly Arg Arg

1 5 10 15

Phe Arg Tyr Arg Arg Tyr Asn Lys Tyr Lys Thr Arg Arg Arg Arg Arg

20 25 30

Ile Pro Arg Arg Arg Asn Arg Arg Phe Thr Lys Thr Arg Arg Arg Arg

35 40 45

Lys Arg Lys Lys Val Arg Arg Lys Leu Lys Lys

50 55

<210> 921

<211> 201

<212> PRT

<213> Gammatorquevirus sp.

<400> 921

Ile Thr Ile Lys Gln Trp Gln Pro Asp Ser Val Lys Lys Cys Lys Ile

1 5 10 15

Lys Gly Tyr Ser Thr Leu Val Met Gly Ala Gln Gly Lys Gln Tyr Asn

20 25 30

Cys Tyr Thr Asn Gln Ala Ser Asp Tyr Val Gln Pro Lys Ala Pro Gln

35 40 45

Gly Gly Gly Phe Gly Cys Glu Val Phe Asn Leu Lys Trp Leu Tyr Gln

50 55 60

Glu Tyr Thr Ala His Arg Asn Ile Trp Thr Lys Thr Asn Glu Tyr Thr

65 70 75 80

Asp Leu Cys Arg Tyr Thr Gly Ala Gln Ile Ile Leu Tyr Arg His Pro

85 90 95

Asp Val Asp Phe Ile Val Ser Trp Asp Asn Gln Pro Pro Phe Leu Leu

100 105 110

Asn Lys Tyr Thr Tyr Pro Glu Leu Gln Pro Gln Asn Leu Leu Leu Ala

115 120 125

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

130 135 140

Leu Arg Ile Lys Leu Arg Ile Pro Pro Pro Lys Gln Met Ile Thr Lys

145 150 155 160

Trp Phe Phe Gln Arg Asp Phe Cys Asp Val Asn Leu Phe Lys Leu Cys

165 170 175

Ala Ser Ala Ala Ser Phe Arg Tyr Pro Gly Ile Ser His Gly Ala Gln

180 185 190

Ser Thr Ile Phe Ser Ala Tyr Ala Leu

195 200

<210> 922

<211> 96

<212> PRT

<213> Gammatorquevirus sp.

<400> 922

Asn Thr Asp Phe Tyr Gln Cys Ser Asp Trp Cys Gln Thr Asn Thr Glu

1 5 10 15

Thr Gly Tyr Leu Asn Ile Lys Thr Gln Gln Met Pro Leu Trp Phe His

20 25 30

Tyr Arg Glu Gly Gly Lys Glu Lys Trp Tyr Lys Tyr Thr Asn Lys Glu

35 40 45

His Arg Pro Tyr Thr Asn Thr Tyr Leu Lys Ser Ile Ser Tyr Asn Asp

50 55 60

Gly Leu Phe Ser Pro Lys Ala Met Phe Ala Phe Glu Val Lys Ala Gly

65 70 75 80

Gly Glu Gly Thr Thr Glu Pro Pro Gln Gly Ala Gln Leu Ile Ala Asn

85 90 95

<210> 923

<211> 161

<212> PRT

<213> Gammatorquevirus sp.

<400> 923

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

1 5 10 15

Gly Asn Glu Ile Tyr Leu Thr Ser Thr Phe Lys Gly Thr Tyr Asp Lys

20 25 30

Pro Lys Val Thr Asp Ala Leu Tyr Phe Asn Asn Val Pro Leu Trp Met

35 40 45

Gly Phe Tyr Gly Tyr Trp Asp Phe Ile Leu Gln Glu Thr Lys Asn Lys

50 55 60

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

65 70 75 80

Pro Ile Ser Gln Val Thr Lys Gln Val Tyr Tyr Pro Leu Val Asp Met

85 90 95

Asp Phe Cys Ser Gly Arg Leu Pro Phe Asp Glu Tyr Leu Ser Lys Asp

100 105 110

Ile Lys Ser His Trp Tyr Pro Thr Ala Glu Arg Gln Thr Val Thr Ile

115 120 125

Asn Asn Phe Val Thr Ala Gly Pro Tyr Met Pro Lys Phe Glu Pro Thr

130 135 140

Asp Lys Asp Ser Thr Trp Gln Leu Asn Tyr His Tyr Lys Phe Phe Phe

145 150 155 160

Lys

<210> 924

<211> 160

<212> PRT

<213> Gammatorquevirus sp.

<400> 924

Trp Gly Gly Pro Gln Val Thr Asp Pro Thr Val Glu Asp Pro Cys Ser

1 5 10 15

Arg Asn Lys Tyr Pro Val Pro Asp Thr Met Gln Gln Thr Ile Gln Ile

20 25 30

Lys Asn Pro Glu Lys Leu His Pro Ala Thr Leu Phe His Asp Trp Asp

35 40 45

Leu Arg Arg Gly Phe Ile Thr Gln Ala Ala Ile Lys Arg Met Ser Glu

50 55 60

Asn Leu Gln Ile Asp Ser Ser Phe Glu Ser Asp Gly Thr Glu Ser Pro

65 70 75 80

Lys Lys Lys Lys Arg Cys Thr Lys Glu Ile Pro Thr Gln Asn Gln Lys

85 90 95

Gln Glu Glu Ile Gln Glu Cys Leu Leu Ser Leu Cys Glu Glu Pro Thr

100 105 110

Cys Gln Glu Glu Thr Glu Asp Leu Gln Leu Phe Ile Gln Gln Gln Gln

115 120 125

Gln Gln Gln Tyr Lys Leu Arg Lys Asn Leu Phe Lys Leu Leu Thr His

130 135 140

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

145 150 155 160

<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
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> 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

<211> 728

<212> PRT

<213> Alphatorquevirus sp.

<400> 931

Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg
1 5 10 15
Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro
20 25 30
Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr
35 40 45
Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg

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

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

180 185 190

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

245 250 255
Met Ser Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser
260 265 270
Ile Tyr Thr Asn Leu Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu
275 280 285
Asn Ile Ile Asn Asn Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn
290 295 300
Asn Lys Lys Trp Gln Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr

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

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

435 440 445

Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile Lys Asn Thr Gly Val

450 455 460

Ser Ser Leu Val Lys Asp Ala Arg Ile Cys Ile Arg Cys Pro Tyr Thr

465 470 475 480

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

485 490 495

Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val Leu Ala Pro Tyr Ile

500 505 510

Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn Ile Ala His Gln Lys

515 520 525

Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro Phe Met Pro Arg Asp

530 535 540

Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly Tyr Lys Met Asp Phe

545 550 555 560

Leu Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys

565 570 575

Gln Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu

580 585 590

Leu Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His

595 600 605

Lys Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg

610 615 620

Val Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro

625 630 635 640

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

645 650 655

Gln Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr

660 665 670

Pro Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu

675 680 685

Leu Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg

690 695 700

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

705 710 715 720

Val His Trp Asn Pro Glu Leu Thr

725

<210> 932

<211> 66

<212> PRT

<213> Alphatorquevirus sp.

<400> 932

Thr Ala Trp Trp Trp Gly Arg Trp Arg Arg Arg Trp Arg Arg Arg Arg

1 5 10 15

Pro Tyr Thr Thr Arg Leu Arg Arg Arg Arg Ala Arg Arg Ala Phe Pro

20 25 30

Arg Arg Arg Arg Arg Arg Phe Val Ser Arg Arg Trp Arg Arg Pro Tyr

35 40 45

Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg

50 55 60

His Lys

65

<210> 933

<211> 211

<212> PRT

<213> Alphatorquevirus sp.

<400> 933

Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg His Cys

1 5 10 15

Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly Ser Thr

20 25 30

Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg Gly Ala

35 40 45

Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu Ala Ile
50 55 60
Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser Asn His
65 70 75 80

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

Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu Met Asn
145 150 155 160
Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu Phe Gln
165 170 175
Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg Met Ser
180 185 190
Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser Ile Tyr
195 200 205

Thr Asn Leu
210

<210> 934
<211> 118
<212> PRT
<213> Alphatorquevirus sp.
<400> 934
Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu Asn Ile Ile Asn Asn
1 5 10 15
Ile Leu His Pro Gln Glu Ile Thr Gly Thr Asn Asn Lys Lys Trp Gln
20 25 30
Tyr Thr Tyr Thr Lys Leu Met Ala Pro Ile Tyr Tyr Ser Ala Asn Arg

35 40 45

Ala Ser Thr Tyr Asp Trp Glu Asn Tyr Ser Lys Glu Thr Asn Tyr Asn

50 55 60

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

65 70 75 80

Ile Arg Lys Glu Trp Gln Met Leu Tyr Pro Gln Gln Pro Thr Ala Leu

85 90 95

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

100 105 110

Tyr Leu Asn Pro Thr Arg

115

<210> 935

<211> 166

<212> PRT

<213> Alphatorquevirus sp.

<400> 935

Ile Asn Leu Asp Trp Met Thr Pro Tyr Thr His Val Arg Tyr Asn Pro

1 5 10 15

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

20 25 30

Glu Ala Asp Val Ser Tyr Asn Arg Thr Lys Ser Lys Cys Leu Leu Gln

35 40 45

Asp Met Pro Leu Phe Phe Met Cys Tyr Gly Tyr Ile Asp Trp Ala Ile

50 55 60

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

65 70 75 80

Arg Cys Pro Tyr Thr Glu Pro Gln Leu Val Gly Ser Thr Glu Asp Ile

85 90 95

Gly Phe Val Pro Ile Ser Glu Thr Phe Met Arg Gly Asp Met Pro Val

100 105 110

Leu Ala Pro Tyr Ile Pro Leu Ser Trp Phe Cys Lys Trp Tyr Pro Asn

115 120 125

Ile Ala His Gln Lys Glu Val Leu Glu Ser Ile Ile Ser Cys Ser Pro

130 135 140

Phe Met Pro Arg Asp Gln Asp Met Asn Gly Trp Asp Ile Thr Ile Gly

145 150 155 160

Tyr Lys Met Asp Phe Leu

165

<210> 936

<211> 167

<212> PRT

<213> Alphatorquevirus sp.

<400> 936

Trp Gly Gly Ser Pro Leu Pro Ser Gln Pro Ile Asp Asp Pro Cys Gln

1 5 10 15

Gln Gly Thr His Pro Ile Pro Asp Pro Asp Lys His Pro Arg Leu Leu

20 25 30

Gln Val Ser Asn Pro Lys Leu Leu Gly Pro Arg Thr Val Phe His Lys

35 40 45

Trp Asp Ile Arg Arg Gly Gln Phe Ser Lys Arg Ser Ile Lys Arg Val

50 55 60

Ser Glu Tyr Ser Ser Asp Asp Glu Ser Leu Ala Pro Gly Leu Pro Ser

65 70 75 80

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

85 90 95

Lys Glu Cys Tyr Ser Leu Leu Lys Ala Leu Glu Glu Glu Glu Thr Pro

100 105 110

Glu Glu Glu Glu Pro Ala Pro Gln Glu Lys Ala Gln Lys Glu Glu Leu

115 120 125

Leu His Gln Leu Gln Leu Gln Arg Arg His Gln Arg Val Leu Arg Arg

130 135 140

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

145 150 155 160

His Trp Asn Pro Glu Leu Thr

165

<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

<400> 950

000

<210> 951

<400> 951

000

<210> 952

<400> 952

000

<210> 953

<400> 953

000

<210> 954

<400> 954

000

<210> 955

<400> 955

000

<210> 956

<400> 956

000

<210> 957

<400> 957

000

<210> 958

<400> 958

000

<210> 959

<400> 959

000

<210> 960

<400> 960

000

<210> 961

<400> 961

000

<210> 962

<400> 962

000

<210> 963

<400> 963

000

<210> 964

<400> 964

000

<210> 965

<400> 965

000

<210> 966

<400> 966

000

<210> 967

<400> 967

000

<210> 968

<400> 968

000

<210> 969

<400> 969

000

<210> 970

<400> 970

000

<210> 971

<400> 971

000

<210> 972

<400> 972

000

<210> 973

<400> 973

000

<210> 974

<400> 974

000

<210> 975

<400> 975

000

<210> 976

<400> 976

000

<210

> 977

<400> 977

000

<210> 978

<400> 978

000

<210> 979

<400> 979

000

<210> 980

<400> 980

000

<210> 981

<400> 981

000

<210> 982

<400> 982

000

<210> 983

<400> 983

000

<210> 984

<400> 984

000

<210> 985

<400> 985

000

<210> 986

<400> 986

000

<210> 987

<400> 987

000

<210> 988

<400> 988

000

<210> 989

<400> 989

000

<210> 990

<400> 990

000

<210> 991

<400> 991

000

<210> 992

<400> 992

000

<210> 993

<400> 993

000

<210> 994

<400> 994

000

<210> 995

<400> 995

000

<210> 996

<400> 996

000

<210> 997

<400> 997

000

<210> 998

<400> 998

000

<210> 999

<400> 999

000

<210> 1000

<400> 1000

000

<210> 1001

<211> 2845

<212> DNA

<213> Betatorquevirus sp.

<400> 1001

ttattaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactaa 60

cttccttatt taacattact tccctttaa ccaatgaata ttcatacaac acatcacact 120

tcctgggagg agacataaaa ctatataact aactacacag acgaatggct gagtttatgc 180

cgctagacgg aggacgcaca gctactgctg cgacctgaac ttgggcgggt gccgaaggtg 240

agtgtaacca ccgtagtcaa ggggcaattc gggctagttc agtctagcgg aacgggcaag 300

attattaata caaacttatt tttacagatg agcaaacaac taaaaccaac tttatacaaa 360

gacaaatcat tggaattaca atggctaaac aacattttta gctctcacga cctgtgctgc 420

ggctgcaacg atccagtttt acatttactg attttaatta acaaaaccgg agaagcacct 480

aaaccagaag aagacattaa aaatataaaa tgcctcctta ctggcgccaa aaatactacc 540

gaagaagata tagacctttc tcctggagaa ctagaagaat tattcaaaga agaaaaagat 600

ggagataccg caaaccaaga aaaacatact ggagaagaaa actgcgggta agaaaacgtt 660

tttataaaag aaagttaaaa aaaattgtac ttaaacagtt tcaaccaaaa attattagaa 720

gatgtacaat atttggaaca atctgcctat ttcaaggctc tccagaaaga gccaacaata 780

attatattca aacaatctac tcctacgtac cagataaaga accaggagga gggggatgga 840

ctttaataac tgaaagctta agtagtttat gggaagactg ggaacattta aaaaatgtat 900

ggactcaaag taacgctggt ttaccacttg taagatacgg gggagtaaca ttatactttt 960

atcaatctgc ctatactgac tatattgctc aagttttcaa ctgttatcct atgacagaca 1020

caaaatacac acatgcagac tcagcaccaa acagaatgtt attaaaaaaa catgtaataa 1080

gagtacctag cagagaaaca cgcaaaaaaa gaaagccata caaaagagtt agagtaggac 1140

ctccttctca aatgcaaaac aaatggtact ttcaaagaga catatgtgaa ataccattaa 1200

taatgattgc agccacagcc gttgacttta gatatccctt ttgtgcaagc gactgtgcta 1260

gtaacaactt aactctaaca tgtttaaacc cactattgtt tcaaaaccaa gactttgacc 1320

acccatccga tacacaaggc tactttccaa aacctggagt atatctatac tcaacacaaa 1380

gaagtaacaa gccaagttct tcagactgta tatacttagg aaacacaaaa gacaatcaag 1440

aaggtaaatc tgcaagtagt ctaatgactc taaaaacaca aaaaataaca gattggggaa 1500

atccattttg gcattattat atagacggtt ctaaaaaaat attttcttac tttaaacccc 1560

catcacaatt agacagcagc gactttgaac acatgacaga attagcagaa ccaatgttta 1620

tacaagttag atacaaccca gaaagagaca caggacaagg aaacttaata tacgtaacag 1680

aaaactttag aggacaacac tgggaccctc catctagtga caacctaaaa ttagatggat 1740

ttcccttata tgacatgtgc tggggtttca tagactggat agaaaaagtt catgaaacag 1800

aaaacttact taccaactac tgcttctgta ttagaagcag cgctttcaat gaaaaaaaaa 1860

cagtttttat acctgtagat cattcatttt taacaggttt tagcccatat gaaactccag 1920

ttaaatcatc agaccaagct cactggcacc cacaaataag atttcaaaca aaatcaataa 1980

atgacatttg tttaacaggc cccggttgtg ctaggtcccc atatggcaat tacatgcagg 2040

caaaaatgag ttataaattt catgtaaaat ggggaggatg tccaaaaact tatgaaaaac 2100

catatgatcc ttgttcacag cccaattgga ctattcccca taacctcaat gaaacaatac 2160

aaatccagaa tccaaacaca tgcccacaaa cagaactcca agaatgggac tggcgacgtg 2220

atattgttac aaaaaaagct atcgaaagaa ttagacaaca cacggaacct catgaaactt 2280

tgcaaatctc tacaggttcc aaacacaacc caccagtaca cagacaaaca tcaccgtgga 2340

cggactcaga aacggactcg gaagaggaaa aagaccaaac acaagagatc cagatccagc 2400

tcaacaagct cagaaagcat caacagcatc tcaagcagca gctcaagcag tacctgaaac 2460

cccaaaatat agaatagttg caagcaacat aaaagttgaa ctttttcctа ctaaaaaacc 2520

ttttaaaaac agacgcttta ctccttctga aagagaaaca gaaagacaat gtgctaaagc 2580

tttttgtaga ccagaaagac atttctttta tgatcctcct ttttaccctt actgtgtacc 2640

tgaacctatt gtaaactttg ctttgggata taaaatttaa ggccaacaaa tttcacttag 2700

tggtgtctgt ttattaaagt ttaaccttaa taagcatact ccgcctccct acattaaggc 2760

gccaaaaggg ggctccgccc ccttaaaccc caaggggggct ccgccccctt aaaccccccaa 2820

gggggctccg ccccccttaca ccccc 2845

<210> 1002

<211> 107

<212> PRT

<213> Betatorquevirus sp.

<400> 1002

Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu
1 5 10 15

Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly
20 25 30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
35 40 45

Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu
50 55 60

Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly
65 70 75 80

Glu Leu Glu Glu Leu Phe Lys Glu Glu Lys Asp Gly Asp Thr Ala Asn
85 90 95

Gln Glu Lys His Thr Gly Glu Glu Asn Cys Gly
100 105

<210> 1003

<211> 218

<212> PRT

<213> Betatorquevirus sp.

<400> 1003

Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu
1 5 10 15

Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly
20 25 30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
35 40 45

Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu

50 55 60

Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly

65 70 75 80

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

85 90 95

Gln Glu Lys His Thr Gly Glu Glu Asn Cys Gly Pro Ile Gly Leu Phe

100 105 110

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

115 120 125

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

130 135 140

Lys Lys Leu Ser Lys Glu Leu Asp Asn Thr Arg Asn Leu Met Lys Leu

145 150 155 160

Cys Lys Ser Leu Gln Val Pro Asn Thr Thr His Gln Tyr Thr Asp Lys

165 170 175

His His Arg Gly Arg Thr Gln Lys Arg Thr Arg Lys Arg Lys Lys Thr

180 185 190

Lys His Lys Arg Ser Arg Ser Ser Ser Thr Ser Ser Glu Ser Ile Asn

195 200 205

Ser Ile Ser Ser Ser Ser Ser Ser Ser Thr

210 215

<210> 1004

<211> 234

<212> PRT

<213> Betatorquevirus sp.

<400> 1004

Met Ser Lys Gln Leu Lys Pro Thr Leu Tyr Lys Asp Lys Ser Leu Glu

1 5 10 15

Leu Gln Trp Leu Asn Asn Ile Phe Ser Ser His Asp Leu Cys Cys Gly

20 25 30

Cys Asn Asp Pro Val Leu His Leu Leu Ile Leu Ile Asn Lys Thr Gly
35 40 45
Glu Ala Pro Lys Pro Glu Glu Asp Ile Lys Asn Ile Lys Cys Leu Leu
50 55 60
Thr Gly Ala Lys Asn Thr Thr Glu Glu Asp Ile Asp Leu Ser Pro Gly
65 70 75 80
Glu Leu Glu Glu Leu Phe Lys Glu Glu Lys Asp Gly Asp Thr Ala Asn

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

145 150 155 160
Glu Thr Pro Lys Tyr Arg Ile Val Ala Ser Asn Ile Lys Val Glu Leu
165 170 175
Phe Pro Thr Lys Lys Pro Phe Lys Asn Arg Arg Phe Thr Pro Ser Glu
180 185 190
Arg Glu Thr Glu Arg Gln Cys Ala Lys Ala Phe Cys Arg Pro Glu Arg
195 200 205
His Phe Phe Tyr Asp Pro Pro Phe Tyr Pro Tyr Cys Val Pro Glu Pro

210 215 220
Ile Val Asn Phe Ala Leu Gly Tyr Lys Ile
225 230
<210> 1005
<211> 655
<212> PRT
<213> Betatorquevirus sp.
<400> 1005
Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg Pro

1 5 10 15

Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg

20 25 30

Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val Arg

35 40 45

Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln Phe

50 55 60

Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys Leu

65 70 75 80

Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr Ile

85 90 95

Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr Leu

100 105 110

Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu Lys

115 120 125

Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr Gly

130 135 140

Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile Ala

145 150 155 160

Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His Ala

165 170 175

Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg Val

180 185 190

Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val Arg

195 200 205

Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg Asp

210 215 220

Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp Phe

225 230 235 240

Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr Leu

245 250 255

Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His Pro

260 265 270

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

275 280 285

Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu Gly

290 295 300

Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met Thr

305 310 315 320

Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His Tyr

325 330 335

Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro Ser

340 345 350

Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu Pro

355 360 365

Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln Gly

370 375 380

Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp Pro

385 390 395 400

Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp Met

405 410 415

Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu Asn

420 425 430

Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn Glu

435 440 445

Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly Phe

450 455 460

Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp His

465 470 475 480

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

485 490 495

Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala Lys

500 505 510

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

515 520 525

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

530 535 540

Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro Gln

545 550 555 560

Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys Lys

565 570 575

Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu Gln

580 585 590

Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr Ser

595 600 605

Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln Thr

610 615 620

Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln His

625 630 635 640

Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu

645 650 655

<210> 1006

<211> 164

<212> PRT

<213> Betatorquevirus sp.

<400> 1006

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

1 5 10 15

Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg

20 25 30

Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Pro Asn

35 40 45

Trp Thr Ile Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro

50 55 60

Asn Thr Cys Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp
65 70 75 80

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

Lys His Gln Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro
145 150 155 160
Gln Asn Ile Glu

<210> 1007
<211> 99
<212> PRT
<213> Betatorquevirus sp.
<400> 1007
Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg Pro
1 5 10 15
Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp Arg
20 25 30
Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val Pro

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

<210> 1008

<211>

2912

<212> DNA

<213> Betatorquevirus sp.

<400> 1008

taataaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactta 60

gttcctcttt ttccacaact tcctctttta ctaatgaata ttcatgtaat taattaataa 120

tcaccgtaat tccggggagg agcctttaaa ctataaaact aactacacat tcgaatggct 180

gagtttatgc cgccagacgg agacgggatc acttcagtga ctccaggctg atcaagggcg 240

ggtgccgaag gtgagtgaaa ccaccgtagt caaggggcaa ttcgggctag atcagtctgg 300

cggaacgggc aagaaactta aaatgtactt tattttacag aaatgttcaa atctccaaca 360

tacttaacaa ctaaaggcaa aaacaatgcc ttaatcaact gcttcgttgg agaccacgat 420

cttctgtgca gctgtaacaa tcctgcctac cattgcctcc aaatacttgc aactacctta 480

gcacctcaac taaaacaaga agaaaaacaa caaataatac aatgccttgg tggtacagac 540

gccgtagcta caacccgtgg agacgaagaa attggtttag aagacctaga aaaactattt 600

acagaagata cagaagaaga cgccgctggg taagaagaaa acctttttac aaacgtaaaa 660

ttaagagact aaatatagta gaatggcaac ctaaatcaat tagaaaatgt agaataaaag 720

gaatgctatg cttgtttcaa acgacagaag acagactgtc atataacttt gatatgtatg 780

aagagtctat tataccagaa aaactgccgg gagggggggg atttagcatt aagaatataa 840

gcttatatgc cttataccaa gaacacatac atgcacacaa catatttaca cacacaaaca 900

cagacagacc actagcaaga tacacaggct gttctttaaa attctaccaa agcaaagaca 960

tagactacgt agtaacatat tctacatcac tcccactaag aagctcaatg ggaatgtaca 1020

actccatgca accatccata catctaatgc aacaaaacaa actaattgta ccaagcaaac 1080

aaacacaaaa aagaagaaaa ccatatatta aaaaacatat atcaccacca acacaaatga 1140

aatctcaatg gtactttcaa cataacattg caaacatacc gctactaatg ataagaacca 1200

cagcattaac attagataat tactatatag gaagcagaca attaagtaca aatgtcacta 1260

tacatacact taacacaaca tacatccaaa acagagactg gggagacaga aataaaactt 1320

actactgcca aacattagga acacaaagat acttcctata tggaacacat tcaactgcac 1380

aaaatattaa tgacataaag ctacaagaac taatacctttt aacaaacaca caagactatg 1440

tacaaggctt tgattggaca gaaaaagaca aacataacat aacaacctac aaagaattct 1500

taactaaagg agcaggaaat ccatttcacg cagaatggat aacagcacaa aacccagtaa 1560

tacacacagc aaacagtcct acacaaatag aacaaatata caccgcttca acaacaacat 1620

tccaaaacaa aaaactaaca gacctaccaa cgccaggata tatatttata actccaacag 1680

taagcttaag atacaaccca tacaaagacc tagcagaaag aaacaaatgc tactttgtaa 1740

gaagcaaaat aaatgcacac gggtgggacc cagaacaaca ccaagaatta ataaacagtg 1800

acctaccaca atggttacta ttatttggct acccagacta cataaaaaga acacaaaact 1860

ttgcattagt agacacaaat tacatactag tagaccactg cccatacaca aatccagaaa 1920

aaacaccatt tatacttta agcacatcat ttatagaagg tagaagccca tacagtcctt 1980

cagacacaca tgaaccagat gaagaagacc aaaacaggtg gtacccatgc taccaatatc 2040

aacaagaatc aataaattca atatgtctta gcggtccagg cacaccaaaa ataccaaaag 2100

gaataacagc agaagcaaaa gtaaaatatt cctttaattt taagtggggt ggtgacctac 2160

caccaatgtc tacaattaca aacccgacag accagccaac atatgttgtt cccaataact 2220

tcaatgaaac aacttcgtta cagaatccaa ccaccagacc agagcacttc ttgtactcct 2280

ttgacgaaag gaggggacaa cttacagaaa aagctacaaa acgcttgctt aaagactggg 2340

aaactaaaga aacttcttta ttgtctacag aatacagatt cgcggagcca acacaaacac 2400

aagccccaca agaggacccg tcctcggaag aagaagaaga gagcaacctc ttcgagcgac 2460

tcctccgaca gcgaaccaag cagctccagc tcaagcgcag aataatacaa acattgaaag 2520

acctacaaaa attagaataa ctaacagcaa aaacaccgtt tacctatttc cacctgaaca 2580

aaagaacaga agactaacac catgggaaat acaagaagac aaagaaatag ccaatttatt 2640

tggcagacca catagatact ttttaaaaga cattcctttc tattgggata tacccccaga 2700

gcctaaagta aactttgatt taaattttca ataaagaaat aaagggcaag gccccattaa 2760

ctcaaagtcg gtgtctacct ctttaagttt aactttacta aacggactcc gcctccctaa 2820

atttgggcgc caaaaggggg ctccgccccc ttaaacccca gggggctccg ccccctaaaa 2880

cccccaaggg ggctacgccc ccttacaccc cc 2912

<210> 1009

<211> 96

<212> PRT

<213> Betatorquevirus sp.

<400> 1009

Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala

1 5 10 15

Leu Ile Asn Cys Phe Val Gly Asp His Asp Leu Leu Cys Ser Cys Asn
20 25 30
Asn Pro Ala Tyr His Cys Leu Gln Ile Leu Ala Thr Thr Leu Ala Pro
35 40 45

Gln Leu Lys Gln Glu Glu Lys Gln Gln Ile Ile Gln Cys Leu Gly Gly
50 55 60
Thr Asp Ala Val Ala Thr Thr Arg Gly Asp Glu Glu Ile Gly Leu Glu
65 70 75 80
Asp Leu Glu Lys Leu Phe Thr Glu Asp Thr Glu Glu Asp Ala Ala Gly
85 90 95

<210> 1010
<211> 198
<212> PRT
<213> Betatorquevirus sp.
<400> 1010

Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala

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

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

Gly Asp Asn Leu Gln Lys Lys Leu Gln Asn Ala Cys Leu Lys Thr Gly

130 135 140

Lys Leu Lys Lys Leu Leu Tyr Cys Leu Gln Asn Thr Asp Ser Arg Ser
145 150 155 160
Gln His Lys His Lys Pro His Lys Arg Thr Arg Pro Arg Lys Lys Lys
165 170 175
Lys Arg Ala Thr Ser Ser Ser Asp Ser Ser Asp Ser Glu Pro Ser Ser
180 185 190
Ser Ser Ser Ser Ala Glu
195

<210> 1011
<211> 216
<212> PRT
<213> Betatorquevirus sp.
<400> 1011
Met Phe Lys Ser Pro Thr Tyr Leu Thr Thr Lys Gly Lys Asn Asn Ala
1 5 10 15
Leu Ile Asn Cys Phe Val Gly Asp His Asp Leu Leu Cys Ser Cys Asn
20 25 30
Asn Pro Ala Tyr His Cys Leu Gln Ile Leu Ala Thr Thr Leu Ala Pro
35 40 45
Gln Leu Lys Gln Glu Glu Lys Gln Gln Ile Ile Gln Cys Leu Gly Gly

50 55 60

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

115 120 125

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

180 185 190
Phe Leu Lys Asp Ile Pro Phe Tyr Trp Asp Ile Pro Pro Glu Pro Lys
195 200 205
Val Asn Phe Asp Leu Asn Phe Gln
210 215
<210> 1012
<211> 672
<212> PRT
<213> Betatorquevirus sp.
<400> 1012
Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg
1 5 10 15
Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg

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

85 90 95
Gly Gly Gly Gly Phe Ser Ile Lys Asn Ile Ser Leu Tyr Ala Leu Tyr
100 105 110
Gln Glu His Ile His Ala His Asn Ile Phe Thr His Thr Asn Thr Asp

115 120 125

Arg Pro Leu Ala Arg Tyr Thr Gly Cys Ser Leu Lys Phe Tyr Gln Ser

130 135 140

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

145 150 155 160

Ser Ser Met Gly Met Tyr Asn Ser Met Gln Pro Ser Ile His Leu Met

165 170 175

Gln Gln Asn Lys Leu Ile Val Pro Ser Lys Gln Thr Gln Lys Arg Arg

180 185 190

Lys Pro Tyr Ile Lys Lys His Ile Ser Pro Pro Thr Gln Met Lys Ser

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ser Thr Asn Val Thr Ile His Thr Leu Asn Thr Thr Tyr Ile Gln

245 250 255

Asn Arg Asp Trp Gly Asp Arg Asn Lys Thr Tyr Tyr Cys Gln Thr Leu

260 265 270

Gly Thr Gln Arg Tyr Phe Leu Tyr Gly Thr His Ser Thr Ala Gln Asn

275 280 285

Ile Asn Asp Ile Lys Leu Gln Glu Leu Ile Pro Leu Thr Asn Thr Gln

290 295 300

Asp Tyr Val Gln Gly Phe Asp Trp Thr Glu Lys Asp Lys His Asn Ile

305 310 315 320

Thr Thr Tyr Lys Glu Phe Leu Thr Lys Gly Ala Gly Asn Pro Phe His

325 330 335

Ala Glu Trp Ile Thr Ala Gln Asn Pro Val Ile His Thr Ala Asn Ser

340 345 350

Pro Thr Gln Ile Glu Gln Ile Tyr Thr Ala Ser Thr Thr Thr Phe Gln

355 360 365

Asn Lys Lys Leu Thr Asp Leu Pro Thr Pro Gly Tyr Ile Phe Ile Thr
370 375 380
Pro Thr Val Ser Leu Arg Tyr Asn Pro Tyr Lys Asp Leu Ala Glu Arg
385 390 395 400
Asn Lys Cys Tyr Phe Val Arg Ser Lys Ile Asn Ala His Gly Trp Asp

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

465 470 475 480
Arg Ser Pro Tyr Ser Pro Ser Asp Thr His Glu Pro Asp Glu Glu Asp
485 490 495
Gln Asn Arg Trp Tyr Pro Cys Tyr Gln Tyr Gln Gln Glu Ser Ile Asn
500 505 510
Ser Ile Cys Leu Ser Gly Pro Gly Thr Pro Lys Ile Pro Lys Gly Ile
515 520 525
Thr Ala Glu Ala Lys Val Lys Tyr Ser Phe Asn Phe Lys Trp Gly Gly

530 535 540
Asp Leu Pro Pro Met Ser Thr Ile Thr Asn Pro Thr Asp Gln Pro Thr
545 550 555 560
Tyr Val Val Pro Asn Asn Phe Asn Glu Thr Thr Ser Leu Gln Asn Pro
565 570 575
Thr Thr Arg Pro Glu His Phe Leu Tyr Ser Phe Asp Glu Arg Arg Gly
580 585 590
Gln Leu Thr Glu Lys Ala Thr Lys Arg Leu Leu Lys Asp Trp Glu Thr

595 600 605
Lys Glu Thr Ser Leu Leu Ser Thr Glu Tyr Arg Phe Ala Glu Pro Thr

610 615 620

Gln Thr Gln Ala Pro Gln Glu Asp Pro Ser Ser Glu Glu Glu Glu Glu

625 630 635 640

Ser Asn Leu Phe Glu Arg Leu Leu Arg Gln Arg Thr Lys Gln Leu Gln

645 650 655

Leu Lys Arg Arg Ile Ile Gln Thr Leu Lys Asp Leu Gln Lys Leu Glu

660 665 670

<210> 1013

<211> 80

<212> PRT

<213> Betatorquevirus sp.

<400> 1013

Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg

1 5 10 15

Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg

20 25 30

Arg Arg Arg Trp Asn Thr Asp Ser Arg Ser Gln His Lys His Lys Pro

35 40 45

His Lys Arg Thr Arg Pro Arg Lys Lys Lys Lys Arg Ala Thr Ser Ser

50 55 60

Ser Asp Ser Ser Asp Ser Glu Pro Ser Ser Ser Ser Ser Ser Ala Glu

65 70 75 80

<210> 1014

<211> 3853

<212> DNA

<213> Alphatorquevirus sp.

<400> 1014

ggcttagtgc gtcaccaccc acgtgacccg cctccgccaa ttaacaggta cttcgtacac 60

ttcctgggcg ggcttataag actaatataa gtagctgcac ttccgaatgg ctgagttttc 120

cacgcccgtc cgcagcggtg aagccacgga gggagctcag cgcgtcccga gggcgggtgc 180

cggaggtgag tttacacacc gcagtcaagg ggcaattcgg gctcgggact ggccgggctt 240

tgggcaaggc tcttaaaaaa gctatgttta ttggcaggca ctaccgaaag aaaagggcgc 300

tgctactgct atctgtgcat tctacaaaga caaaagggaa acttctaata gctatgtgga 360

ctcccccacg caatgatcaa caatacctta actggcaatg gtacacttct gtacttagct 420

cccactctgc tatgtgcggg tgttccgacg ctatcgctca tcttaatcat cttgctaatc 480

tgcttcgtgc cccgcaaaat ccgcccccgc ctgataatcc aagaccccta cccgtgcgag 540

cactgcctgc tccccggct gcccacgagg cagccggtga tcgagcacca tggcctatgg 600

gtggtggagg agacgccgga ggcgctggcg caggtggaga cgccgaccat ggaggcgccg 660

ctggaggacc cgcagacgca gacctgctag acgccgtggc cgccgcagaa acgtaaggag 720

acggcgcaga gggaggtgga gaaggaggta caggaggtgg aaaagaaagg gcagacgtag 780

aagaaaagca aaaataataa taagacagtg gcagccaaac tacagaagaa gatgtaatat 840

agtgggctac ctccctatac ttatctgtgg tggaaatact gtttctagaa actatgccac 900

acactcagac gatactaact atccaggacc ctttgggggа ggcatgacca cagacaaatt 960

cagccttaga atactatatg atgaatacaa aagatttatg aactactgga cagcctcaaa 1020

tgaggaccta gatctctgta gatatctagg atgcactttt tacttcttta gacaccctga 1080

agtagacttt attataaaaa taaacaccat gcccccattc ttagatacaa ccataacagc 1140

acctagcata cacccaggcc tcatggccct agacaaaaga gccagatgga ttccttctct 1200

taaaaataga ccaggtaaaa aacactatat aaaaattaga gtaggggctc ctaaaatgtt 1260

cacagataaa tggtaccctc aaacagacct ctgtgacatg acactgctaa ctatctatgc 1320

aaccgcagcg gatatgcaat atccgttcgg ctcaccacta actgacactg tggttgttaa 1380

ctcccaagtt ctgcaatcca tgtatgatga aacaattagc atattacctg atgaaaaaac 1440

taaaagaaat agccttctta cttctataag aagctacata ccttttttata atactacaca 1500

aacaatagct caattaaaac catttgtaga tgcaggagga cacacaacag gctcaacaac 1560

aactacatgg ggacaactat taaacacaac taaatttacc actaccacaa caaccacata 1620

cacataccct ggcaccacaa atacagcagt aacatttata acagccaatg atacctggta 1680

caggggaaca gcatataaag ataacattaa agatgtacca caaaaagcag cacaattata 1740

ctttcaaaca acacaaaaac tactaggaaa cacattccat ggctcagatg aaacacttga 1800

ataccatgca ggcctataca gctctatctg gctatcacca ggtagatcct actttgaaac 1860

accaggtgca tacacagaca ttaaatataa ccctttaca gacagaggag aaggcaacat 1920

gctgtggata gactggctaa gtaaaaaaaa catgaaatat gacaaagtgc aaagtaagtg 1980

cctagtagca gacctaccac tgtgggcagc agcatatggt tatgtagaat tctgctctaa 2040

aagcacagga gacacaaaca tacacatgaa tgccagacta ctaataagaa gtccttttac 2100

agacccccag ctaatagtac acacagaccc cactaaaggc tttgtaccct attctttaaa 2160

ctttggaaat ggtaaaatgc caggaggtag cagcaatgtt cccataagaa tgagagctaa 2220

gtggtacccc actttatccc accaacaaga agttctagag gccttagcac agtcaggacc 2280

ctttgcttat cactcagaca ttaaaaaagt atctctaggc ataaaatacc gttttaagtg 2340

gatctggggt ggaaaccccg ttcgccaaca ggttgttaga aatccctgca aggaacccca 2400

ctcctcgggc aatagagtcc ctagaagcat acaaatcgtt gacccgagat acaactcacc 2460

ggaacttacc atccatgcct gggacttcag acgtggcttc tttggcccga aagctattca 2520

aagaatgcaa caacaaccaa ctgctactga atttttttca gcaggccgca agagacccag 2580

aagggacaca gaagtgtatc agtccgacca agaaaaggag caaaaagaaa gctcgctttt 2640

ccccccagtc aagctcctcc gaagagtccc cccgtgggag gactcggaac aggagcaaag 2700

cgggtcgcaa agctcagagg aagagacggc gaccctctcc cagcagctca aacagcagct 2760

gcagcagcag cgagtcttgg gagtcaaact cagactcctg ttcaaccaag tccaaaaaat 2820

ccaacaaaat caagatatca accctacctt gttaccaagg gggggggatc tagtatcctt 2880

ctttcaggct gtaccataaa tatgtttcca gaccctaaac cttactgccc ctccagcaat 2940

gactggaaag aagagtatga ggcctgtaaa tattgggata gacctcccag acacaacctt 3000

agagaccccc ccttttaccc ctgggcccct aaaaacaatc cttgcaatgt aagctttaaa 3060

cttggcttca aataaactag gccgtgggag tttcacttgt cggtgtctac ctctataagt 3120

cactaagcac tccgagcgca gcgaggagtg cgacccttcc ccctggtgca acgccctcgg 3180

cggccgcgcg ctacgccttc ggctgcgcgc ggcacctcgg acccccgctc gtgctgacac 3240

gcttgcgcgt gtcagaccac ttcgggctcg cgggggtcgg gaaatttgct aaacagactc 3300

cgagttgcca ttggacactg tagctatgaa tcagtaacga aagtgagtgg ggccagactt 3360

cgccataagg cctttatctt cttgccattt gtcagtattg ggggtcgcca taaactttgg 3420

gctccatttt aggccttccg gactacaaaa atcgccatat ttgtgacgtc agagccgcca 3480

ttttaagtca gctctgggga ggcgtgactt ccagttcaaa ggtcatcctc accataactg 3540

gcacaaaatg gccgccaact tcttccgggt caaaggtcac tgctacgtca taggtgacgt 3600

ggggggggac ctacttaaac acggaagtag gccccgacac gtcactgtca cgtgacagta 3660

cgtcacagcc gccattttgt tttacaaaat agccgacttc cttcctcttt tttaaaaaaa 3720

ggcgccaaaa aaccgtcggc gggggggccg cgcgctgcgc gcgcggcccc cgggggaggc 3780

acagcctccc ccccccgcgc gcatgcgcgc gggtcccccc ccctccgggg ggctccgccc 3840

cccggccccc ccc 3853

<210> 1015

<211> 120

<212> PRT

<213> Alphatorquevirus sp.

<400> 1015

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

Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65 70 75 80
Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
85 90 95
Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
100 105 110
Asp Ala Val Ala Ala Ala Glu Thr
115 120

<210> 1016

<211> 286

<212> PRT

<213> Alphatorquevirus sp.

<400> 1016

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

50 55 60

Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65 70 75 80
Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
85 90 95
Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
100 105 110
Asp Ala Val Ala Ala Ala Glu Thr Leu Leu Glu Ile Pro Ala Arg Asn
115 120 125

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

Gln Lys Cys Ile Ser Pro Thr Lys Lys Arg Ser Lys Lys Lys Ala Arg
195 200 205
Phe Ser Pro Gln Ser Ser Ser Ser Glu Glu Ser Pro Arg Gly Arg Thr
210 215 220
Arg Asn Arg Ser Lys Ala Gly Arg Lys Ala Gln Arg Lys Arg Arg Arg
225 230 235 240
Pro Ser Pro Ser Ser Ser Asn Ser Ser Cys Ser Ser Ser Glu Ser Trp
245 250 255

Glu Ser Asn Ser Asp Ser Cys Ser Thr Lys Ser Lys Lys Ser Asn Lys
260 265 270
Ile Lys Ile Ser Thr Leu Pro Cys Tyr Gln Gly Gly Gly Ile
275 280 285
<210> 1017
<211> 289

<212> PRT

<213> Alphatorquevirus sp.

<400> 1017

Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1 5 10 15

Tyr Thr Ser Val Leu Ser Ser His Ser Ala Met Cys Gly Cys Ser Asp
20 25 30

Ala Ile Ala His Leu Asn His Leu Ala Asn Leu Leu Arg Ala Pro Gln
35 40 45

Asn Pro Pro Pro Pro Asp Asn Pro Arg Pro Leu Pro Val Arg Ala Leu
50 55 60

Pro Ala Pro Pro Ala Ala His Glu Ala Ala Gly Asp Arg Ala Pro Trp
65 70 75 80

Pro Met Gly Gly Gly Gly Asp Ala Gly Gly Ala Gly Ala Gly Gly Asp
85 90 95

Ala Asp His Gly Gly Ala Ala Gly Gly Pro Ala Asp Ala Asp Leu Leu
100 105 110

Asp Ala Val Ala Ala Ala Glu Thr Pro Gln Glu Thr Gln Lys Gly His
115 120 125

Arg Ser Val Ser Val Arg Pro Arg Lys Gly Ala Lys Arg Lys Leu Ala
130 135 140

Phe Pro Pro Ser Gln Ala Pro Pro Lys Ser Pro Pro Val Gly Gly Leu
145 150 155 160

Gly Thr Gly Ala Lys Arg Val Ala Lys Leu Arg Gly Arg Asp Gly Asp
165 170 175

Pro Leu Pro Ala Ala Gln Thr Ala Ala Ala Ala Ala Ala Ser Leu Gly
180 185 190

Ser Gln Thr Gln Thr Pro Val Gln Pro Ser Pro Lys Asn Pro Thr Lys
195 200 205

Ser Arg Tyr Gln Pro Tyr Leu Val Thr Lys Gly Gly Gly Ser Ser Ile
210 215 220

Leu Leu Ser Gly Cys Thr Ile Asn Met Phe Pro Asp Pro Lys Pro Tyr
225 230 235 240
Cys Pro Ser Ser Asn Asp Trp Lys Glu Glu Tyr Glu Ala Cys Lys Tyr
245 250 255
Trp Asp Arg Pro Pro Arg His Asn Leu Arg Asp Pro Pro Phe Tyr Pro
260 265 270
Trp Ala Pro Lys Asn Asn Pro Cys Asn Val Ser Phe Lys Leu Gly Phe

275 280 285
Lys

<210> 1018
<211> 185
<212> PRT
<213> Alphatorquevirus sp.
<400> 1018
Met Trp Thr Pro Pro Arg Asn Asp Gln Gln Tyr Leu Asn Trp Gln Trp
1 5 10 15
Pro Gln Glu Thr Gln Lys Gly His Arg Ser Val Ser Val Arg Pro Arg
20 25 30
Lys Gly Ala Lys Arg Lys Leu Ala Phe Pro Pro Ser Gln Ala Pro Pro
35 40 45

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

Thr Lys Gly Gly Gly Ser Ser Ile Leu Leu Ser Gly Cys Thr Ile Asn
115 120 125
Met Phe Pro Asp Pro Lys Pro Tyr Cys Pro Ser Ser Asn Asp Trp Lys

130 135 140

Glu Glu Tyr Glu Ala Cys Lys Tyr Trp Asp Arg Pro Pro Arg His Asn

145 150 155 160

Leu Arg Asp Pro Pro Phe Tyr Pro Trp Ala Pro Lys Asn Asn Pro Cys

165 170 175

Asn Val Ser Phe Lys Leu Gly Phe Lys

180 185

<210> 1019

<211> 105

<212> PRT

<213> Alphatorquevirus sp.

<400> 1019

Met Ile Asn Asn Thr Leu Thr Gly Asn Gly Thr Leu Leu Tyr Leu Ala

1 5 10 15

Pro Thr Leu Leu Cys Ala Gly Val Pro Thr Leu Ser Leu Ile Leu Ile

20 25 30

Ile Leu Leu Ile Cys Phe Val Pro Arg Lys Ile Arg Pro Arg Leu Ile

35 40 45

Ile Gln Asp Pro Tyr Pro Cys Glu His Cys Leu Leu Pro Arg Leu Pro

50 55 60

Thr Arg Gln Pro Val Ile Glu His His Gly Leu Trp Val Val Glu Glu

65 70 75 80

Thr Pro Glu Ala Leu Ala Gln Val Glu Thr Pro Thr Met Glu Ala Pro

85 90 95

Leu Glu Asp Pro Gln Thr Gln Thr Cys

100 105

<210> 1020

<211> 769

<212> PRT

<

213> Alphatorquevirus sp.

<400> 1020

Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp

1 5 10 15

Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro

20 25 30

Ala Arg Arg Arg Gly Arg Arg Arg Asn Val Arg Arg Arg Arg Arg Gly

35 40 45

Arg Trp Arg Arg Arg Tyr Arg Arg Trp Lys Arg Lys Gly Arg Arg Arg

50 55 60

Arg Lys Ala Lys Ile Ile Ile Arg Gln Trp Gln Pro Asn Tyr Arg Arg

65 70 75 80

Arg Cys Asn Ile Val Gly Tyr Leu Pro Ile Leu Ile Cys Gly Gly Asn

85 90 95

Thr Val Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Asn Tyr Pro

100 105 110

Gly Pro Phe Gly Gly Gly Met Thr Thr Asp Lys Phe Ser Leu Arg Ile

115 120 125

Leu Tyr Asp Glu Tyr Lys Arg Phe Met Asn Tyr Trp Thr Ala Ser Asn

130 135 140

Glu Asp Leu Asp Leu Cys Arg Tyr Leu Gly Cys Thr Phe Tyr Phe Phe

145 150 155 160

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

165 170 175

Phe Leu Asp Thr Thr Ile Thr Ala Pro Ser Ile His Pro Gly Leu Met

180 185 190

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

195 200 205

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

210 215 220

Thr Asp Lys Trp Tyr Pro Gln Thr Asp Leu Cys Asp Met Thr Leu Leu

225 230 235 240

Thr Ile Tyr Ala Thr Ala Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro

245 250 255

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

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

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

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

500 505 510

Asp Pro Thr Lys Gly Phe Val Pro Tyr Ser Leu Asn Phe Gly Asn Gly
515 520 525

Lys Met Pro Gly Gly Ser Ser Asn Val Pro Ile Arg Met Arg Ala Lys
530 535 540

Trp Tyr Pro Thr Leu Ser His Gln Gln Glu Val Leu Glu Ala Leu Ala
545 550 555 560

Gln Ser Gly Pro Phe Ala Tyr His Ser Asp Ile Lys Lys Val Ser Leu
565 570 575

Gly Ile Lys Tyr Arg Phe Lys Trp Ile Trp Gly Gly Asn Pro Val Arg
580 585 590

Gln Gln Val Val Arg Asn Pro Cys Lys Glu Pro His Ser Ser Gly Asn
595 600 605

Arg Val Pro Arg Ser Ile Gln Ile Val Asp Pro Arg Tyr Asn Ser Pro
610 615 620

Glu Leu Thr Ile His Ala Trp Asp Phe Arg Arg Gly Phe Phe Gly Pro
625 630 635 640

Lys Ala Ile Gln Arg Met Gln Gln Gln Pro Thr Ala Thr Glu Phe Phe
645 650 655

Ser Ala Gly Arg Lys Arg Pro Arg Arg Asp Thr Glu Val Tyr Gln Ser
660 665 670

Asp Gln Glu Lys Glu Gln Lys Glu Ser Ser Leu Phe Pro Pro Val Lys
675 680 685

Leu Leu Arg Arg Val Pro Pro Trp Glu Asp Ser Glu Gln Glu Gln Ser
690 695 700

Gly Ser Gln Ser Ser Glu Glu Glu Thr Ala Thr Leu Ser Gln Gln Leu
705 710 715 720

Lys Gln Gln Leu Gln Gln Gln Arg Val Leu Gly Val Lys Leu Arg Leu
725 730 735

Leu Phe Asn Gln Val Gln Lys Ile Gln Gln Asn Gln Asp Ile Asn Pro
740 745 750

Thr Leu Leu Pro Arg Gly Gly Asp Leu Val Ser Phe Phe Gln Ala Val
755 760 765

Pro

<210> 1021
<211> 216
<212> PRT
<213> Alphatorquevirus sp.
<400> 1021
Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp
1 5 10 15
Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro
20 25 30
Ala Arg Arg Arg Gly Arg Arg Arg Asn Val Val Arg Asn Pro Cys Lys
35 40 45
Glu Pro His Ser Ser Gly Asn Arg Val Pro Arg Ser Ile Gln Ile Val

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

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

Leu Gly Val Lys Leu Arg Leu Leu Phe Asn Gln Val Gln Lys Ile Gln

180 185 190

Gln Asn Gln Asp Ile Asn Pro Thr Leu Leu Pro Arg Gly Gly Asp Leu

195 200 205

Val Ser Phe Phe Gln Ala Val Pro

210 215

<210> 1022

<211> 143

<212> PRT

<213> Alphatorquevirus sp.

<400> 1022

Met Ala Tyr Gly Trp Trp Arg Arg Arg Arg Arg Arg Trp Arg Arg Trp

1 5 10 15

Arg Arg Arg Pro Trp Arg Arg Arg Trp Arg Thr Arg Arg Arg Arg Pro

20 25 30

Ala Arg Arg Arg Gly Arg Arg Arg Asn Ala Ala Arg Asp Pro Glu Gly

35 40 45

Thr Gln Lys Cys Ile Ser Pro Thr Lys Lys Arg Ser Lys Lys Lys Ala

50 55 60

Arg Phe Ser Pro Gln Ser Ser Ser Ser Glu Glu Ser Pro Arg Gly Arg

65 70 75 80

Thr Arg Asn Arg Ser Lys Ala Gly Arg Lys Ala Gln Arg Lys Arg Arg

85 90 95

Arg Pro Ser Pro Ser Ser Ser Asn Ser Ser Cys Ser Ser Ser Glu Ser

100 105 110

Trp Glu Ser Asn Ser Asp Ser Cys Ser Thr Lys Ser Lys Lys Ser Asn

115 120 125

Lys Ile Lys Ile Ser Thr Leu Pro Cys Tyr Gln Gly Gly Gly Ile

130 135 140

<210> 1023

<400> 1023

000

<210> 1024

<400> 1024

000

<210> 1025

<400> 1025

000

<210> 1026

<400> 1026

000

<210> 1027

<400> 1027

000

<210> 1028

<400> 1028

000

<210> 1029

<400> 1029

000

<210> 1030

<211> 1932

<212> DNA

<213> Adeno-associated virus

<400> 1030

atgccggggt tttacgagat tgtgattaag gtccccagcg accttgacga gcatctgccc 60

ggcatttctg acagctttgt gaactgggtg gccgagaagg aatgggagtt gccgccagat 120

tctgacatgg atctgaatct gattgagcag gcacccctga ccgtggccga gaagctgcag 180

cgcgactttc tgacggaatg gcgccgtgtg agtaaggccc cggaggccct tttctttgtg 240

caatttgaga agggagagag ctacttccac atgcacgtgc tcgtggaaac caccggggtg 300

aaatccatgg ttttgggacg tttcctgagt cagattcgcg aaaaactgat tcagagaatt 360

taccgcggga tcgagccgac tttgccaaac tggttcgcgg tcacaaagac cagaaatggc 420

gccggaggcg ggaacaaggt ggtggatgag tgctacatcc ccaattactt gctccccaaa 480

acccagcctg agctccagtg ggcgtggact aatatggaac agtatttaag cgcctgtttg 540

aatctcacgg agcgtaaacg gttggtggcg cagcatctga cgcacgtgtc gcagacgcag 600

gagcagaaca aagagaatca gaatcccaat tctgatgcgc cggtgatcag atcaaaaact 660

tcagccaggt acatggagct ggtcgggtgg ctcgtggaca aggggattac ctcggagaag 720

cagtggatcc aggaggacca ggcctcatac atctccttca atgcggcctc caactcgcgg 780

tcccaaatca aggctgcctt ggacaatgcg ggaaagatta tgagcctgac taaaaccgcc 840

cccgactacc tggtgggcca gcagcccgtg gaggacattt ccagcaatcg gatttataaa 900

attttggaac taaacgggta cgatccccaa tatgcggctt ccgtctttct gggatgggcc 960

acgaaaaagt tcggcaagag gaacaccatc tggctgtttg ggcctgcaac taccgggaag 1020

accaacatcg cggaggccat agcccacact gtgcccttct acgggtgcgt aaactggacc 1080

aatgagaact ttcccttcaa cgactgtgtc gacaagatgg tgatctggtg ggaggagggg 1140

aagatgaccg ccaaggtcgt ggagtcggcc aaagccattc tcggaggaag caaggtgcgc 1200

gtggaccaga aatgcaagtc ctcggcccag atagacccga ctcccgtgat cgtcacctcc 1260

aacaccaaca tgtgcgccgt gattgacggg aactcaacga ccttcgaaca ccagcagccg 1320

ttgcaagacc ggatgttcaa atttgaactc acccgccgtc tggatcatga ctttgggaag 1380

gtcaccaagc aggaagtcaa agacttttc cggtgggcaa aggatcacgt ggttgaggtg 1440

gagcatgaat tctacgtcaa aaagggtgga gccaagaaaa gacccgcccc cagtgacgca 1500

gatataagtg agcccaaacg ggtgcgcgag tcagttgcgc agccatcgac gtcagacgcg 1560

gaagcttcga tcaactacgc agacaggtac caaaacaaat gttctcgtca cgtgggcatg 1620

aatctgatgc tgtttccctg cagacaatgc gagagaatga atcagaattc aaatatctgc 1680

ttcactcacg gacagaaaga ctgtttagag tgctttcccg tgtcagaatc tcaacccgtt 1740

tctgtcgtca aaaaggcgta tcagaaactg tgctacattc atcatatcat gggaaaggtg 1800

ccagacgctt gcactgcctg cgatctggtc aatgtggatt tggatgactg catctttgaa 1860

caataaatga tttaaatcag gtatggctgc cgatggttat cttccagatt ggctcgagga 1920

cactctctct ga 1932

<210> 1031

<211> 621

<212> PRT

<213> Adeno-associated virus

<400> 1031

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

1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu

20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile

35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu

50 55 60

Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val

65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu

85 90 95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile

100 105 110

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

115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly

130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys

145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu

165 170 175

Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His

180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn

195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr

210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys

225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

245 250 255

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

260 265 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
275 280 285
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
290 295 300

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

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

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

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
500 505 510
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp

515 520 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu

530 535 540

Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys

545 550 555 560

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu

565 570 575

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr

580 585 590

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp

595 600 605

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln

610 615 620

<210>

1032

<211> 536

<212> PRT

<213> Adeno-associated virus

<400> 1032

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

1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu

20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile

35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu

50 55 60

Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val

65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu

85 90 95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile

100 105 110

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

115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly

130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys

145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu

165 170 175

Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His

180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn

195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr

210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys

225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

245 250 255

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

260 265 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln

275 280 285

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

290 295 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala

305 310 315 320

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

325 330 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro

340 345 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

355 360 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala

370 375 380

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

385 390 395 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val

405 410 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser

420 425 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe

435 440 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln

450 455 460

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

465 470 475 480

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala

485 490 495

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val

500 505 510

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp

515 520 525

Arg Leu Ala Arg Gly His Ser Leu

530 535

<210> 1033

<211> 397

<212> PRT

<213> Adeno-associated virus

<400> 1033

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys

1 5 10 15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

20 25 30

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

35 40 45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln

50 55 60

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

65 70 75 80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala

85 90 95

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

100 105 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro

115 120 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

130 135 140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala

145 150 155 160

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

165 170 175

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val

180 185 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser

195 200 205

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe

210 215 220

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln

225 230 235 240

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

245 250 255

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala

260 265 270

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val

275 280 285

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp

290 295 300

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu

305 310 315 320

Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys

325 330 335

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu

340 345 350

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr

355 360 365

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp

370 375 380

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln

385 390 395

<210> 1034

<211> 312

<212> PRT

<213> Adeno-associated virus

<400

> 1034

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys

1 5 10 15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

20 25 30

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

35 40 45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln

50 55 60

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

65 70 75 80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
85 90 95
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
100 105 110
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
115 120 125

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

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

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
260 265 270
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
275 280 285
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
290 295 300
Arg Leu Ala Arg Gly His Ser Leu
305 310
<210> 1035
<211> 1941

<212> DNA

<213> Adeno-associated virus

<400> 1035

atgccggggt tctacgagat tgtcctgaag gtcccgagtg acctggacga gcgcctgccg 60

ggcatttcta actcgtttgt taactgggtg gccgagaagg aatgggacgt gccgccggat 120

tctgacatgg atccgaatct gattgagcag gcacccctga ccgtggccga aaagcttcag 180

cgcgagttcc tggtggagtg gcgccgcgtg agtaaggccc cggaggccct ctttttttgtc 240

cagttcgaaa agggggagac ctacttccac ctgcacgtgc tgattgagac catcggggtc 300

aaatccatgg tggtcggccg ctacgtgagc cagattaaag agaagctggt gacccgcatc 360

taccgcgggg tcgagccgca gcttccgaac tggttcgcgg tgaccaaaac gcgaaatggc 420

gccgggggcg ggaacaaggt ggtggacgac tgctacatcc ccaactacct gctccccaag 480

acccagcccg agctccagtg ggcgtggact aacatggacc agtatttaag cgcctgtttg 540

aatctcgcgg agcgtaaacg gctggtggcg cagcatctga cgcacgtgtc gcagacgcag 600

gagcagaaca aagagaatca gaaccccaat tctgacgcgc cggtcatcag gtcaaaaacc 660

tcagccaggt acatggagct ggtcgggtgg ctggtggacc gcgggatcac gtcagaaaag 720

caatggattc aggaggacca ggcctcgtac atctccttca acgccgcctc caactcgcgg 780

tcccagatca aggccgcgct ggacaatgcc tccaagatca tgagcctgac aaagacggct 840

ccggactacc tggtgggcag caacccgccg gaggacatta ccaaaaatcg gatctaccaa 900

atcctggagc tgaacgggta cgatccgcag tacgcggcct ccgtcttcct gggctgggcg 960

caaaagaagt tcgggaagag gaacaccatc tggctctttg ggccggccac gacgggtaaa 1020

accaacatcg cggaagccat cgcccacgcc gtgcccttct acggctgcgt aaactggacc 1080

aatgagaact ttcccttcaa cgattgcgtc gacaagatgg tgatctggtg ggaggagggc 1140

aagatgacgg ccaaggtcgt ggagagcgcc aaggccattc tgggcggaag caaggtgcgc 1200

gtggaccaaa agtgcaagtc atcggcccag atcgaaccca ctcccgtgat cgtcacctcc 1260

aacaccaaca tgtgcgccgt gattgacggg aacagcacca ccttcgagca tcagcagccg 1320

ctgcaggacc ggatgtttga atttgaactt acccgccgtt tggaccatga ctttgggaag 1380

gtcaccaaac aggaagtaaa ggacttttc cggtgggctt ccgatcacgt gactgacgtg 1440

gctcatgagt tctacgtcag aaagggtgga gctaagaaac gccccgcctc caatgacgcg 1500

gatgtaagcg agccaaaacg ggagtgcacg tcacttgcgc agccgacaac gtcagacgcg 1560

gaagcaccgg cggactacgc ggacaggtac caaaacaaat gttctcgtca cgtgggcatg 1620

aatctgatgc tttttccctg taaaacatgc gagagaatga atcaaatttc caatgtctgt 1680

tttacgcatg gtcaaagaga ctgtggggaa tgcttccctg gaatgtcaga atctcaaccc 1740

gtttctgtcg tcaaaaagaa gacttatcag aaactgtgtc caattcatca tatcctggga 1800

agggcacccg agattgcctg ttcggcctgc gatttggcca atgtggactt ggatgactgt 1860

gtttctgagc aataaatgac ttaaaccagg tatggctgct gacggttatc ttccagattg 1920

gctcgaggac aacctttctg a 1941

<210> 1036

<211> 624

<212> PRT

<213> Adeno-associated virus

<400> 1036

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

1 5 10 15

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

20 25 30

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile

35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu

50 55 60

Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val

65 70 75 80

Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu

85 90 95

Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile

100 105 110

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

115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly

130 135 140

Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys

145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
165 170 175
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
180 185 190
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn

195 200 205
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
210 215 220
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
225 230 235 240
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
245 250 255
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys

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

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

385 390 395 400
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val

405 410 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser

420 425 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe

435 440 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln

450 455 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val

465 470 475 480

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala

485 490 495

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu

500 505 510

Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp

515 520 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu

530 535 540

Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys

545 550 555 560

Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser

565 570 575

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

580 585 590

Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser

595 600 605

Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln

610 615 620

<210> 1037

<211> 536

<212> PRT

<213> Adeno-associated virus

<400> 1037

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
1 5 10 15

Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
20 25 30

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
50 55 60

Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65 70 75 80

Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
85 90 95

Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
100 105 110

Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
130 135 140

Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
165 170 175

Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

245 250 255

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

260 265 270

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

275 280 285

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu

290 295 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala

305 310 315 320

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

325 330 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro

340 345 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

355 360 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala

370 375 380

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

385 390 395 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val

405 410 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser

420 425 430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe

435 440 445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln

450 455 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val

465 470 475 480

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala

485 490 495

Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
500 505 510
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
515 520 525

Arg Leu Ala Arg Gly Gln Pro Phe
530 535

<210> 1038
<211> 400
<212> PRT
<213> Adeno-associated virus
<400> 1038
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
1 5 10 15
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
20 25 30
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
35 40 45

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

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

165 170 175

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

Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
245 250 255
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
260 265 270
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
275 280 285
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
290 295 300

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

Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
370 375 380
Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln
385 390 395 400

<210> 1039

<211> 312

<212> PRT

<213> Adeno-associated virus

<400> 1039

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
1 5 10 15

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
20 25 30

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
35 40 45

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
50 55 60

Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
65 70 75 80

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
85 90 95

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
100 105 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
115 120 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
130 135 140

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
145 150 155 160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
165 170 175

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
180 185 190

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
195 200 205

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

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
225 230 235 240
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
245 250 255
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
260 265 270
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu

275 280 285
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
290 295 300
Arg Leu Ala Arg Gly Gln Pro Phe
305 310

<210> 1040
<211> 610
<212> PRT
<213> Adeno-associated virus
<400> 1040

Met Ala Thr Phe Tyr Glu Val Ile Val Arg Val Pro Phe Asp Val Glu
1 5 10 15
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asp Trp Val Thr Gly

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

85 90 95
Ser Gly Ile Ser Ser Met Val Leu Gly Arg Tyr Val Ser Gln Ile Arg
100 105 110
Ala Gln Leu Val Lys Val Val Phe Gln Gly Ile Glu Pro Gln Ile Asn

115 120 125

Asp Trp Val Ala Ile Thr Lys Val Lys Lys Gly Gly Ala Asn Lys Val

130 135 140

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

145 150 155 160

Glu Leu Gln Trp Ala Trp Thr Asn Leu Asp Glu Tyr Lys Leu Ala Ala

165 170 175

Leu Asn Leu Glu Glu Arg Lys Arg Leu Val Ala Gln Phe Leu Ala Glu

180 185 190

Ser Ser Gln Arg Ser Gln Glu Ala Ala Ser Gln Arg Glu Phe Ser Ala

195 200 205

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

210 215 220

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

225 230 235 240

Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr Gly Asn Ser Arg

245 250 255

Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys Ile Met Ser Leu

260 265 270

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

275 280 285

Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn Gly Tyr Asp

290 295 300

Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys Gln Arg Ser Phe

305 310 315 320

Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys

325 330 335

Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro Phe Tyr Gly Cys

340 345 350

Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Asp Lys

355 360 365

Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys Val Val Glu
370 375 380
Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg Val Asp Gln Lys
385 390 395 400
Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val Ile Val Thr Ser

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

465 470 475 480
Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu Lys Ser Leu
485 490 495
Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr Lys Ser Leu Glu
500 505 510
Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro Arg Ser Ser Asp
515 520 525
Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn Trp Asn Ser Arg

530 535 540
Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp Asn Ile Ser Asn
545 550 555 560
Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn Gly Cys Ile
565 570 575
Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile Pro Pro Trp
580 585 590
Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp Ala Asn Lys

595 600 605
Glu Gln

610

<210> 1041

<211> 390

<212> PRT

<213> Adeno-associated virus

<400> 1041

Met Ala Leu Val Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys
1 5 10 15

Gln Trp Ile Gln Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr
20 25 30

Gly Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys
35 40 45

Ile Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser
50 55 60

Val Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met
65 70 75 80

Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys
85 90 95

Gln Arg Ser Phe Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala
100 105 110

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
115 120 125

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
130 135 140

Cys Val Asp Lys Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn
145 150 155 160

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
165 170 175

Val Asp Gln Lys Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val
180 185 190

Ile Val Thr Ser Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser
195 200 205

Thr Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe
210 215 220
Glu Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln
225 230 235 240

Glu Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val
245 250 255
Thr His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala
260 265 270
Glu Lys Ser Leu Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr
275 280 285
Lys Ser Leu Glu Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro
290 295 300

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

Ile Pro Pro Trp Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp
370 375 380
Asp Ala Asn Lys Glu Gln
385 390

<210> 1042
<211> 341
<212> PRT
<213> Adeno-associated virus
<400> 1042
Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser Val
1 5 10 15
Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn

20 25 30

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

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

Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe Glu
165 170 175
Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln Glu
180 185 190
Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val Thr
195 200 205
His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu
210 215 220

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

Asn Ser Arg Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp Asn
275 280 285

Ile Ser Asn Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn
290 295 300
Gly Cys Ile Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile
305 310 315 320
Pro Pro Trp Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp
325 330 335
Ala Asn Lys Glu Gln
340

<210> 1043

<400> 1043

000

<210> 1044

<400> 1044

000

<210> 1045

<400> 1045

000

<210> 1046

<400> 1046

000

<210> 1047

<400> 1047

000

<210> 1048

<400> 1048

000

<210> 1049

<400> 1049

000

<210> 1050

<400> 1050

000

<210> 1051

<211> 62

<212> DNA

<213> Artificial Sequence

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

<400> 1051

aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60

cc 62

<210> 1052

<211> 62

<212> DNA

<213> Artificial Sequence

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

<400> 1052

cgggcgggtg gtggcggcgg ttggggctcg gcgctcgctc gctcgctggg cgggcgggcg 60

gt 62

<210> 1053

<211> 16

<212> PRT

<213> Artificial Sequence

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

<400> 1053

Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr
1 5 10 15

<210> 1054

<211> 20

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 1054

Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr Gly Met Gly Tyr

1 5 10 15

Gly Met Gly Tyr

20

<210> 1055

<211> 16

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 1055

gcgcgctcgc tcgctc 16

<210> 1056

<211> 16

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 1056

gctcgctcgc tcgctg 16

<210> 1057

<211> 6

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<220><221> modified_base

<222> (1)..(1)

<223> a, c, t, or g

<400> 1057

ngttgg 6

<210> 1058

<211> 6

<212> DNA

<213> Artificial Sequence

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

<400> 1058

agttgg 6

<210> 1059

<211> 6

<212> DNA

<213> Artificial Sequence

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

<400> 1059

ggttgg 6

<210> 1060

<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> 1060

Tyr Asn Pro Xaa Xaa Asp Xaa Gly Xaa Xaa Asn
1 5 10

<210> 1061

<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> 1061

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> 1062

<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

<220>

<223> See specification as filed for detailed description of substitutions and preferred embodiments

<400> 1062

aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agggncagtc t 51

<210> 1063

<211> 50

<212> DNA

<213> Alphatorquevirus sp.

<400> 1063

aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 50

<210> 1064

<211> 50

<212> DNA

<213> Betatorquevirus sp.

<

400> 1064

aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agatcagtct 50

<210> 1065

<211> 50

<212> DNA

<213> Gammatorquevirus sp.

<400> 1065

aggtgagtga aaccaccgag gtctaggggc aattcgggct agggcagtct 50

<210> 1066

<211> 150

<212> PRT

<213> Betatorquevirus sp.

<400> 1066

Met Pro Trp Trp Tyr Arg Arg Arg Ser Tyr Asn Pro Trp Arg Arg Arg
1 5 10 15

Asn Trp Phe Arg Arg Pro Arg Lys Thr Ile Tyr Arg Arg Tyr Arg Arg

20 25 30

Arg Arg Arg Trp Pro Thr Tyr Val Val Pro Asn Asn Phe Asn Glu Thr
35 40 45

Thr Ser Leu Gln Asn Pro Thr Thr Arg Pro Glu His Phe Leu Tyr Ser
50 55 60

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

65 70 75 80

Leu Lys Asp Trp Glu Thr Lys Glu Thr Ser Leu Leu Ser Thr Glu Tyr

85 90 95

Arg Phe Ala Glu Pro Thr Gln Thr Gln Ala Pro Gln Glu Asp Pro Ser

100 105 110

Ser Glu Glu Glu Glu Glu Ser Asn Leu Phe Glu Arg Leu Leu Arg Gln

115 120 125

Arg Thr Lys Gln Leu Gln Leu Lys Arg Arg Ile Ile Gln Thr Leu Lys

130 135 140

Asp Leu Gln Lys Leu Glu

145 150

<210> 1067

<211> 62

<212> DNA

<213> Artificial Sequence

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

<400> 1067

aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60

cc 62