



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0038669  
(43) 공개일자 2022년03월29일

(51) 국제특허분류(Int. Cl.)  
C12N 15/86 (2006.01) A61K 39/00 (2006.01)  
A61K 39/12 (2006.01) A61P 31/14 (2006.01)  
C07K 14/005 (2006.01)  
(52) CPC특허분류  
C12N 15/86 (2013.01)  
A61K 39/12 (2013.01)  
(21) 출원번호 10-2022-7001607  
(22) 출원일자(국제) 2020년06월18일  
심사청구일자 없음  
(85) 번역문제출일자 2022년01월17일  
(86) 국제출원번호 PCT/GB2020/051465  
(87) 국제공개번호 WO 2020/254804  
국제공개일자 2020년12월24일  
(30) 우선권주장  
1908729.5 2019년06월18일 영국(GB)

(71) 출원인  
임피리얼 컬리지 이노베이션스 리미티드  
영국 런던 에스더블유7 2에이제트 이그지비션 로드  
임피리얼 컬리지 앞 본관1층  
(72) 발명자  
샤록 로빈  
영국 런던 더블유2 1파워 패딩턴 노퍽 플레이스  
생 메리즈 캠퍼스 임피리얼 컬리지 런던 디파트먼트  
트 오브 인팩서스 디지즈  
블랙크니 안나  
영국 런던 더블유2 1파워 패딩턴 노퍽 플레이스  
생 메리즈 캠퍼스 임피리얼 컬리지 런던 디파트먼트  
트 오브 인팩서스 디지즈  
맥케이 폴  
영국 런던 더블유2 1파워 패딩턴 노퍽 플레이스  
생 메리즈 캠퍼스 임피리얼 컬리지 런던 디파트먼트  
트 오브 인팩서스 디지즈  
(74) 대리인  
특허법인태평양

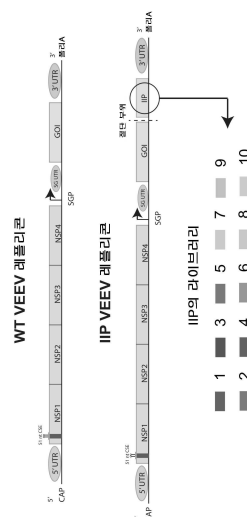
전체 청구항 수 : 총 36 항

(54) 발명의 명칭 RNA 구조체

(57) 요약

본 발명은 (i) 적어도 하나의 치료 생체분자; 및 (ii) 적어도 하나의 선천적 억제제 단백질(IIP);을 암호화하는 RNA 구조체에 관한 것이다. 상기 구조체는 RNA 레플리콘 및 saRNA 분자이고, 본 발명은 이러한 RNA 레플리콘을 암호화하는 유전자 구조체 또는 벡터를 포함한다. 본 발명은 치료법, 예를 들면 질환을 치료하거나 및/또는 백신을 운반하는데 이러한 RNA 구조체 및 레플리콘을 사용하는 것으로 확장된다. 본 발명은 이러한 RNA 구조체를 포함하는 약학적 조성물, 및 그의 방법 및 용도로 확장된다.

대표도 - 도1



(52) CPC특허분류

**A61P 31/14** (2018.01)

**C07K 14/005** (2013.01)

A61K 2039/53 (2013.01)

A61K 2039/70 (2013.01)

C12N 2760/18022 (2013.01)

C12N 2760/20134 (2013.01)

C12N 2770/20022 (2013.01)

C12N 2770/20034 (2013.01)

C12N 2770/36143 (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

(i) 적어도 하나의 치료 생체분자; 및 (ii) 적어도 하나의 선천적 억제제 단백질(IIP)을 암호화하는 RNA 구조체.

#### 청구항 2

청구항 1에 있어서,

상기 RNA 구조체는 자가-증폭 RNA(saRNA)를 포함하고, 바람직하게는 saRNA 구조체인 RNA 구조체.

#### 청구항 3

청구항 1 또는 청구항 2에 있어서,

상기 RNA 구조체는 알파바이러스; 피코르나바이러스; 플라비바이러스; 루비바이러스; 페스티바이러스; 헤파시바이러스; 칼리시바이러스 및 코로나바이러스;로 이루어진 속의 군으로부터 선택되는 양성 가닥 RNA 바이러스를 포함하거나 이로부터 유래되는 RNA 구조체.

#### 청구항 4

청구항 1 내지 청구항 3 중 어느 한 항에 있어서,

상기 RNA 구조체는 베네주엘라 말 뇌염 바이러스(VEEV); 엔테로바이러스 71; 뇌심근염 바이러스; 쿨진 바이러스; 및 중동 호흡기 증후군 바이러스;로 이루어진 종의 군으로부터 선택되는 바이러스를 포함하거나 이로부터 유래되는 RNA 구조체.

#### 청구항 5

청구항 1 내지 청구항 4 중 어느 한 항에 있어서,

상기 구조체는 VEEV로부터 유래되는 RNA 구조체.

#### 청구항 6

청구항 1 내지 청구항 5 중 어느 한 항에 있어서,

상기 적어도 하나의 치료 생체분자는 백신 구조체, 또는 치료 단백질인 RNA 구조체.

#### 청구항 7

청구항 1 내지 청구항 6 중 어느 한 항에 있어서,

상기 적어도 하나의 치료 생체분자는:

(i) 내인성 숙주 유전자의 발현을 조절할 수 있는 RNA 분자이거나;

(ii) 박테리아, 바이러스, 진균, 원생동물 또는 기생충으로부터 유래되거나;

(iii) 효소; 효소 억제제; 호르몬; 면역 시스템 단백질; 수용체; 결합 단백질; 전사 또는 번역 인자; 종양 성장 저해 단백질; 구조 단백질; 및 혈액 단백질;로 이루어진 군으로부터 선택되거나;

(iv) 항원, 선택적으로는 종양 항원인 RNA 구조체.

#### 청구항 8

청구항 1 내지 청구항 7 중 어느 한 항에 있어서,

상기 RNA 분자에 의해 암호화되는 적어도 하나의 치료 생체분자는 박테리아, 바이러스, 진균, 원생동물; 및 기생충으로 이루어진 군으로부터 선택되는 병원체로부터 유래되는 단백질 또는 펩티드인 RNA 구조체.

#### 청구항 9

청구항 8에 있어서,

상기 단백질 또는 펩티드는 오르토믹소바이러스; 파라믹소비리대 바이러스; 메타뉴모바이러스 및 모르빌리바이러스; 뉴모바이러스; 파라믹소바이러스; 폭스비리대; 메타뉴모바이러스; 모르빌리바이러스; 피코르나바이러스; 엔테로바이러스; 분야바이러스; 플레보바이러스; 나이로바이러스; 헤파르나바이러스; 토가바이러스; 알파바이러스; 아르테리바이러스; 플라비바이러스; 페스티바이러스; 헤파르나바이러스; 라브도바이러스; 칼리시비리대; 코로나바이러스; 레트로바이러스; 레오바이러스; 파르보바이러스; 델타 간염 바이러스(HDV); E형 간염 바이러스(HEV); 인간 헤르페스바이러스 및 파포바바이러스;로 이루어진 군으로부터 선택되는 바이러스로부터 유래되는 바이러스 항원인 RNA 구조체.

#### 청구항 10

청구항 9에 있어서,

상기 코로나바이러스는 SARS CoV-1, SARS-CoV-2, MERS, 인간 호흡기 코로나바이러스, 조류 전염성 기관지염(IBV), 마우스 간염 바이러스(MHV), 또는 돼지 전염성 위장염 바이러스(TGEV)인 RNA 구조체.

#### 청구항 11

청구항 1 내지 청구항 10 중 어느 한 항에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 (i) 흑색종 분화-연관 단백질 5(MDA5)의 작용을 감소 또는 차단하거나, 및/또는 (ii) RNA에 대한 PKR 활성화 단백질의 결합을 차단 또는 감소시킬 수 있는 RNA 구조체.

#### 청구항 12

청구항 1 내지 청구항 11 중 어느 한 항에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 중등 호흡기 증후군 코로나바이러스 MERS 코로나바이러스(ORF4a)인 RNA 구조체.

#### 청구항 13

청구항 1 내지 청구항 12 중 어느 한 항에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 파라인플루엔자 바이러스 타입 5 V 단백질(PIV5 V)인 RNA 구조체.

#### 청구항 14

청구항 1 내지 청구항 13 중 어느 한 항에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 코로나바이러스 ORF3b, 바람직하게는 SARS-CoV-2 ORF3b인 RNA 구조체.

#### 청구항 15

청구항 1 내지 청구항 14 중 어느 한 항에 있어서,

상기 RNA 구조체는 서열번호 11, 서열번호 15 및/또는 서열번호 20을 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함하는 RNA 구조체.

#### 청구항 16

청구항 1 내지 청구항 15 중 어느 한 항에 있어서,

상기 RNA 구조체는 실질적으로 서열번호 47, 서열번호 48 및/또는 서열번호 56으로 나타낸 것과 같은 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함하는 RNA 구조체.

#### 청구항 17

청구항 1 내지 청구항 16 중 어느 한 항에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 MDA5 활성화의 하나 이상의 하류 경로를 억제하거나, dsRNA의 MDA/PACT 인식의 하나 이상의 하류 경로를 차단할 수 있는 RNA 구조체.

#### 청구항 18

청구항 17에 있어서,

상기 적어도 하나의 선천적 억제제 단백질은 HSV-2 Us1; HSV-1 Us1; HSV-1 Us11; OV20.0L; BVDV Npro; 란가트 바이러스 NS5; 및 인플루엔자 NS1;로 이루어진 군으로부터 선택되는 RNA 구조체.

#### 청구항 19

청구항 1 내지 청구항 18 중 어느 한 항에 있어서,

상기 RNA 구조체는 상기 적어도 하나의 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열에 작동가능하게 연결된 프로모터 또는 서브게놈 프로모터를 포함하여서, 상기 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 뉴클레오타이드 서열의 전사를 가능하게 하는 RNA 구조체.

#### 청구항 20

청구항 19에 있어서,

상기 프로모터는 26S이고, 선택적으로 상기 프로모터는 실질적으로 서열번호 57에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함하는 RNA 구조체.

#### 청구항 21

청구항 1 내지 청구항 20 중 어느 한 항에 있어서,

상기 RNA 구조체는 상기 치료 생체분자를 암호화하는 서열 및 상기 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열 사이에 배치된 링커 서열을 포함하고, 선택적으로 상기 링커 서열은 소화됨으로써 상기 적어도 하나의 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 분리하도록 구성되는 펩티드 스페이서를 암호화하고, 바람직하게는 상기 펩티드 스페이서는 2A 펩티드 또는 푸린/2A 펩티드인 RNA 구조체.

#### 청구항 22

청구항 1 내지 청구항 21 중 어느 한 항에 있어서,

상기 RNA 구조체는 실질적으로 서열번호 38 또는 서열번호 39에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 단편 또는 변이체를 포함하는 RNA 구조체.

#### 청구항 23

청구항 1 내지 청구항 22 중 어느 한 항의 RNA 구조체를 암호화하는 핵산 서열로서,

선택적으로 상기 서열은 실질적으로 서열번호 40 또는 서열번호 41에 나타난 것, 또는 그의 단편 또는 변이체인 핵산 서열.

#### 청구항 24

청구항 23에 따른 핵산 서열을 포함하는 발현 카세트.

#### 청구항 25

청구항 24에 따른 발현 카세트를 포함하는 재조합 벡터로서,

선택적으로 상기 벡터는 실질적으로 서열번호 35 내지 서열번호 37 중 어느 하나에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함하는 재조합 벡터.

## 청구항 26

청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산 서열, 청구항 24에 따른 발현 카세트 또는 청구항 25에 따른 벡터, 및 약학적으로 허용가능한 비히클을 포함하는 약학적 조성물.

## 청구항 27

a) i) 숙주 세포 내로 청구항 25에 따른 벡터를 도입하는 단계; 및

ii) 청구항 1 내지 청구항 22 중 어느 한 항의 RNA 구조체의 생산을 야기하는 조건 하에 상기 숙주 세포를 배양하는 단계; 또는

b) 청구항 25에 따른 벡터로부터 RNA 구조체를 전사시키는 단계;를 포함하는 청구항 1 내지 청구항 22 중 어느 한 항의 RNA 구조체의 제조 방법.

## 청구항 28

약제로서 또는 치료법에 사용하기 위한 청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물.

## 청구항 29

원생동물, 진균, 박테리아 또는 바이러스 감염의 예방, 개선 또는 치료에 사용하기 위한 청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물.

## 청구항 30

암의 예방, 개선 또는 치료에 사용하기 위한 청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물.

## 청구항 31

청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물, 및 선택적으로 어주번트를 포함하는 백신.

## 청구항 32

대상체에서 면역 반응을 자극하는데 사용하기 위한 청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물.

## 청구항 33

줄기 세포 치료법에 사용하기 위한 청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물.

## 청구항 34

청구항 1 내지 청구항 22 중 어느 한 항에 따른 RNA 구조체, 청구항 23에 따른 핵산, 청구항 24에 따른 발현 카세트, 청구항 25에 따른 벡터 또는 청구항 26에 따른 약학적 조성물을 세포에 전달하는 단계를 포함하는 생체의 또는 시험관내에서 세포를 변형시키는 방법.

## 청구항 35

청구항 28의 방법으로부터 수득되거나, 상기 방법에 의해 수득될 수 있는 변형된 세포.

## 청구항 36

청구항 35에 있어서,

치료법, 선택적으로 세포 치료법에 사용하기 위한 변형된 세포.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 RNA 구조체에 관한 것이며, 특히, 배타적인 것은 아니지만, RNA 레플리콘(replicon) 및 saRNA 분자, 및 이러한 RNA 레플리콘을 암호화하는 유전자 구조체 또는 벡터(vector)에 관한 것이다. 본 발명은 치료법, 예를 들면 질환의 치료 및/또는 백신 운반에 있어서의 이러한 RNA 구조체 및 레플리콘의 용도로 확장된다. 본 발명은 이러한 RNA 구조체를 포함하는 약학적 조성물, 및 이의 방법 및 용도로 확장된다.

### 배경 기술

[0002] 메신저 RNA(mRNA)는 생물치료제(biotherapeutics)를 위한 유망한 도구이다. 그러나, mRNA 치료제는 소동물에서 매우 효과적인 것으로 나타났지만, 그 결과는 상기 제형이 인간에서 용량-확대 연구로 번역될 때 선형으로 규모가 커지지 않는다는. 아울러, 인터페론 반응의 유도 및 연관된 부작용은 증가된 용량의 RNA가 인간에서 효과적일 가능성과 관련한 속도-제한요인(rate-limiting)인 것으로 나타나 있다. 이러한 불일치의 이유는 명확하지 않지만, 본 발명자들은 인간 선천적 감지(innate sensing)에서의 내재하는 차이가 실험실로부터 임상으로 RNA 치료제를 번역하기 위한 장애물을 제공한다는 가설을 세운다. 아울러, RNA의 선천적 감지는 단백질 발현의 억제와 연관된다. 지금까지, 외인성(exogenous) RNA의 선천적 인식을 극복하기 위한 주된 접근법은 선천적 감지 메커니즘에 의해 덜 검출가능한 변형된 리보뉴클레오타이드를 사용하는 것이다. 그러나, 변형된 mRNA는 완전히 검출불가능하지는 않으며, 인간에게 사용할 때 여전히 인터페론의 일부 활성화, 단백질 사일런싱(silencing) 및 감소된 내약성(tolerability)을 야기한다.

[0003] 다른 접근법은 전형적으로 그 비-구조 단백질 내에 폴리머라아제(polymerase) 활성을 암호화함으로써 그 자신의 RNA를 자가-증폭하는 능력을 갖고 있는 알파바이러스 백본(alphavirus backbone)에 기반하는 자가-증폭형(self-amplifying) 또는 saRNA 벡터를 사용하는 것이다. 종래 기술의 방법은 상기 벡터의 구조 단백질을 관심있는 유전자(GOI)로 교체하여 이를 백신 구조체로 하거나, 치료 단백질을 암호화하는 것을 수반한다. 다른 버전의 saRNA는 피코르나바이러스(picornavirus), 플라비바이러스(flavivirus), 및 코로나바이러스(coronavirus)에 기반한다. saRNA가 표적 세포의 세포질 내로 들어갈 때, 이것은 상기 암호화된 폴리머라아제 기구(machinery)에 의한 RNA의 증폭 및 GOI의 매우 높은 발현 레벨을 유도한다. 결과적으로, saRNA는 mRNA보다 낮은(10 내지 100배 낮은) 용량의 saRNA로 면역 반응을 유도하는 것으로 나타났고, 마우스에서 최대 60일까지 연장된 단백질 발현을 야기한다.

[0004] 그러나, saRNA를 이용하는 결점은 이것은 또한 선천적 인식에 의해 감지되어 상기 종래 기술의 saRNA의 단백질 발현 및 자가-증폭을 제한하는 항-바이러스 반응을 촉발하게 된다는 것이다. saRNA의 선천적 감지는 그 큰 크기(전형적으로 >5,000 염기) 및 이중 가닥 영역(dsRNA)을 포함하는 고도한 2차 구조로 인해 mRNA의 경우와 상이하다. 길고 이중 가닥인 RNA는 MDA5(흑색종 분화-연관 단백질 5) 경로를 통해 선천적 반응을 촉발한다. 이것은 길고 dsRNA인 RNA에 대한 PACT(PKR 활성화 단백질)의 결합에 의해 MDA5의 올리고머화(oligomerization)를 촉진하고 이어서 saRNA의 복제 및 발현을 억제하는 하류의 신호전달 캐스케이드(cascade)를 촉발함으로써 용이하게 된다.

[0005] 따라서, 본 기술분야에는 RNA 치료제가 환자에게 운반 및 발현되어 RNA의 선천적 면역 시스템 감지를 극복할 수 있는 새로운 수단을 생성하기 위한 필요성이 있다.

### 발명의 내용

[0006] 본 발명자들은 선천적 면역 시스템 기구를 차단 또는 감소시키는 선천적 억제제 단백질을 발현함으로써 RNA의 선천적 면역 시스템 감지를 유리하게 극복하고, 개선된 단백질 발현 및 자가-증폭을 야기하는 신규한 자가-증폭 RNA(saRNA)를 개발하였다.

[0007] 따라서, 본 발명의 제1 측면에서, (i) 적어도 하나의 치료 생체분자(biomolecule); 및 (ii) 적어도 하나의 선천적 억제제 단백질(IIP);을 암호화하는 RNA 구조체가 제공된다.

[0008] RNA 레플리콘 또는 구조체는 백신 및 치료제에 대한 관심있는 유전자의 운반 및 발현을 위한 잠재적인 도구로 상정되어 왔다. 그러나, 이중 가닥 RNA(dsRNA)는 단백질 번역을 억제하는 반응을 촉발하는 선천적 감지 메커니

증에 의해 세포내에서 인지된다. 결과적으로, 레플리콘에서 암호화되는 관심있는 유전자의 발현은 현저하게 손상되고, 따라서 RNA 레플리콘의 치료 잠재력은 제한된다. 따라서, 본 발명의 RNA 구조체는 전이유전자(transgene) 발현의 하류 억제를 제거하는 하나 이상의 선천적 억제 단백질(들), 즉 IIP를 암호화하기 때문에 이러한 문제점을 극복한다. 인터페론의 유도는 선천적 인식의 하류 결과이지만, 다른 분자 및 경로가 유도되거나 유도될 수 있고, 이들 중 임의의 것은 RNA 구조체가 갖고 있는 하나 이상의 IIP에 의해 억제될 것임이 인식될 것이다. saRNA를 이용해 인터페론 반응을 제거하는 이전에 공개된 유일한 접근법은 백신니아(vaccinia) 바이러스 유래의 인터페론 억제 단백질인 E3, K3 및 B18을 사용하였다. 그러나, 상기 연구에서, 인터페론 억제 단백질은 saRNA와 조합된 별도의 mRNA 분자로서 운반 및 체형화되었다. 이것은 saRNA 및 mRNA 모두의 제조를 필요로 하고, 동일한 세포 내로 공동-운반하고 단백질 발현에서 임의의 관찰가능한 향상을 제공하는 것을 보장하기 위하여 본 발명에 따른 레플리콘 RNA 구조체보다 3-6배 더 많은 백신니아 mRNA를 사용하는 것을 필요하게 한다. 아울러, 발현의 동역학이 mRNA 및 saRNA의 경우에 상이하여서, mRNA로부터 발현된 IIP의 임의의 유익한 효과는 본 발명의 RNA 구조체(최대 60일까지)와 비교하여 매우 짧은 기간(최대 72시간)뿐일 것이다.

[0009] 유리하게는, 제1 측면의 RNA 구조체에서, 하나 이상의 IIP의 존재는 관심있는 펩티드 또는 단백질을 이용한 이중 단백질 발현을 가능하게 한다. 종래 기술에서 기술된 것과 같이 하나는 관심있는 펩티드/단백질을 암호화하고 하나는 IIP를 암호화하는 2개의 상이한 가닥의 RNA를 운반하는 것과 대조적으로, 본 발명의 구조체를 이용함으로써 단지 단일 가닥만이 표적 세포로 운반되고, 이로 인해 RNA 및 선천적 억제 단백질의 동시국소화(colocalization)를 보장한다. 상기 IIP는 RNA의 선천적 감지를 억제하고, 따라서 더 높은 단백질 발현을 가능하게 하며, 상기 IIP 발현 자체는 관심있는 유전자, 즉 치료 생체분자를 갖는 서브게놈(subgenome) 가닥에 공동-발현됨으로써 자가-증폭된다. 실시예에 기술된 것과 같이, ("스틸리콘(Stealthicon)"으로도 알려져 있는) 루시페라아제(luciferase)를 암호화하는 본 발명의 RNA 구조체는 놀랍게도 인간 세포주에서 시험관내 루시페라아제 단백질 발현 레벨을 최대 2 자릿수 크기까지 증가시키고, 또한 BL/6 마우스에서 종래 VEEV RNA 레플리콘과 비교하여 생체내 루시페라아제의 단백질 발현의 크기 및 기간 모두를 증가시키는 것으로 나타났다. 상기 RNA 구조체는 생체내에서 본 발명의 RNA 분자가 갖고 있는 유전자를 발현할 수 있는 것으로 판명되었기 때문에, 숙련된 기술자는 루시페라아제 리포터(reporter)가 치료 생체분자를 정확히 대표함을 용이하게 인식할 것이다. 이와 같이, 상기 루시페라아제는 본 발명의 saRNA 구조체가 임의의 치료적으로 활성인 생체분자를 발현하기 위해 사용될 수 있다는 개념 증명(proof of concept)의 강력한 증거를 제공한다.

[0010] 숙련된 기술자는 RNA 구조체가 또한 자가-복제(self-replicating) RNA 바이러스 벡터, 또는 RNA 레플리콘으로 나타낼 수 있음을 이해할 것이다. 상기 RNA 구조체는 이중-가닥 또는 단일-가닥일 수 있다. 바람직하게는, 상기 RNA 구조체는 자가-증폭 RNA(saRNA)를 포함하고, 바람직하게는 saRNA 구조체이다.

[0011] 바람직하게는, 상기 RNA 구조체는 알파바이러스; 피코르나바이러스; 플라비바이러스; 루비바이러스(*rubivirus*); 페스티바이러스(*pestivirus*); 헤파시바이러스(*hepacivirus*); 칼리시바이러스 또는 코로나바이러스로 이루어진 속(genus)의 군으로부터 선택되는 양성 가닥(positive stranded) RNA 바이러스를 포함하거나 이로부터 유래된다.

[0012] 적합한 야생형 알파바이러스 서열은 잘 알려져 있다. 적합한 알파바이러스의 대표적인 예는 아우라(*Aura*), 베바루(*Bebaru*) 바이러스, 카바소우(*Cabassou*), 치쿤군야(*Chikungunya*) 바이러스, 동부 말 뇌척수염 바이러스, 포트모건(*Fort Morgan*), 게타(*Getah*) 바이러스, 키질라가치(*Kyzylgach*), 마야로(*Mayaro*), 마야로 바이러스, 미들버그(*Middleburg*), 무캄보(*Mucambo*) 바이러스, 은두무(*Ndumu*), 픽수나(*Pixuna*) 바이러스, 로스 리버(*Ross River*) 바이러스, 쉼리키 포레스트(*Semliki Forest*), 신드비스(*Sindbis*) 바이러스, 토나테(*Tonate*), 트리니티(*Trinit*), 우나(*Una*), 베네주엘라 말 뇌척수염, 서부 말 뇌척수염, 와타로아(*Whataroa*), 및 Y-62-33을 포함한다.

[0013] 바람직하게는, 상기 RNA 구조체는 베네주엘라 말 뇌염 바이러스(VEEV); 엔테로바이러스(*enterovirus*) 71; 뇌심근염 바이러스; 쿤진(*Kunjin*) 바이러스; 및 중동 호흡기 증후군 바이러스로 이루어진 종의 군으로부터 선택되는 바이러스를 포함하거나 이로부터 유래된다. 바람직하게는, 상기 벡터는 VEEV로부터 유래된다.

[0014] 상기 RNA 구조체는 적어도 하나의 치료 생체분자를 암호화하는 서열을 포함한다. 상기 적어도 하나의 치료 생체분자는 백신 구조체, 또는 치료 단백질이거나 이를 포함할 수 있다. 숙련된 기술자는 치료 단백질이 바람직하게는 인간에서 치료 적용분야를 갖는 임의의 단백질에 관한 것임을 이해할 것이다. 상기 RNA 분자에 의해 암호화될 수 있는 예시적인 치료 생체분자는 박테리아, 바이러스, 진균, 원생동물/또는 기생충과 같은 병원체로부터 유래되는 단백질 및 펩티드를 포함한다. 바람직하게는, 상기 단백질 및 펩티드는 항원이다.

- [0015] 바이러스로부터 유래되는 단백질 및 캡티드는 바이러스 항원일 수 있다. 상기 바이러스 항원은 오르토믹소바이러스(*Orthomyxovirus*); 파라믹소비리데(*Paramyxoviridae*) 바이러스; 메타뉴모바이러스(*Metapneumovirus*) 및 모르빌리바이러스(*Morbillivirus*); 뉴모바이러스; 파라믹소바이러스(*Paramyxovirus*); 포스비리데(*Poxviridae*); 메타뉴모바이러스; 모르빌리바이러스; 피코르나바이러스; 엔테로바이러스; 분야바이러스(*Bunyavirus*); 플레보바이러스(*Phlebovirus*); 나이로바이러스(*Nairovirus*); 헤파르나바이러스(*Heparnavirus*); 토가바이러스(*Togavirus*); 알파바이러스; 아르테리바이러스(*Arterivirus*); 플라비바이러스; 페스티바이러스; 헤파드나바이러스(*Hepadnavirus*); 라브도바이러스(*Rhabdovirus*); 칼리시비리데(*Caliciviridae*); 코로나바이러스; 레트로바이러스(*Retroviruses*); 레오바이러스(*Reovirus*); 파르보바이러스(*Parvovirus*); 델타 간염 바이러스(HDV); E형 간염 바이러스(HEV); 인간 헤르페스바이러스(*Herpesvirus*) 및 파포마바이러스(*Papovavirus*);로 이루어진 군으로부터 선택되는 바이러스로부터 유래될 수 있다.
- [0016] 상기 오르토믹소바이러스는 인플루엔자(Influenza) A, B 및 C일 수 있다. 상기 파라믹소비리데 바이러스는 뉴모바이러스(RSV), 파라믹소바이러스(PIV)일 수 있다. 상기 메타뉴모바이러스는 모르빌리바이러스(예컨대, 홍역)일 수 있다. 상기 뉴모바이러스는 호흡기 세포융합 바이러스(RSV), 소 호흡기 세포융합 바이러스, 마우스의 폐렴 바이러스, 또는 터키 비기관염(*rhinotracheitis*) 바이러스일 수 있다. 상기 파라믹소바이러스는 파라인쿠엔자(*Parainkuenza*) 바이러스 타입 1 - 4(PIV), 볼거리(Mumps), 센다이(*Sendai*) 바이러스, 원숭이 바이러스 5, 소 파라인쿠엔자 바이러스, 니파바이러스(*Nipahvirus*), 헨이파바이러스(*Henipavirus*) 또는 뉴캐슬(Newcastle) 질환 바이러스일 수 있다. 상기 포스비리데는 바리올라 베라(*Variola vera*), 예를 들면 바리올라 메이저(major) 및 바리올라 마이너(minor)일 수 있다. 상기 메타뉴모바이러스는 인간 메타뉴모바이러스(hMPV) 또는 조류 메타뉴모바이러스(aMPV)일 수 있다. 상기 모르빌리바이러스는 홍역일 수 있다. 상기 피코르나바이러스는 엔테로바이러스, 리노바이러스(*Rhinovirus*), 헤파르나바이러스, 파레코바이러스(*Par에코바이러스*), 카디오바이러스(*Cardiovirus*) 및 아프토바이러스(*Aphthovirus*)일 수 있다. 상기 엔테로바이러스는 폴리오바이러스(*Poliovirus*) 타입 1, 2 또는 3, 콕사키(Coxsackie) A 바이러스 타입 1 내지 22 및 24, 콕사키 B 바이러스 타입 1 내지 6, 에코바이러스(Echovirus)(ECHO) 바이러스 타입 1 내지 9, 11 내지 27 및 29 내지 34 또는 엔테로바이러스 68 내지 71일 수 있다. 상기 분야바이러스는 캘리포니아 뇌염 바이러스일 수 있다. 상기 플레보바이러스는 리프트 계곡열 바이러스일 수 있다. 상기 나이로바이러스는 크림-콩고(Crimean-Congo) 출혈열 바이러스일 수 있다. 상기 헤파르나바이러스는 A형 간염 바이러스(HAV)일 수 있다. 상기 토가바이러스는 루비바이러스일 수 있다. 상기 플라비바이러스는 진드기-매개 뇌염(TBE) 바이러스, 탕기(타입 1, 2, 3 또는 4) 바이러스, 황열 바이러스, 일본 뇌염 바이러스, 키아사누 삼림 바이러스(Kyasanur Forest Virus), 서나일(West Nile) 뇌염 바이러스, 세인트 루이스 뇌염 바이러스, 러시아 봄-여름 뇌염 바이러스 또는 포와산(Powassan) 뇌염 바이러스일 수 있다. 상기 페스티바이러스는 소 바이러스 설사(BVDV), 고전 돼지열(CSFV) 또는 보더(Border) 질환(BDV)일 수 있다. 상기 헤파드나바이러스는 B형 간염 바이러스 또는 C형 간염 바이러스일 수 있다. 상기 라브도바이러스는 리사바이러스(Lyssavirus)(광견병 바이러스) 또는 수포성바이러스(Vesiculovirus)(VSV)일 수 있다. 상기 칼리시비리데는 노워크(Norwalk) 바이러스, 또는 노워크-유사 바이러스, 예컨대 하와이 바이러스 및 스노우 마운틴 바이러스(Snow Mountain Virus)일 수 있다. 상기 코로나바이러스는 SARS CoV-1, SARS-CoV-2, MERS, 인간 호흡기 코로나바이러스, 조류 전염성 기관지염(IBV), 마우스 간염 바이러스(MHV), 또는 돼지 전염성 위장염 바이러스(TGEV)일 수 있다. 상기 레트로바이러스는 온코바이러스(Oncovirus), 렌티바이러스(Lentivirus) 또는 스푸마바이러스(Spumavirus)일 수 있다. 상기 레오바이러스는 오르토레오(Orthoreo) 바이러스, 로타바이러스(Rotavirus), 오르비바이러스(Orbivirus), 또는 콜티바이러스(Coltivirus)일 수 있다. 상기 파르보바이러스는 파르보바이러스 B 19일 수 있다. 상기 인간 헤르페스바이러스는 헤르페스 심플렉스 바이러스(Herpes Simplex Virus)(HSV), 바리셀라-조스터(Varicella-zoster) 바이러스(VZV), 엡스테인-바(Epstein-Barr) 바이러스(EBV), 사이토메갈로바이러스(CMV), 인간 헤르페스바이러스 6(HHV6), 인간 헤르페스바이러스 7(HHV7), 또는 인간 헤르페스바이러스 8(HHV8)일 수 있다. 상기 파포마바이러스는 유두종 바이러스, 폴리오마바이러스(Polyomavirus), 아데노바이러스(Adenovirus) 또는 아레나바이러스(Arenavirus)일 수 있다.
- [0017] 실시예 및 도 13에 나타난 것과 같이, 바람직한 구현예에서, 상기 바이러스 항원은 광견병 바이러스 항원, 바람직하게는 광견병 바이러스 당단백질(glycoprotein)일 수 있다. 다른 바람직한 구현예에서, 실시예 및 도 29에 나타난 것과 같이, 상기 바이러스 항원은 코로나바이러스 항원일 수 있다. 바람직하게는, 상기 코로나바이러스 항원은 표면 당단백질, 보다 바람직하게는 SARS-CoV-2 표면 당단백질이다.
- [0018] 박테리아로부터 유래되는 단백질 및 캡티드는 박테리아 항원일 수 있다.
- [0019] 상기 박테리아 항원은 나이세리아 메닝기티데스(*Neisseria meningitides*), 스트렙토코커스 뉴모니에

(*Streptococcus pneumoniae*), 스트렙토코커스 피로게네스(*Streptococcus pyogenes*), 모락셀라 카타랄리스(*Moraxella catarrhalis*), 보르데텔라 페르투스시스(*Bordetella pertussis*), 부르크홀데리아 속(*Burkholderia* sp.)(예컨대, 부르크홀데리아 말레이, 부르크홀데리아 슈도말레이 및 부르크홀데리아 세파시아), 스태필로코커스 아우레우스(*Staphylococcus aureus*), 해모필러스 인쿠엔자에(*Haemophilus inkuenzae*), 클로스트리디움 테타니(*Clostridium tetani*)(파상풍), 클로스트리디움 퍼프린젠스(*Clostridium perfringens*), 클로스트리디움 보툴리눔(*Clostridium botulinum*), 코리네박테리움 디프테리아에(*Corynebacterium diphtheriae*)(디프테리아), 슈도모나스 애루기노사(*Pseudomonas aeruginosa*), 레지오넬라 뉴모필라(*Legionella pneumophila*), 콕시엘라 부르네티이(*Coxiella burnetii*), 브루셀라 속(*Brucella* sp.)(예컨대, B. 아보르투스, B. 카니스, B. 멜리텐시스, B. 네오토마에, B. 오비스, B. 수이스 및 B. 피니페디에), 프란시셀라 속(*Francisella* sp.)(예컨대, F. 노비시다, F. 필로마리자 및 F. 툴라렌시스), 스트렙토코커스 아갈락티아에(*Streptococcus agalactiae*), 나이세리아 고노레아(*Neisseria gonorrhoeae*), 클라미디아 트라코마티스(*Chlamydia trachomatis*), 트레포네마 팔리둠(*Treponema pallidum*)(매독), 해모필러스 두크레이이(*Haemophilus ducreyi*), 엔테로코커스 패칼리스(*Enterococcus faecalis*), 엔테로코커스 패시움(*Enterococcus faecium*), 헬리코박터 파일로리(*Helicobacter pylori*), 스태필로코커스 사프로피티쿠스(*Staphylococcus saprophyticus*), 예르시니아 엔테로콜리티카(*Yersinia enterocolitica*), 대장균(*E. coli*), 바실러스 안트라시스(*Bacillus anthracis*)(탄저병), 예르시니아 페스티스(*Yersinia pestis*)(페스트), 미코박테리움 투베르쿨로시스(*Mycobacterium tuberculosis*), 리케차(*Rickettsia*), 리스테리아(*Listeria*), 클라미디아 뉴모니아(*Chlamydia pneumoniae*), 비브리오 콜레라에(*Vibrio cholerae*), 살모넬라 타이피(*Salmonella typhi*)(장티푸스), 보렐리아 부르그도르퍼(*Borrelia burgdorferi*), 포르필라모나스 속(*Porphyromonas* s) 및 클레브시엘라 속(*Klebsiella* sp.)으로 이루어진 군으로부터 선택되는 박테리아로부터 유래될 수 있다.

[0020] 진균으로부터 유래되는 단백질 및 펩티드는 진균 항원일 수 있다.

[0021] 상기 진균 항원은 에피더모피톤 코쿠숨(*Epidermophyton koccusum*), 마이크로스포룸 아우도우이니(*Microsporium audouini*), 마이크로스포룸 카니스(*Microsporium canis*), 마이크로스포룸 디스토르툼(*Microsporium distortum*), 마이크로스포룸 에퀴눔(*Microsporium equinum*), 마이크로스포룸 기프숨(*Microsporium gypsum*), 마이크로스포룸 나눔(*Microsporium nanum*), 트리코피톤 콘센트리쿰(*Trichophyton concentricum*), 트리코피톤 에퀴눔(*Trichophyton equinum*), 트리코피톤 갈리나에(*Trichophyton gallinae*), 트리코피톤 기프세움(*Trichophyton gypsum*), 트리코피톤 메그니니(*Trichophyton megnini*), 트리코피톤 멘타그로피테스(*Trichophyton mentagrophytes*), 트리코피톤 퀴нке아눔(*Trichophyton quinckeanum*), 트리코피톤루브룸(*Trichophyton rubrum*), 트리코피톤 쇼엔레이니(*Trichophyton schoenleini*), 트리코피톤 톤수란스(*Trichophyton tonsurans*), 트리코피톤 베루코숨(*Trichophyton verrucosum*), T 베루코숨 var. album, var. 디스코이테스, var. 오크라세움, 트리코피톤 비올라세움(*Trichophyton violaceum*), 및/또는 트리코피톤 파비포르메(*Trichophyton faviforme*)를 포함하는 피부사상균(*Dermatophytes*); 또는 아스페르길루스 푸미가투스(*Aspergillus fumigatus*), 아스페르길루스 카부스(*Aspergillus kavus*), 아스페르길루스 나이거(*Aspergillus niger*), 아스페르길루스 니둘란스(*Aspergillus nidulans*), 아스페르길루스 테레우스(*Aspergillus terreus*), 아스페르길루스 시도위(*Aspergillus sydowi*), 아스페르길루스 카바투스(*Aspergillus kavatus*), 아스페르길루스 글라우쿠스(*Aspergillus glaucus*), 블라스토키조마이세스 카피타투스(*Blastoschizomyces capitatus*), 칸디다 알비칸스(*Candida albicans*), 칸디다 에놀라세(*Candida enolase*), 칸디다 트로피칼리스(*Candida tropicalis*), 칸디다 글라브라타(*Candida glabrata*), 칸디다 크루세이(*Candida krusei*), 칸디다 파라프실로시스(*Candida parapsilosis*), 칸디다 스텔라토이데아(*Candida stellatoidea*), 칸디다 쿠세이(*Candida kusei*), 칸디다 파라쿠세이(*Candida parakwsei*), 칸디다 루시타니아(*Candida lusitaniae*), 칸디다 슈도트로피칼리스(*Candida pseudotropicalis*), 칸디다 구일리에르몽디(*Candida guilliermondi*), 클라도스포리움 카리오니(*Cladosporium carrionii*), 코시디오이데스 이미티스(*Coccidioides immitis*), 블라스토마이세스 더마티디스(*Blastomyces dermatidis*), 크립토코커스 네오포르만스(*Cryptococcus neoformans*), 게오텐리쿰 클라바툼(*Geotrichum clavatum*), 히스토플라스마 카프슐라툼(*Histoplasma capsulatum*), 클레브시엘라 뉴모니아(*Klebsiella pneumoniae*), 마이크로스포리디아(*Microsporidia*), 엔세팔리토존 속(*Encephalitozoon spp.*), 세프타타 인테스티날리스(*Septata intestinalis*) 및 엔테로시토존 비에네우시(*Enterocytozoon bienewsi*); 브라키올라 속(*Brachiola spp.*), 마이크로스포룸 속(*Microsporidium spp.*), 노세마 속(*Nosema spp.*), 플레이스토포라 속(*Pleistophora spp.*), 트라키플레이스토포라 속(*Trachipleistophora spp.*), 비타포르마 속(*Vittaforma spp.*), 파라코시디오이데스 브라실리엔시스(*Paracoccidioides brasiliensis*), 뉴모시스티스 카리니(*Pneumocystis carinii*), 피티움 인시디오숨(*Pythium insidiosum*), 피티로스포룸 오발레(*Pityrosporum ovale*), 사카로마이세스 세레비지애(*Sacharomyces cerevisiae*), 사카로마이세

스 보울라르디아(*Saccharomyces boulardii*), 사카로마이세스 폼베(*Saccharomyces pombe*), 세도스포리움 아디오스페툼(*Scedosporium apiospermum*), 스포로트릭스 센키이(*Sporothrix schenckii*), 트리코스포론 베이젤리이(*Trichosporon beigelii*), 톡소플라스마 곤다이(*Toxoplasma gondii*), 페니실리움 마르네페이(*Penicillium marneffei*), 말라세지아 속(*Malassezia spp.*), 폰세카에 속(*Fonsecaea spp.*), 완지엘라 속(*Wangiella spp.*), 스포로트릭스 속(*Sporothrix spp.*), 바시디오볼루스 속(*Basidiobolus spp.*), 코니디오볼루스 속(*Conidiobolus spp.*), 리조푸스 속(*Rhizopus spp.*), 무코르 속(*Mucor spp.*), 아브시디아 속(*Absidia spp.*), 모르티에렐라 속(*Mortierella spp.*), 쿠닝가멜라 속(*Cunninghamella spp.*), 사크세나에 속(*Saksenaea spp.*), 알테르나리아 속(*Alternaria spp.*), 쿠르볼라리아 속(*Curvularia spp.*), 헬민토스포리움 속(*Helminthosporium spp.*), 푸사리움 속(*Fusarium spp.*), 아스페르길루스 속(*Aspergillus spp.*), 페니실리움 속(*Penicillium spp.*), 모놀리니아 속(*Monolinia spp.*), 리조코토니아 속(*Rhizoctonia spp.*), 패실로마이세스 속(*Paecilomyces spp.*), 피토마이세스 속(*Pithomyces spp.*), 및 클라도스포리움 속(*Cladosporium spp.*)으로 이루어진 군으로부터 선택되는 진균으로부터 유래될 수 있다.

- [0022] 원생동물로부터 유래되는 단백질 및 펩티드는 원생동물 항원일 수 있다.
- [0023] 상기 원생동물 항원은 엔타모에바 히스톨리티카(*Entamoeba histolytica*), 지아르디아 람블리(*Giardia lamblia*), 크립토스포리디움 파르븀(*Cryptosporidium parvum*), 시클로스포라 카야타넨시스(*Cyclospora cayatanensis*) 및 톡소플라스마(*Toxoplasma*)로 이루어진 군으로부터 선택되는 원생동물로부터 유래될 수 있다.
- [0024] 상기 치료 생체분자는 식물로부터 유래되는 단백질 및 펩티드일 수 있다. 바람직하게는, 상기 단백질 및 펩티드는 식물 항원이다. 상기 식물 항원은 리시누스 코무니스(*Ricinus communis*)로부터 유래될 수 있다.
- [0025] 다른 구현예에서, 상기 치료 생체분자는 면역원 또는 항원일 수 있다. 바람직하게는, 상기 면역원 또는 항원은 종양 면역원 또는 항원, 또는 암 면역원 또는 항원이다. 상기 종양 면역원 및 항원은 펩티드-함유 종양 항원, 예컨대 폴리펩티드 종양 항원 또는 당단백질 종양 항원일 수 있다.
- [0026] 상기 종양 항원은 (a) 암 세포와 연관된 전장(full length) 분자, (b) 결실, 부가 및/또는 치환된 부분을 갖는 분자를 포함하는 상동체(homolog) 및 그의 변형된 형태, 및 (c) 그의 단편일 수 있다.
- [0027] 적합한 종양 면역원은 CD8+ 림프구에 의해 인식되는 클래스(class) I-제한(restricted) 항원 또는 CD4+ 림프구에 의해 인식되는 클래스 II-제한 항원을 포함한다.
- [0028] 상기 종양 항원은 고환암, 흑색종, 폐암, 두경부암, NSCLC, 유방암, 위장암, 방광암, 결장직장암, 췌장암, 림프종, 백혈병, 신장암, 간암, 난소암, 위암 및 전립선암으로 이루어진 군으로부터 선택되는 암과 연관된 항원일 수 있다.
- [0029] 상기 종양 항원은 다음으로부터 선택될 수 있다:
- [0030] (a) 암-고환 항원, 예컨대 NY-ESO-I, SSX2, SCP1뿐만 아니라 RAGE, BAGE, GAGE 및 MAGE 패밀리의 폴리펩티드, 예를 들면, (예를 들면, 흑색종, 폐, 두경부, NSCLC, 유방, 위장, 및 방광 종양을 다루기 위해 사용될 수 있는) GAGE-1, GAGE-2, MAGE-1, MAGE-2, MAGE-3, MAGE-4, MAGE-5, MAGE-6, 및 MAGE-12;
- [0031] (b) 돌연변이된 항원, 예를 들면, (다양한 고형 종양, 예컨대, 결장직장암, 폐암, 두경부암과 연관되는) p53, (예컨대, 흑색종, 췌장암 및 결장직장암과 연관되는) p21/Ras, (예컨대, 흑색종과 연관되는) CDK4, (예컨대, 흑색종과 연관되는) MUM1, (예컨대, 두경부암과 연관되는) caspase-8, (예컨대, 방광암과 연관되는) CIA 0205, HLA-A2-R1701, (예컨대, 흑색종과 연관되는) 베타 카테닌, (예컨대, T-세포 비-호지킨 림프종과 연관되는) TCR, (예컨대, 만성 골수성 백혈병과 연관되는) BCR-abl, 트리오스포스페이트 아이소머라아제(triosephosphate isomerase), KIA 0205, CDC-27, 및 LDLR-FUT;
- [0032] (c) 과발현된 항원, 예를 들면, (예컨대, 결장직장암과 연관되는) 갈렉틴(Galectin) 4, (예컨대, 호지킨 질환과 연관되는) 갈렉틴 9, (예컨대, 만성 골수성 백혈병과 연관되는) 프로티나아제 3, (예컨대, 다양한 백혈병과 연관되는) WT 1, (예컨대, 신장암과 연관되는) 탄산 탈수효소(carbonic anhydrase), (예컨대, 폐암과 연관되는) 알도라아제(aldolase) A, (예컨대, 흑색종과 연관되는) PRAME, (예컨대, 유방암, 결장암, 폐암 및 난소암과 연관되는) HER-2/neu, (예컨대, 간암과 연관되는) 알파-페토단백질, (예컨대, 결장직장암과 연관되는) KSA, (예컨대, 췌장암 및 위암과 연관되는) 가스트린(gastrin), 텔로머라아제 촉매화 단백질, (예컨대, 유방암 및 난소암과 연관되는) MUC-I, (예컨대, 신장 세포 암종과 연관되는) G-250, (예컨대, 유방암, 결장암과 연관되는) p53, 및 (예컨대, 유방암, 폐암, 및 결장직장암과 같은 위장관의 암과 연관되는) 암태아성 항원;

- [0033] (d) 공유된 항원, 예를 들면, (예컨대, 흑색종과 연관되는) 흑색종-멜라닌세포 분화 항원, 예컨대 MART-1/멜란 (Melan) A, gp100, MC1R, 멜라닌세포-자극 호르몬 수용체, 티로시나아제, 티로시나아제 관련 단백질-1/TRP1 및 티로시나아제 관련 단백질-2/TRP2;
- [0034] (e) 예컨대, 전립선 암과 연관되는 전립선-연관 항원, 예컨대 PAP, PSA, PSMA, PSH-P1, PSM-P1, PSM-P2; 및/또는
- [0035] (f) (예를 들면, 골수종 및 B 세포 림프종과 연관되는) 면역글로불린 이디오타입(idiotypic).
- [0036] 상기 치료 생체분자는 진핵생물 폴리펩티드일 수 있다. 한 구현예에서, 상기 진핵생물 폴리펩티드는 포유동물 폴리펩티드이다. 상기 포유동물 폴리펩티드는 효소; 효소 억제제; 호르몬; 면역 시스템 단백질; 수용체; 결합 단백질; 전사 또는 번역 인자; 종양 성장 저해 단백질; 구조 단백질 및 혈액 단백질;로 이루어진 군으로부터 선택될 수 있다.
- [0037] 상기 효소는 키모신; 위 리파아제; 조직 플라스미노겐 활성화제; 스트렙토키나아제; 콜레스테롤 생합성 또는 분해성 스테로이드생성 효소; 키나아제; 포스포디에스테라아제; 메틸라아제; 탈-메틸라아제; 탈수소효소; 셀룰라아제; 프로테아제; 리파아제; 포스포리파아제; 아로마타아제; 시토크롬; 아데닐레이트 또는 구아닐레이트 시클라아제 및 뉴라미다아제;로 이루어진 군으로부터 선택될 수 있다.
- [0038] 상기 효소 억제제는 메탈로프로티나아제(TIMPS)의 조직 억제제일 수 있다. 상기 호르몬 성장 호르몬일 수 있다.
- [0039] 상기 면역 시스템 단백질은 사이토카인; 케모카인; 림포카인; 에리트로포이에틴; 인테그린; 아드레신; 셀렉틴; 귀소(homing) 수용체; T 세포 수용체 및 면역글로불린;으로 이루어진 군으로부터 선택될 수 있다.
- [0040] 상기 사이토카인은 인터루킨, 예를 들면 IL-2, IL-4 및/또는 IL-6, 콜로니 자극 인자(CSF), 과립구 콜로니 자극 인자(G-CSF), 과립구-대식세포 콜로니 자극 인자(GM-CSF) 또는 종양 괴사 인자(TNF)일 수 있다.
- [0041] 상기 케모카인은 대식세포 염증성 단백질-2 및/또는 플라스미노겐 활성화제일 수 있다.
- [0042] 상기 림포카인은 인터페론일 수 있다.
- [0043] 상기 면역글로불린은 자연형, 변형된 또는 키메라 면역글로불린 또는 그의 단편일 수 있다. 바람직하게는, 상기 면역글로불린은 이중 활성을 갖는 키메라 면역글로불린, 예컨대 항체 효소 또는 항체-독소 키메라이다.
- [0044] 상기 호르몬은 인슐린, 갑상선 호르몬, 카테콜아민, 고나도트로핀, 자극 호르몬, 프로락틴, 옥시토신, 도파민, 소 소마토트로핀, 렙틴; 성장 호르몬(예컨대, 인간 성장 호르몬), 성장 인자(예컨대, 표피 성장 인자, 신경 성장 인자, 인슐린-유사 성장 인자 등);로 이루어진 군으로부터 선택될 수 있다.
- [0045] 상기 수용체는 스테로이드 호르몬 수용체 또는 펩티드 수용체일 수 있다. 바람직하게는, 상기 수용체는 성장 인자 수용체이다.
- [0046] 상기 결합 단백질은 성장 인자 결합 단백질일 수 있다.
- [0047] 상기 종양 성장 저해 단백질은 혈관형성을 억제하는 단백질일 수 있다.
- [0048] 상기 구조 단백질은 콜라겐; 피브로인; 피브리노겐; 엘라스틴; 튜블린; 액틴; 및 미오신;으로 이루어진 군으로부터 선택될 수 있다.
- [0049] 상기 혈액 단백질은 트롬빈; 혈청 알부민; 인자 VII; 인자 VIII; 인슐린; 인자 IX; 인자 X; 조직 플라스미노겐 활성화제; 단백질 C; 본 빌레브란트 인자; 항-트롬빈 III; 글루코세레브로시다아제; 에리트로포이에틴 과립구 콜로니 자극 인자(G-CSF) 또는 변형된 인자 VIII; 및 항응고제;로 이루어진 군으로부터 선택될 수 있다.
- [0050] 한 바람직한 구현예에서, 상기 치료 생체분자는 림프 항상성을 조절할 수 있는 사이토카인, 바람직하게는 T 세포의 발생, 프라이밍, 확장, 분화 및/또는 생존에 수반되고, 바람직하게는 이를 유도 또는 향상시키는 사이토카인이다. 따라서, 바람직하게는, 상기 사이토카인은 인터루킨이다. 가장 바람직하게는, IL-2, IL-7, IL-12, IL-15, 또는 IL-21이다.
- [0051] 상기 치료 생체분자는 줄기 세포 특징을 갖는 세포로 체세포의 재프로그래밍(reprogramming)을 향상시킬 수 있는 단백질일 수 있다.
- [0052] 줄기 세포 특징을 갖는 세포로 체세포의 재프로그래밍을 향상시킬 수 있는 단백질은 OCT4, SOX2, NANOG, LIN28, p53, ART-4, BAGE, ss-카테닌/m, Bcr-abl CAMEL, CAP-1, CASP-8, CDC27/m, CD 4/m, CEA, CLAUDIN-12, c-MYC,

CT, Cyp-B, DAM, ELF2M, ETV6-AML1, G250, GAGE, GnT-V, Gap100, HAGE, HER-2/neu, HPV-E7, HPV-E6, HAST-2, hTERT(또는 hTRT), LAGE, LDLR/FUT, MAGE-A, MAGE-B, MAGE-C, MART-1/멜란-A, MC1R, 마이오신/m, MUC1, MUM-1, -2, -3, NA88-A, NF1, NY-ESO-1, NY-BR-1, p190 마이너 BCR-abL, Plac-1, Pml/RARa, PRAME, 프로티나아제 3, PSA, PSM, RAGE, RU1 또는 RU2, SAGE, SART-1 또는 SART-3, SCGB3A2, SCP1, SCP2, SCP3, SSX, SURVIVIN, TEL/AML1, TPI/m, TRP-1, TRP-2, TRP-2/INT2, TPTE 및 WT로 이루어진 군으로부터 선택될 수 있고, 바람직하게는 WT-1이다.

[0053] 바람직하게는, MAGE-A는 MAGE-A1, MAGE-A2, MAGE-A3, MAGE-A4, MAGE-A5, MAGE-A6, MAGE-A7, MAGE-A8, MAGE-A9, MAGE-A10, MAGE-A11, 또는 MAGE-A12로 이루어진 군으로부터 선택된다.

[0054] 바람직하게는, 줄기 세포 특징을 갖는 세포로 체세포의 재프로그래밍을 향상시킬 수 있는 단백질은 OCT4, SOX2, LF4; c-MYC; NANOG; LIN28;이다.

[0055] 상기 치료 생체분자는 세포-치료법 적응증을 위해 생체외(ex vivo)에서 세포를 변형하기 위해 이용될 수 있는 생체분자일 수 있다. 따라서, 바람직하게는 상기 치료 생체분자는 면역글로불린, T-세포 수용체 및 NK 수용체로 이루어진 군으로부터 선택될 수 있다.

[0056] 상기 치료 생체분자는 내인성(endogenous) 숙주 유전자의 발현을 조절할 수 있는 RNA 분자, 예를 들면 간접 RNA, 예컨대 작은 RNA, siRNA 또는 microRNA일 수 있다.

[0057] 바람직하게는, 상기 RNA 구조체는 선천적 면역 반응을 RNA로 감소 또는 차단할 수 있는 적어도 하나의 선천적 억제제 단백질(IIP)을 암호화하는 유전자를 포함한다. 선천적 면역 반응을 RNA로 감소 또는 차단하는 것은 바람직하게는 본 발명의 RNA 구조체를 갖고 있는 숙주 세포에 의한 RNA(바람직하게는 길이가 적어도 1 kb인 RNA를 의미하는 것으로 숙련된 기술자에 의해 이해될 수 있는) 긴 RNA 또는 dsRNA)의 인식을 감소 또는 차단함으로써 IIP에 의해 달성된다. 보다 바람직하게는, 상기 선천적 억제제 단백질은 선천적 억제 단백질이어서, 선천적 반응을 RNA, 바람직하게는 제1 측면의 RNA 구조체의 RNA로 감소 또는 차단할 수 있다. 상기 선천적 억제제 단백질은 인터페론 조절 인자 3 및 7(IRF3 및 IRF7) 및 NF- $\kappa$ B 전사 인자의 활성화를 유도하여 다양한 항-바이러스 유전자(예컨대, saRNA 발현을 저해하는 것으로 알려진 IFIT1-3, Mx1, Mx2), 그 생성물이 선천적 면역 반응을 조직하는 전염증성(proinflammatory) 유전자를 직접 촉발하고, 임의의 인터페론 의존성 캐스케이드 상류의 정규적(canonical) IFN-자극(stimulated) 유전자(ISGs)의 직접적인 활성화를 유도하는 패턴 인식 수용체에 의한 세포기질 saRNA의 인식을 감소 또는 방지할 수 있다. 상기 경로는 많은 항-바이러스 반응을 추가로 증폭하는 양성 피드백 루프(feedback loop)를 제공하는 타입 I & III 인터페론의 유도에 의해 향상될 수 있다.

[0058] 상기 RNA는 단일 가닥 RNA 또는 이중 가닥 RNA일 수 있다. 바람직하게는, 상기 RNA는 saRNA이다.

[0059] 적어도 하나의 선천적 억제제 단백질은 (i) 예를 들면 MDA5의 올리고머화 및 RNA에 대한 MDA5의 결합을 방지함으로써 흑색종 분화-연관 단백질 5(MDA5)의 작용을 감소 또는 차단하거나, 및/또는 (ii) RNA에 대한 PKR 활성화 단백질로도 불릴 수 있는 PACT의 결합을 차단 또는 감소시킬 수 있다. 숙련된 기술자는 상기 센서들이 신호를 전사 인자(NF- $\kappa$ B, IRF-3, -7)의 활성화를 포함하는 하류의 캐스케이드를 활성화시키는 하류의 미토콘드리아 어댑터인 미토콘드리아 항-바이러스 신호전달(MAVS)로 전달하기 위해 신호를 전송함을 이해할 것이다. 이것은 차례로 적절한 항-바이러스 신호전달 반응 및 saRNA 발현을 저해하는 것으로 알려진 IFIT1을 포함하는 항-바이러스 활성을 갖는 분자를 암호화하는 타입-I 인터페론 자극 유전자의 활성화를 유도한다.

[0060] MDA5의 작용을 차단하는 적어도 하나의 선천적 억제제 단백질은 파라믹소바이러스 V 단백질, 콕사키바이러스 A16, 콕사키바이러스 A6, 및 엔테로바이러스 D68 바이러스의 3C 단백질; 버나바이러스의 VP3 단백질; 돼지 델타 코로나바이러스의 약세서리 단백질 NS6; 및 뇌심근염 바이러스의 2C 단백질; 및 그의 동원체(orthologue);로 이루어진 군으로부터 선택될 수 있다.

[0061] 바람직하게는, MDA5의 작용을 차단하는 적어도 하나의 선천적 억제제 단백질은 파라믹소바이러스 V 단백질이다. 가장 바람직하게는, MDA5의 작용을 차단하는 적어도 하나의 선천적 억제제 단백질은 파라인플루엔자 바이러스 타입 5 V 단백질(PIV5 V)이다.

[0062] RNA에 대한 PACT의 결합을 차단 또는 감소시키는 적어도 하나의 선천적 억제제 단백질은 임의의 코로나바이러스의 ORF4a(NS4a), 임의의 코로나바이러스의 ORF3b, 또는 마우스 간염 바이러스 및 SARS(코로나바이러스)의 뉴클레오텍시드 단백질; 및 그의 동원체;로 이루어진 군으로부터 선택될 수 있다.

[0063] 바람직하게는, 상기 ORF4a(NS4a)는 중동 호흡기 증후군 코로나바이러스인 MERS 코로나바이러스(ORF4a)이다.

- [0064] 바람직하게는, 상기 코로나바이러스 ORF3b는 SARS-CoV2 ORF3b이다.
- [0065] 각각의 PIV5V, ORF4a 및 ORF3b에 대한 단백질, DNA 및 RNA 서열은 아래에 제공된다.
- [0066] 한 구현예에서, 상기 PIV5 V 폴리펩티드는 다음과 같은 서열번호 11로서 본 명세서에 제공된다:
- [0067] MDPTDLSFSPDEINKLIETGLNTVEYFTSQVVTGTSSLGKNTIPPGVTGLLTNAAEAKIQESTNHQKGSVGGGAKPKKPRPKIAIVPADDKTVPGKPIPNPL  
LGLDSTPSTQTVLDLSGKTLPSGSYKGVKLAKFGKENLMTRFIEEPRENPIATSSPIDFKRGRDTGGFHRREYSIGWVGDEVKVTWCNPSCSPITAAARRF  
ECTCHQCPVTCSECERDT [서열번호 11]
- [0068] 따라서, 바람직하게는 상기 PIV5 V 폴리펩티드는 실질적으로 서열번호 11에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 11을 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.
- [0069] 한 구현예에서, 상기 PIV5 V 폴리펩티드는 다음과 같은 서열번호 12의 뉴클레오티드 서열에 의해 암호화된다:
- [0070] ATGGACCCTACCGACCTGAGCTTCAGCCCCGACGAGATCAACAAGCTGATCGAGACAGGCCTGAACACCGTGGAATACTTCACCAGCCAGCAAGTGACCGGC  
ACAAGCAGCCTGGGCAAGAACAATTCCTCCAGGCGTGACCGGCTGCTGACAAATGCTGCCGAGGCCAAGATCCAAGAGAGCACCAACCACCAGAAGGGC  
TCTGTGTGAGGCGGAGCCAAGCCTAAGAAGCCGACCTAAGATCGCCATCGTGCCCGCCGACGATAAGACAGTGCCTGGCAAGCCCATTCCTAATCCTCTG  
CTGGGCTCGACAGACCCCTAGCACACAGACAGTGTGATCTGAGCGGCAAGACACTGCCTAGCGGCAGCTATAAGGGCGTGAAGCTGGCCAAGTTCGGC  
AAAGAAAACCTGATGACCCGTTTCATCGAGGAACCCAGAGAGAACCCCTATCGCCACCGAGTCTCCCATCGACTTCAAGAGAGGCAGAGACACCGCGCGGCTTC  
CACAGAAGAGAGTACAGCATTGGCTGGTTCGAGATGAAGTGAAAGTGACCGAGTGTGCAACCCGAGTGCAGCCCTATTACAGCCGCCGCTAGAAGATTC  
GAGTGACCTGTACCAGTGTCTGTGACCTGTAGCGAGTGCAGCGGGACACA [서열번호 12]
- [0071] 따라서, 바람직하게는 상기 PIV5 V 폴리펩티드는 실질적으로 서열번호 12에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.
- [0072] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 47의 RNA 뉴클레오티드 서열을 포함할 수 있다:
- [0073] AUGGACCCUACCGACCGAGCUUCAGCCCCGACGAGAUACAAGCUGAUCGAGACAGGCCUGAACACCGUGGAUACUUCACCAGCCAGCAAGUGACCGGC  
ACAAGCAGCCUGGCAAGAACAACAAUCCUCCAGGCGUGACCGGCCUGCUGACAAUUGCUGCCGAGGCCAAGAUCCAAGAGAGCACCAACCACCAGAAGGGC  
UCUGUUGGAGGCGGAGCCAAGCCUAGAAGCCGAGACCUAAGAUCGCCAUCGUGCCCGCCGACGAUAAGACAGUGCCUGGCAAGCCAUUCCUAAUCCUCUG  
CUGGGCCUCGACAGCACCCCUAGCACACAGACAGUGCUGGAUCUGAGCGGCAAGACACUGCCUAGCGGCAGCUAUAAGGGCGUGAAGCUGGCCAAGUUCGGC  
AAAGAAAACCUGAUGACCCGUUCAUCGAGGAACCCAGAGAGAACCCUUAUCGCCACCGAGCUCUCCUAGCAGCUAUAAGAGAGGCAGAGACACCGCGCGCUUC  
CACAGAAGAGAGUACAGCAUUGGCGGGUGGAGAUAGAUGAAAGUGACCGAGUGGUGCAACCCGAGCUGCAGCCCUAUAACAGCCGCCGCUAGAAGAUUC  
GAGUGCACCUGUACCCAGUGUCCUGUGACCUAGCGAGUGCGAGCGGGACACA [서열번호 47]
- [0074] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 47로 나타난 것과 같은 RNA 뉴클레오티드 서열 또는 그의 변이체 또는 단편을 포함한다.
- [0075] 한 구현예에서, 상기 MERS-CoV ORF4a 폴리펩티드는 다음과 같은 서열번호 15로서 본 명세서에 제공된다:
- [0076] MDVYSLNLIWQKYLNSPYTTCCLYIPKPTAKYTPLVGTSLHPVLWNCQLSFAGYTESAVNSTKALAKQDAAQRIAWLLHKDGGIPDGC SLYL RHSSLFAQSE  
EEESFSN [서열번호 15]
- [0077] 따라서, 바람직하게는 상기 MERS-CoV ORF4a 폴리펩티드는 실질적으로 서열번호 15에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 15를 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.
- [0078] 한 구현예에서, 상기 MERS-CoV ORF4a 폴리펩티드는 다음과 같은 서열번호 16의 뉴클레오티드 서열에 의해 암호화된다:
- [0079] ATGGA CTACGTGCTCCCTGCTGAACAGATTTGGCAGAAGTACCTGAACAGCCCTACACCACCTGTCTGTACATCCCCAAGCCTACCGCAAGTACACACCT  
CTCGTGGGCACATCTCTGCACCCGCTGCTGTGAATTGCCAGCTGAGCTTTGCCGGCTACACCGAGTCTGCCGTGAACAGCACAAGGCCCTGGCCAAACAG  
GACGCCGCTCAGAGAATTGCCTGGTGTCTGCACAAGGATGGCGGCATCCCTGATGGTGTAGCCTGTACCTGAGACACAGCAGCCTGTTGCCAGAGCGAG  
GAAGAGGAATCCTTCAGCAAC [서열번호 16]
- [0080] 따라서, 바람직하게는 상기 MERS-CoV ORF4a 폴리펩티드는 실질적으로 서열번호 16에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

- [0081] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 48의 RNA 뉴클레오티드 서열을 포함할 수 있다:
- [0082] AUGGACUACGUGUCCUGUGAACCAGAUUUGGCAGAGUACCUGAACAGCCCCUACACCACCUGUCUGUACAUCCCCAAGCCUACCGCCAAGUACACACCU CUCGUGGGCACAUUCUGCACCCTGUGUGGAAUUGCCAGCUGAGCUUUGCCGGCUACACCGAGUCUGCCGUGAACAGCACAAAGGCCUGGCCAAACAG GACGCGCUCAGAGAAUUGCCUGGCUGCUGCACAAGGAUGGCGCAUCCUGAUGGCUGUAGCCUGUACCUGAGACACAGCAGCCUGUUCGCCAGAGCGAG GAAGAGGAUCCUUCAGCAAC [서열번호 48]
- [0083] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 48로 나타낸 것과 같은 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.
- [0084] 한 구현예에서, 상기 SARS-CoV-2 ORF3b 폴리펩티드는 다음과 같은 서열번호 20으로서 본 명세서에 제공된다:
- [0085] MMPTIFFAGILIVTTIVYLTIVQLQLSLLQVMAQVFLNMTTRLVVLKNGNLEQKTVLYYTVTSLQTITSTCTQLN [서열번호 20]
- [0086] 따라서, 바람직하게는 상기 SARS-CoV-2 ORF3b 폴리펩티드는 실질적으로 서열번호 20에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 20을 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.
- [0087] 한 구현예에서, 상기 SARS-CoV-2 ORF3b 폴리펩티드는 다음과 같은 서열번호 55의 뉴클레오티드 서열(Wuhan-Hu-1 접근 번호 NC\_045512.2; 뉴클레오티드 25814-26050)에 의해 암호화된다:
- [0088] ATGATGCCAACTATTTTCTTGTGCTGCATACCTAATGTTAGCTATTGTATACCTTACAATAGTGCAACTTCTTCAATTGTCATTACTTCAGGTGATGGCA CAACAAGTCCTATTTCTGAACATGACTACCAGATTGGTGGTTATACTGAAAAATGGGAATCTGGAGCAAAAGACTGTGTTGTATTACAGTTACTTCACTT CAGACTATTACCAGCTGTACTCAACTCAATTGA [서열번호 55]
- [0089] 따라서, 바람직하게는 상기 SARS-CoV-2 ORF3b 폴리펩티드는 실질적으로 서열번호 55에 나타낸 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.
- [0090] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 56의 RNA 뉴클레오티드 서열을 포함할 수 있다:
- [0091] AUGAUGCCAACUAAUUUCUUGCUGGCAUACUAAUUGUACGACUAAUUGUAAUACCUUACAAUAGUGCAACUUCUCAAUUGUCAUUACUUCAGGUGAUGGCA CAACAAGUCCUAAUUUCUGAACAUAGACUACCAGAUUGGUGGUUUAUACUGAAAAUUGGGAUUCUGGAGCAAAAGACUGUGUUGUAAUACAGUUACUUCACUU CAGACUAAUACCAGCUGUACUACUACUAAUUGA [서열번호 56]
- [0092] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 56으로 나타낸 것과 같은 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.
- [0093] 다른 구현예에서, 적어도 하나의 선천적 억제제 단백질은 MDA5 활성화의 하류 경로를 억제하거나 dsRNA의 MDA/PACT인식의 하류 경로를 차단할 수 있다.
- [0094] MDA5 활성화의 하류 경로를 억제하거나 dsRNA의 MDA/PACT 인식의 하류 경로를 차단할 수 있는 적어도 하나의 선천적 억제제 단백질은 HSV-2 Us1; HSV-1 Us1; HSV-1 Us11; OV20.0L; BVDV Npro; 란가트(Langat) 바이러스 NS5; 및 인플루엔자 NS1;로 이루어진 군으로부터 선택될 수 있다.
- [0095] 숙련된 기술자는 MDA5 활성화의 하류 경로를 억제하거나 dsRNA의 MDA/PACT인식의 하류 경로를 차단할 수 있는 적어도 하나의 선천적 억제제 단백질이 MDA5의 작용을 차단하는 적어도 하나의 선천적 억제제 단백질 및/또는 RNA에 대한 PACT의 결합을 차단 또는 감소시키는 적어도 하나의 선천적 억제제 단백질과 조합하여 사용될 수 있음을 이해할 것이다.
- [0096] 한 구현예에서, 상기 HSV-2 Us1 폴리펩티드는 다음과 같은 서열번호 1로서 본 명세서에 제공된다:
- [0097] VRDCYLMGYCRTRLGPRTWGRLLQISGGTWDVRLRNAIREVEAHFEPAAEPVCELPCLNARRYGPECVDGNLETNGGSTSDDEISDATSDDDLASHSDTEG GPSPAGRENPEASGGAIAARLECEFGTFDWTSEEGSQPWL.SAVVADTSSAERSGLPAPGACRATEAPEREDGCRKMRFPACPYPCGHTFLRP [서열번호 1]
- [0098] 따라서, 바람직하게는 상기 HSV-2 Us1 폴리펩티드는 실질적으로 서열번호 1에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 1을 암호화하는 RNA 뉴클레오티드 서열 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.
- [0099] 한 구현예에서, 상기 HSV-2 Us1 폴리펩티드는 다음과 같은 서열번호 2의 뉴클레오티드 서열에 의해 암호화된다:
- [0100] GTGAGAGACTGCTACCTGATGGGCTACTGCCGACAAGACTGGGACCTAGAACATGGGGCAGACTGCTGCAGATCAGCGGCGGAACATGGGATGTGCGGCTG

AGAAACGCCATCAGAGAGGTGGAAGCCACTTCGAGCCTGCCGCTGAACCTGTGTGTGAACCTGCCCTGTCTGAACGCTAGAAGATACGGCCCTGAGTGCAGAC  
GTGGGCAACCTGGAACAAATGGCGGCAGCACCAGCGACGACGAGATTTCGATGCCACCGACAGCGACGATACACTGGCCAGCCACAGCGATACAGAAGGC  
GGACCATCTCTGCCGAAGAGAGAATCCTGAGTCTGCCTCTGGCGGAGCCATTGCCGCTAGACTGGAATGCGAGTTCGGCACCTTCGACTGGACAAGCGAG  
GAAGGCTCTCAGCCTTGCTGTCTGTGTGGTGGCCGATACAAGCTCTGCCGAGAGAAGTGGACTTCCTGCTCCTGGCGCCTGTAGAGTACAGAGGCTCTT  
GAAAGAGAGGACGGCTGCAGAAAGATGCGGTTCCTTGCCGCTGTCTTATCCTTGCGGCCACACATTTCTGCGGCC [서열번호 2]

[0101] 따라서, 바람직하게는 상기 HSV-2 Us1 폴리펩티드는 실질적으로 서열번호 2에 나타낸 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0102] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 42의 RNA 뉴클레오티드 서열을 포함할 수 있다:

[0103] GUGAGAGACUGCUACCUGAUGGGCUACUGCCGACAAGACUGGGACCUAGAACAUGGGGACAGACUGCUGCAGAUACAGCGCGGAACAUGGGAUGUGCGGCUG  
AGAAACGCCAUCAGAGAGGUGGAAGCCACUUCGAGCCUGCCGUGAACUGUGUGUAACUGCCUGUCUGAACGCUAGAAGAUACGGCCUGAGUGCGAC  
GUGGGCAACCUGGAACAAAUGGCGGCAGCACCAGCGACGACGAGAUUUCGAGUCCACCGACAGCGACGAUACUGGCCAGCCACAGCGAUACAGAAGGC  
GGACCAUCCUGCCGAAGAGAGAAUCCUGAGUCUGCCUCUGCGGAGCCAUUGCCGCUAGACUGGAUUGCGAGUUCGGCACCUUCGACUGGACAAGCGAG  
GAAGGCUUCAGCCUUGGUGUCUGUGUGGCGCAUACAAGCUCUGCCGAGAGAAGUGGACUUCUGCUCCUGGCCGCGUGUAGAGCUACAGAGGCUCCU  
GAAAGAGAGGACGGCUGCAGAAAGAUUGCGGUUCCUGCCGCGUGUCCUUAUCCUUGCGGCCACACAUUUCUGCGGCC [서열번호 42]

[0104] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 42로 나타낸 것과 같은 RNA 뉴클레오티드 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

[0105] 한 구현예에서, 상기 HSV-1 Us1 폴리펩티드는 다음과 같은 서열번호 3으로서 본 명세서에 제공된다:

[0106] VRDYLGMGCRTRLGPRTWGRLLQISGGTWDVRLRNAIREVEAHFEPAAEPVCELPCLNARRYGPECDVGNLETNGGSTDSDEISDATSDDTLASHSDTEG  
GPSPAGRENPEASGGAIAARLECEFGTFDWTSEEGSQPWL SAVVADIRDYLGMGCRARLAPRTWCRLQVSGGTWGMHLRNTIREVEARFDATAEPVCKL  
PCLETRRYGPECDSLNLIEHLSATSDDEISDATDLEAAGSDHTLASQSDTEDAPSPVTLETPEPRGSLAVRLEDEFGEFDWTPQEGSQPWL SAVVADTSSVE  
RPGPSDSGAGRAAEDRCKLDGRCKMRFSTACPYP CSDTFLRPTSSAERSGLPAPGACRATEAPEREDGCRKMRFPAACPYPGHTFLRP [서열번호 3]

[0107] 따라서, 바람직하게는 상기 HSV-1 Us1 폴리펩티드는 실질적으로 서열번호 3에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 3을 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

[0108] 한 구현예에서, 상기 HSV-1 Us1 폴리펩티드는 다음과 같은 서열번호 4의 뉴클레오티드 서열에 의해 암호화된다:

[0109] ATCAGAGACTGCTACCTGATGGGCTACTGCCGGCTAGACTGGCCCTAGAACATGGTGCAGACTGCTGCAAGTGTCTGGCGGCACATGGGGCATGCACCTG  
AGAAACACCATCAGAGAGGTGGAAGCCAGATTTCAGCGCCACAGCCGAGCCTGTGTGCAAGCTGCCTTGTCTGGAACCTCGGAGATACGGCCCCAGTGCAGAC  
CTGAGCAATCTGGAATTCACCTGAGCGCCACCAGCGACGACGAGATTCTGATGCCACCGACCTGGAAGCCCGGATCTGATCATACACTGGCCAGCCAG  
AGCGACACCGAGGATGCTCCATCTCCAGTGACTCTGGAACCCCTGAGCCTAGAGGATCTCTGGCCGTGCGACTGGAAGATGAGTTGCGCGAGTTCGACTGG  
ACCCCTCAAGAGGGATCTCAGCCTTGCTGTCTGCGTGGTGGCCGATACAAGCAGCGTGGAAGACCCGAGCCTAGCGATTCTGTTGCTGGCAGAGCCGCC  
GAGGATAGAAAGTGCTGGATGGCTGCGGAAGATGCGGTTCTCTACCGCCTGTCCATATCCTTGCGAGCGACACCTTCTGCGGCCT [서열번호 4]

[0110] 따라서, 바람직하게는 상기 HSV-1 Us1 폴리펩티드는 실질적으로 서열번호 4에 나타낸 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0111] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 43의 RNA 뉴클레오티드 서열을 포함할 수 있다:

[0112] GUGAGAGACUGCUACCUGAUGGGCUACUGCCGACAAGACUGGGACCUAGAACAUGGGGACAGACUGCUGCAGAUACAGCGCGGAACAUGGGAUGUGCGGCUG  
AGAAACGCCAUCAGAGAGGUGGAAGCCACUUCGAGCCUGCCGUGAACUGUGUGUAACUGCCUGUCUGAACGCUAGAAGAUACGGCCUGAGUGCGAC  
GUGGGCAACCUGGAACAAAUGGCGGCAGCACCAGCGACGACGAGAUUUCGAGUCCACCGACAGCGACGAUACUGGCCAGCCACAGCGAUACAGAAGGC  
GGACCAUCCUGCCGAAGAGAGAAUCCUGAGUCUGCCUCUGCGGAGCCAUUGCCGCUAGACUGGAUUGCGAGUUCGGCACCUUCGACUGGACAAGCGAG  
GAAGGCUUCAGCCUUGGUGUCUGUGUGGCGCAUACAAGAGACUGCUACCUGAUGGGCUACUGCCGGGCUAGACUGGCCCCUAGAACAUGGUGCAGA  
CUGCUGCAAGUGUCUGGCGGCACAUGGGGCAUGCACCUGAGAAACACCAUCAGAGAGGUGGAAGCCAGAUUCGACGCCACAGCCGAGCCUGUGCAAGCUG  
CCUUGUCUGGAAACUCGAGAUACGCCCCGAGUGCGACCUAGACAAUCUGGAAAUACCUAGAGCGCCACCAGCGACGACGAGAUUUCUGAUGCCACCGAC  
CUGGAAGCCGCCGAUCUGAUACAUCACUGGCCAGCCAGAGCGACACCGAGGAUGCUCAUUCUCCAGUGACUCUGGAAACCCUGAGCCUAGAGGAUCUCUG  
GCCGUGCGACUGGAAGAUAGAUUCGGCGAGUUCGACUGGACCCUCAAGAGGGAUCUCAGCCUUGGCUGUCUGCCGUGGUGGCCGAUACAAGCAGCGUGGAA  
AGACCCGACCUAGCGAUUCUGGUGUGGACAGCCGCCGAGGAUAGAAAGUGCCUGGAUGGCUGCCGGAAGAUUGCGGUUCUCUACCGCCUGUCCAUAUCCU  
UGCAGCGACACCUUCCUGCGGCCUGAUAAACAAGCUCUGCCGAGAGAAGUGGACUUCUGCUCCUGGCCGCCUGUAGAGCUACAGAGGCCUCCUGAAAGAGAG

GACGGCUGCAGAAAGAUGCGGUUCCUGCCGCCUGUCCUUAUCCUUGCGGCCACACAUUCUGCGGCC [서열번호 43]

[0113] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 43으로 나타낸 것과 같은 RNA 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0114] 한 구현예에서, 상기 HSV-1 Us11 폴리펩티드는 다음과 같은 서열번호 5로서 본 명세서에 제공된다:

[0115] MSQTQPPAPVGPDPDVLKGVPSAGMHPRGVHAPRGHPRMISGPPQQRGNDQAAGQCGDSGLLRVGADTTISKPSSEAVRPPTIPRTPRVPREPRVPRPPREPREPRVPRAPRDPVRPRDPRDPRQPRSPREPRSPREPRSPREPTPTPTPREPTARGSV [서열번호 5]

[0116] 따라서, 바람직하게는 상기 HSV-1 Us11 폴리펩티드는 실질적으로 서열번호 5에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 5를 암호화하는 RNA 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0117] 한 구현예에서, 상기 HSV-1 Us11 폴리펩티드는 다음과 같은 서열번호 6의 뉴클레오타이드 서열에 의해 암호화된다:

[0118] ATGAGCCAGACACAGCCTCCAGCTCCAGTTGGACCTGGCGACCTGATGTGTATCTGAAGGGCGTGCCAAGCGCCGGCATGCATCCTAGAGGTGTTCATGCCCTAGAGGACACCCAGAATGATCTCTGGCCCTCCTCAGAGAGGCGACAACGATCAGGCTGCTGGACAGTGTGGCGATAGCGGACTGCTGAGAGTGGGCGCCGATACCACAATCAGCAAGCCATCTGAGGCTGTGCGGCCTCTACAATCCCCAGAACCTAGAGTGCCCCGCGAGCCAAGAGTGCCTAGACCTCCTAGAGAGCCCAGAGAACCAGAGTGCCAAGGGCTCCAGAGATCCTAGAGTCCCTCGGGACCTAGGGACCCAAGACAACCTAGATCACCCAGAGAGCCTCGGAGCCCAAGAGAGCCAAGAAGCCCTAGGGAACCCGCGACCAAGAACACCCAGGGAACCTAGAACCAGGAGGCGAGCGTG [서열번호 6]

[0119] 따라서, 바람직하게는 상기 HSV-1 Us11 폴리펩티드는 실질적으로 서열번호 6에 나타낸 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0120] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 44의 RNA 뉴클레오타이드 서열을 포함할 수 있다:

[0121] AUGAGCCAGACACAGCCUCCAGCUCCAGUUGGACCUGGGCAGCCUGAUGUGUAUCUGAAGGGCGUGCCAAGCGCCGGCAUGCAUCCUAGAGGUGUUAUGCCCUAGAGGACACCCAGAAUGAUCUCUGGCCUCCUCAGAGAGGCGACAACGAUCAGGCUGCUGGACAGUGUGGCGAUAGCGGACUGCUGAGAGUGGGCGCCGAUACCAAAUCAGCAAGCAUCUGAGGCGUGCGGCCUCCUACAAUCCCCAGAACACCUAGAGUCCCCGCGAGCCAAGAGUGCCUAGACCUCCUAGAGAGCCCAGAGAACCAGAGUGCCAAGGGCUCCAGAGAUCCUAGAGUCCUCGGGACCCUAGGGACCCAAGACAACCUAGAUACCCAGAGAGCCUCGGAGCCCAAGAGAGCCAAGAAGCCUAGGGAACCCGCGACCAAGAACACCCAGGGAACCUAGAACCAGGAGGCGAGUG [서열번호 44]

[0122] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 44로 나타낸 것과 같은 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0123] 한 구현예에서, 상기 OV20.0L 폴리펩티드는 다음과 같은 서열번호 7로서 본 명세서에 제공된다:

[0124] MACECASLILELLRKSDDKLPAKQIAKELGISKHEANRQLYRLDSDEVCCEDGNPPRWFVECAPSAPTEEDENSDETPMETEAGCDTLFGGDIDIMTQSAVIRLKSINPVSAVNEFCMMTHRPLEFCETRAGGEDHCPRTCTITISGKVVAADGASKKLARHTACSSALTILINNCGISF [서열번호 7]

[0125] 따라서, 바람직하게는 상기 OV20.0L 폴리펩티드는 실질적으로 서열번호 7에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 7을 암호화하는 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0126] 한 구현예에서, 상기 OV20.0L 폴리펩티드는 다음과 같은 서열번호 8의 뉴클레오타이드 서열에 의해 암호화된다:

[0127] ATGGCCTGTGAATGCGCCAGCCTGATCCTGGAAGTCTGAGAAAGAGCGACGACAAGCTGCCCCCAAGCAGATCGCCAAAGAGCTGGGCATCTCTAAGCACGAGGCCAACCAGGCGAGCTGTACCGGCTGCTGGATTCTGACGAAGTGTGCTGCGAGGACGGCAATCCTCCTGTTGGTTCGTGGAATGTGCCCTAGCGCTCCACCGAAGAGGACGAGAATAGCGACACCGAGCCTATGGAACCGAGGCGCGGTGCGATACACTGTTTGGCGGAGACATCGACATCATGACCCAGAGCGCCGTGATCCGGCTGAAGTCCCTGAATCCTGTGTCCGCCGTGAACGAGTCTGCATGATGACCCACCGGCCTCTGGAATTTTGGCAGACAAGAGCCGGCGGAGAGGATCACTGCCCCAGATTACCTGTACCATCACCATCAGCGGCAAGGTGGTGGCTGTTGCCGATGGCGCCTCTAAGAACTGGCCAGACACCGCCTGTAGCAGCGCCCTGACAATCCTGATCAACAACCTGCGGCATCAGCTTC [서열번호 8]

[0128] 따라서, 바람직하게는 상기 OV20.0L 폴리펩티드는 실질적으로 서열번호 8에 나타낸 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0129] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 45의 RNA 뉴클레오타이드 서열을 포함할 수 있다:

[0130] AUGGCCUGUGAAUGCGCCAGCCUGAUCCUGGAACUGCUGAGAAAGAGCGACGACAAGCUGCCCGCAAGCAGAUCCGCAAGAGCUGGGCAUCUCUAAGCAC

GAGGCCAACCGGCAGCUGUACCGGCUGCUGGAUUCUGACGAAGUGUGCUGCGAGGACGGCAAUCCUCCUGUUGGUUCUGUGGAAUGUGCCCCUAGCGCUCCC  
ACCGAAGAGGACGAGAAUAGCGACACCGAGCCUUAUGGAAACCGAGGCCGCGCAUACACUGUUUGCGGAGACAUCGACAUCAGACCCAGAGCGCCGUG  
AUCCGGCUGAAGUCCUGAAUCCUGUGUCCGCCUGAACGAGUUCUGCAUGAUGACCCACCGGCCUCUGGAAUUUGCGAGACAAGAGCCGGCGGAGAGGAU  
CACUGCCCCAGAUUACCCUGUACCAUCACCAUCAGCGGCAAGGUGGUGGUGUCCGAUGGCGCCUCUAGAAGAACUGGCCAGACACCGCCUGUAGCAGC  
GCCCUGACAAUCCUGAUCAACAACUGCGGCAUCAGCUUC [서열번호 45]

[0131] 따라서, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 45로 나타낸 것과 같은 RNA 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0132] 한 구현예에서, 상기 BVDV Npro 폴리펩티드는 다음과 같은 서열번호 9로서 본 명세서에 제공된다:

[0133] MELITNELLYKTYKQKPVGVEEPVYDQAGDPLFGERGAVHPQSTLKLPHKRGEDVPTNLASLPKRGDCRTGNSRGPVSGIYLPKGPFLFYQDYKGPVYHRAP  
LELFEEGSMCETTKRIGRVTSDDGLYHIYVCIDGCI IKSATRSYQVRFRVHNRLDCPLWVTSC [서열번호 9]

[0134] 따라서, 바람직하게는 상기 BVDV Npro 폴리펩티드는 실질적으로 서열번호 9에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 9를 암호화하는 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0135] 한 구현예에서, 상기 BVDV Npro 폴리펩티드는 다음과 같은 서열번호 10의 뉴클레오타이드 서열에 의해 암호화된다:

[0136] ATGGAAGTATGATCACCAACGAGCTGCTGTACAAGACCTACAAGCAGAAACCCGTGGGCGTCGAGGAACCCGTGTATGATCAAGCTGGCGACCTCTGTTTGGC  
GAGAGAGGCGCTGTTACCCCTCAGAGCACACTGAAGCTGCCCAACAGCGGGCGAAAGAGATGTGCCTACCAACCTGGCCAGCCTGCCTAAGAGAGGCGAT  
TGCAGAACCGGCAATAGCAGAGGCCCTGTGTCCGGCATCTACCTGAAACCTGGACCACTGTTCTACCAGGACTACAAGGACCCGTGTACCACAGAGCCCT  
CTGGAAGTGTGTAAGAGGGCAGCATGTGCGAAACCACCAAGCGGATCGGAAGAGTGACCGGCTCTGACGGCAAGCTGTACCACATCTACGTGTGCATCGAC  
GGCTGCATCATCATCAAGAGCGCCACCAGATCTACCAGCGGTGTTTCAGATGGGTGCACAACAGACTGGACTGCCCTCTGTGGGTACCAGCTGC [서열  
번호 10]

[0137] 따라서, 바람직하게는 상기 BVDV Npro 폴리펩티드는 실질적으로 서열번호 10에 나타낸 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0138] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 46의 RNA 뉴클레오타이드 서열을 포함할 수 있다:

[0139] AUGGAACUGAUCACCAACGAGCUGCUGUACAAGACCUACAAGCAGAAACCCGUGGGCGUCGAGGAACCCGUGUAUGAUGAAGCUGGCGACCCUCUGUUUGGC  
GAGAGAGGCGCUGUUCACCCUCAGAGCACACUGAAGCUGCCCCACAAGCGGGCGAAAGAGAUGUGCCUACCAACCUGGCCAGCCUGCCUAGAGAGGCGAU  
UGCAGAACCGGCAUAGCAGAGGCCUGUGUCCGGCAUCUACCUAGAAACCGGACCAUGUUCUACCAGGACUACAAGGACCCGUGUACCACAGAGCCCU  
CUGGAACUGUUUAGAGGGCAGCAUGUGCGAAACCACCAAGCGGAUCGGAAGAGUGACCGGCUCUGACGGCAAGCUGUACCAUACUACUGUGCAUCGAC  
GGCUGCAUCAUCAAGAGCGCCACCAGAUCCUACCAGCGGGUGUUCAGAUUGGUGCACAACAGACUGGACUGCCUCUGUGGGUACACAGCUGC [서열  
번호 46]

[0140] 따라서, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 46으로 나타낸 것과 같은 RNA 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0141] 한 구현예에서, 상기 란가트 NS5 폴리펩티드는 다음과 같은 서열번호 17로서 본 명세서에 제공된다:

[0142] VFCDKVDTKAQEPQPGTKIIMRAVNDWLLERLVKKSRRMCSREEFIKVRNSNAALGAWSDENKWSAREAVEDPEFWSLVEAERERHLQGRCAHCYNNM  
GKREKKLGEFGVAKGSRAIWMWLGSRFLEFEALGFLNEDHWASRASSGAGVEGISLNYLWHLKKLASLSGGLFYADDTAGWDTKITNADLDDEEQILRYM  
DGDHKKLAATVLRKAYHAKVVRVARPSREGGCVMDIITRRDQRSGQVVTYALNITNKKVQLVRMMEGEGVIEVADSHNPRLLRVEKWLEEHGEERLSRML  
VSGDDCVVRPVDDRFSKALYFLNDMAKTRKDTGEWEPSTGFASWEEVPFCSHHFHELVMDGRALVVPDRDQDEL [서열번호 17]

[0143] 따라서, 바람직하게는 상기 란가트 NS5 폴리펩티드는 실질적으로 서열번호 17에 나타낸 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 17을 암호화하는 RNA 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0144] 한 구현예에서, 상기 란가트 NS5 폴리펩티드는 다음과 같은 서열번호 18의 뉴클레오타이드 서열에 의해 암호화된다:

[0145] GTGTTCAAGGACAAGGTGGACACCAAGGCTCAAGAGCCTCAGCCTGGCACCAAGATCATCATGAGAGCCGTGAACGACTGGCTGCTGGAACGGCTGGTCAAG  
AAAAGCAGACCCCGATGTGCAGCCGGGAAGAGTTTATCGCAAAGTGCGGAGCAATGCCGTCTCGGAGCTTGGAGTGACGAGCAGAACAAGTGAAGTCC

GCCAGAGAAGCCGTGGAAGATCCCGAGTTTGGAGCCTGGTGGAGCCGAGAGAGAGAGGCATCTGCAGGGAAGATGTGCCACTGCGTGTACAACATGATG  
GGCAAGAGAGAGAAGAAGCTGGGCGAGTTCCGGAGTGGCCAAAGGCAGCAGAGCCATCTGGTATATGTGGCTGGGCAGCCGCTTCTGGAATTTGAGGCCCTG  
GGCTTCTGAACGAGGATCACTGGGCTAGCAGAGCCTCTTCTGGTGCTGGCGTGAAGGCATCAGCCTGAATTATCTCGGCTGGCACCTGAAGAACTGGCC  
TCTCTGTCTGGCGCCTGTTCTACGCCGATGATACAGCCGATGGGACACAAAGATCACCAACGCCGACCTGGACGACGAGGAACAGATCCTGAGATATATG  
GACGGCGACCACAAAAGCTGGCCGCCACCGTGCTGAGAAAAGCCTATCACGCCAAGGTCGTGAGAGTGCCAGACCTAGTAGAGAAGGCGGCTGCGTGATG  
GACATCATCACCAAGGACCAGCGCGGCTCTGGCCAGGTTGTGACATACGCCCTGAACACCATCACCAACATCAAGGTGCAGCTCGTGCGGATGATGGAA  
GGCGAGGGCGTGATCGAAGTGGCCGACAGCCATAATCCTCGGCTGTGAGAGTGGAAGAGTGCTGGAAGAACCGGCGAAGAACGGCTGAGCAGAATGCTG  
GTGTCCGGCGACGATTGTGTTGTGCGGCCCTGGACGACAGATTGAGCAAGGCCCTGTACTTTCTGAATGACATGGCCAAGACCAGAAAGGACACCGGCGAG  
TGGGAGCCTTCTACAGGCTTTGCCAGCTGGGAAGAAGTGCCTTTCTGCAGCCACCACTTCCACGAGCTGGTCATGAAGGATGGCAGAGCCCTGGTGGTGCCC  
TGCAGAGATCAGGACGAACCTG [서열번호 18]

[0146] 따라서, 바람직하게는 상기 란가트 NS5 폴리펩티드는 실질적으로 서열번호 18에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0147] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 49의 RNA 뉴클레오티드 서열을 포함할 수 있다:

[0148] GUGUUC AAGGACAAGGUGGACACCAAGGCUCAAGAGCCUCAGCCUGGCACCAAGAUCAUCAUGAGAGCCUGAACGACUGGUCUGGAAACGGCUGGUAAG  
AAAAGCAGACCCCGAUGUGCAGCCGGGAAGAGUUUAUCGCCAAAGUGCGGAGCAUUGCCGUCUCGAGCUUGGAGUGACGAGCAGAACAAGUGGAAGUCC  
GCCAGAGAAGCCGUGGAAGAUCCGAGUUUUGGAGCCUGGUGGAAGCCGAGAGAGAGAGGCAUCUGCAGGGAAGAUUGGCCACUGCGUGUACAACAUGAUG  
GGCAAGAGAGAGAAGAAGCUGGGCGAGUUCGAGUGGCCAAAGGCAGCAGAGCCAUCUGGUUAUUGUGGUGGCGAGCCGCUUCCUGGAUUUAGGCCUG  
GGCUUCCUGAACGAGGAUCACUGGGCUAGCAGAGCCUUCUUGGUGUGGCGUGGAAGGCAUCAGCCUGAAUUAUCUGCGUGGACCCUGAAGAACUGGCC  
UCUCUGUCUGGGCGCCUGUUCACGCCGAUGAUACAGCCGAUGGGACACAAAGAUCAACACGCCGACCGUGGACGACGAGGAACAGAUCCUGAGAUUAUG  
GACGGCGACCACAAAAGCUGGCCGCCACCGUGCUGAGAAAAGGCCUAUCACGCCAAGGUCUGCAGAGUGGCCAGACCUAGUAGAGAAGGGCGCUGCGUGAUG  
GACAUCAUCACCAGAAGGACCAGCGCGGCUCUGGCCAGGUUGAGCAUACGCCUGAACACCAUCACCAACAUAAGGUGCAGCUCUGCGGAUGAUGGAA  
GGCGAGGGCGUGAUCGAAGUGGCGACGCCAUAAUCCUGGUGCUGAGAGUGGAAGAGUGGUGGAAGAACCGGCGAAGAACGGCUGAGCAGAAUGCUG  
GUGUCCGGCGACGAUUGUGUGUGGCGCCCGUGGACGACAGAUUCAGCAAGGCCUGUACUUUCUGAAUGACAUGGCCAAGACCAGAAAGGACACCGGCGAG  
UGGAGGCCUUCUACAGGCUUUGCCAGUGGAAGAAGUGCCUUCUGCAGCCACCACUCCACGAGCUGGUGAUGAAGGAUGGCAGAGCCUGGUGGUGCCC  
UGCAGAGAUACAGGACGAACUG [서열번호 49]

[0149] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 49로 나타난 것과 같은 RNA 뉴클레오티드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0150] 한 구현예에서, 상기 인플루엔자 NS1 폴리펩티드(접근 번호 DQ508893)는 다음과 같은 서열번호 13으로서 본 명세서에 제공된다:

[0151] MDSNTVSSSFQVDCFLWHVRKQVADQELGDAPFLDRLRRDQKSLKRGSTLGLNIETATCVGKQIVERILKEESDEAFRMTMASALASRYLTDMTIEEMSRDW  
FMLMPKQKVAGPLCVRMDQAIMDKNIILKANFSVIFDRLETLTLRAFTEEGAIVGEISPLPSLPGHNTEDVKNAIGVLIGGLEWNDNTVRVSETLQRFWR  
SSNENGGPPLTPTQKRKMAGKIRSEV [서열번호 13]

[0152] 따라서, 바람직하게는 상기 인플루엔자 NS1 폴리펩티드는 실질적으로 서열번호 13에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다. 따라서, 제1 측면의 RNA 구조체는 바람직하게는 서열번호 13을 암호화하는 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0153] 한 구현예에서, 상기 인플루엔자 NS1 폴리펩티드는 다음과 같은 서열번호 14의 뉴클레오티드 서열에 의해 암호화된다:

[0154] atggattccaacactgtgtcaagctttcaggtagattgtcttctttggcatgtcgcaacaagaattgcagaccaagagctagtgatgccccattccttgat  
cggttcgccgagatcagaagtccttaagggaagaggcagcactctcggtctgaacatcgaacagccacctgtgttgaaagcaaatagtagagaggatt  
ctgaaggaagaatccgatgaggcatttagaatgaccatggcctcgccacttgcttcgcgatacctaactgacatgactattgaagagatgtcaagggactgg  
ttcatgctcatgccaagcagaagaaggcaggccctctttgtgtcagaatggaccaggcgataatggataagaacatcatactgaaagcgaatttcagtgtg  
atTTTTgaccggttggagactctgacattactaagggttttcaccgaagaggagcaattgttggcgaaatttcaccattgccttctcttcaggacatact  
aatgaggatgtcaaaaatgcaattggggctctcatcgggggacttgaatggaatgataacacagttcgagtctctgaaactctacagagattcgcttgaga  
agcagtaatgagaatgggggacctccactcactccaacacagaaacggaaaaatggcggggaaaaattaggtcagaagttaga [서열번호 14]

[0155] 따라서, 바람직하게는 상기 인플루엔자 NS1 폴리펩티드는 실질적으로 서열번호 14에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

- [0156] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 19의 RNA 뉴클레오티드 서열을 포함할 수 있다:
- [0157] `auggauccaacacugugucaagcuuucagguagauugcuuccuuggcauguccgcaacaaguugcagaccaagagcuaggugaugccccaauccuugaucggcuucgccgagaucagaaguccuaaagggaagaggcagcacucuggcugacaucgaaacagccaccuguguuggaaagcaauagagaggauucugaaggaagaauccgaugaggcauuuagaauagaccuagccucgcacucugcuucgcgauaccuacugacauuugaagagaugucaaggagacugguucugcucaugcccaagcagaaguggcaggccucuuugugucagaauggaccaggcgauaauaggauaagaacaucauacugaaagcgaaauucaguguaauuuugaccgguuggagacucugacauuacuaaggguuucaccgaagaggagcaauuugugcgaaauucaccauugccuucucuccaggacauucaaugaggauucaaaaaugcaauugggguccucaucgggggacuugaauaggauaacaacagauucgagucucugaaacucucacagagauucgcuggagagcaguaaagagaauuggggaccuccacucacuccaacaacagaaacggaaaaauaggucagaaguuuga` [서열번호 19]
- [0158] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 19로 나타낸 것과 같은 RNA 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.
- [0159] 바람직하게는, 상기 선천적 억제제 단백질은 HSV-2 Us1; HSV-1 Us11; OV20.0L; BVDV Npro, PIV5 V; MERS-CoV ORF4a; SARS-CoV-2 ORF3b; 란가트 바이러스 NS5 및 인플루엔자 NS1;으로 이루어진 군으로부터 선택된다.
- [0160] 바람직하게는, 상기 선천적 억제제 단백질은 HSV-2 Us1; HSV-1 Us11; OV20.0L; BVDV Npro, PIV5 V; MERS-CoV ORF4a; 및 란가트 바이러스 NS5;로 이루어진 군으로부터 선택된다.
- [0161] 바람직하게는, 상기 선천적 억제제 단백질은 HSV-2 Us1; HSV-1 Us11; OV20.0L; BVDV Npro, PIV5 V; 및 MERS-CoV ORF4a;로 이루어진 군으로부터 선택된다.
- [0162] 실시예에 기술된 것과 같이, 넓은 범위의 IIP를 스크리닝하는 공정을 통해, 본 발명자들은 놀랍게도 시험관내 및 생체내에서 saRNA 발현을 현저하게 향상시킬 수 있는 2개의 매우 강력한 IIP를 확인하였다: MDA5 활성화의 억제제인 파라인플루엔자 바이러스 타입 5 V 단백질(PIV-5) 및 PACT/MDA 활성화의 억제제인 중동 호흡기 증후군(MERS) 코로나바이러스 ORF4a. 본 발명자들은 레플리콘 RNA의 선천적 감지를 완하시키기 위해 이전에 사용되지 않은 상기 2개의 신규한 IIP인 PIV-5 및 MERS-CoV ORF4a를 이용하였다. 본 발명자들은 상기 구조체가 인간에서 더 높은 효능의 RNA 백신 및 치료제를 가능하게 하는 차세대 RNA 레플리콘일 수 있음을 보였다. 본 발명자들은 이들이 알파바이러스, 피코르나바이러스, 플라비바이러스, 또는 코로나바이러스 레플리콘에서 발현될 때 현저한 유용성을 가질 것으로 여긴다.
- [0163] 따라서, 가장 바람직하게는, 상기 선천적 억제제 단백질은 PIV-5 및/또는 숙련된 기술자가 NS4a로도 나타낼 수 있음을 이해하는 MERS-CoV ORF4a이다.
- [0164] 이러한 구조체는 본 기술분야에 기술된 구조체보다도 다음을 포함하는 많은 이점을 나타낸다:
- [0165] i) VEEV 레플리콘 내로 PIV-5 및 ORF4a 단백질을 직접 삽입하여, VP1 단백질 및 관심있는 유전자의 이중 단백질 발현을 가능하게 함;
- [0166] ii) 관심있는 유전자(GOI), 즉 치료 생체분자를 암호화하는 RNA, 및 IIP를 암호화하는 RNA의 2개의 상이한 가닥의 RNA를 운반하는 것과 대조적으로, 단일 가닥이 운반되고, 이에 따라 RNA 및 선천적 억제 단백질의 동시국소화를 보장함;
- [0167] iii) 상기 IIP는 RNA의 선천적 감지를 억제하고, 따라서 더 높은 단백질 발현을 가능하게 함;
- [0168] iv) 상기 IIP 발현 자체는 GOI와 서브게놈 상에 공동-발현됨으로써 자가-증폭됨; 및/또는
- [0169] v) 종래의 VEEV RNA 레플리콘 구조체와 비교하여 단백질 발현의 크기 및 기간 모두에서 증가됨.
- [0170] 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열은 RNA 구조체 또는 레플리콘 서열 내의 임의의 곳에 배치될 수 있어서, 관심있는 적어도 하나의 펩티드 또는 단백질을 암호화하는 서열은 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열의 5' 또는 3'에 배치될 수 있다.
- [0171] 그러나, 바람직하게는 관심있는 적어도 하나의 펩티드 또는 단백질을 암호화하는 서열은 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열의 5'에 배치된다.
- [0172] 바람직하게는, 제1 측면에 따른 RNA 구조체는 게놈 또는 서브게놈에 적어도 하나의 프로모터(promoter)를 포함한다. 바람직하게는, 상기 프로모터는 서브게놈 프로모터이다.
- [0173] 숙련된 기술자는 상기 서브게놈 프로모터가 적어도 하나의 치료 생체분자 및 적어도 하나의 선천적 억제제 단백

질을 암호화하는 서열에 작동가능하게 연결되어서, 상기 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 뉴클레오타이드 서열을 전사할 수 있는 프로모터에 관한 것임을 이해할 것이다.

- [0174] 바람직하게는, 상기 서브게놈 프로모터는 다음과 같은 서열번호 57로서 본 명세서에 제공되는 26S이다:
- [0175] GGGCCCCTATAACTCTCTACGGCTAACCTGAATGGACTACGACAT [서열번호 57]
- [0176] 따라서, 바람직하게는 상기 프로모터(바람직하게는, 서브게놈 프로모터)는 실질적으로 서열번호 57에 나타낸 것, 또는 그의 변이체 또는 단편이다.
- [0177] 한 구현예에서, 동일한 프로모터가 관심있는 적어도 하나의 펩티드 또는 단백질을 암호화하는 서열 및 적어도 하나의 선천적 억제제를 암호화하는 서열에 작동가능하게 연결된다.
- [0178] GOI(즉, 치료 생체분자) 및 IIP가 모두 단일 가닥에 암호화되는 본 발명자들의 디자인은 유리하게는 훨씬 적은 용량의 RNA를 사용하는 것을 가능하게 하는데, 그 이유는 단백질이 상기 RNA에 의해 감지되는 동일한 세포에서 발현되고, 또한 복제될 수 있으며, 이에 따라 선천적 억제제 성분이 증폭되는 부가적인 측면을 갖기 때문이다.
- [0179] 따라서, 한 구현예에서, 상기 프로모터는 관심있는 펩티드 또는 단백질(즉, 치료 생체분자)을 암호화하는 서열 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열의 5'에 배치되어서, 상기 프로모터는 양쪽 서열 모두에 작동가능하게 연결된다.
- [0180] 다른 구현예에서, 제1 프로모터는 관심있는 적어도 하나의 펩티드 또는 단백질(즉, 치료 생체분자)을 암호화하는 서열에 작동가능하게 연결되고, 제2 프로모터는 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열에 작동가능하게 연결된다.
- [0181] 상기 RNA 구조체는 적어도 2개, 3개, 4개 또는 5개 IIP를 암호화할 수 있다. 선천적 억제제 단백질을 암호화하는 서열에 1개 이상인 구현예에서, 단일 프로모터는 선천적 억제제 단백질을 암호화하는 모든 서열에 작동가능하게 연결될 수 있다. 대안적으로, 프로모터는 선천적 억제제 단백질을 암호화하는 각각의 적어도 하나의 다른 서열에 연결되어서, 각각의 선천적 억제제 단백질이 별도의 프로모터에 작동가능하게 연결될 수 있다. 상기 구현예에서, 별도의 프로모터는 동일한 프로모터 서열 또는 상이한 프로모터 서열을 포함할 수 있다. 다른 구현예에서, 상이한 프로모터는 선천적 억제제 단백질을 암호화하는 각각의 서열에 작동가능하게 연결된다.
- [0182] 상기 RNA 구조체는 관심있는 적어도 하나의 펩티드 또는 단백질(즉, 치료 생체분자)을 암호화하는 서열 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열 사이에 배치된 링커(linker) 서열을 추가로 포함할 수 있다.
- [0183] 한 구현예에서, 상기 링커 서열은 소화됨으로써 관심있는 유전자에 의해 암호화되는 적어도 하나의 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 분리하도록 구성되는 펩티드 스페이스(spacer)를 암호화하는 서열을 포함한다. 따라서, 바람직하게는 상기 스페이스 서열은 관심있는 적어도 하나의 펩티드 또는 단백질을 암호화하는 서열 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열 사이에 배치된다.
- [0184] 이와 같이, 상기 스페이스 서열은 바람직하게는 절단가능한 펩티드, 예를 들면 2A 펩티드이다. 적합한 2A 펩티드는 돼지 테스코바이러스(teschovirus)-1 2A(P2A)-ATNFSLLKQAGDVEENPGP(서열번호 21), 토세아 아시그나(*thosea asigna*) 바이러스 2A(T2A)-QCTNYALLKLKAGDVESNPGP(서열번호 22), 말 비염 A 바이러스 2A(E2A), 및 구제역 바이러스 2A(F2A) VKQTLNFDLLKLAGDVESNPGP(서열번호 23)를 포함한다. 바람직하게는, 상기 2A 펩티드는 토세아 아시그나 바이러스 2A(T2A)이다.
- [0185] 다른 구현예에서, 상기 절단가능한 펩티드는 자가-절단 펩티드이다. 바람직하게는, 상기 자가-절단 펩티드는 푸린(furin)/2A 펩티드이다. 상기 푸린 서열은 상기 2A 서열의 3' 또는 5'에 배치될 수 있다. 그러나, 바람직하게는, 상기 푸린 서열은 상기 2A 서열의 5'에 배치되고, 바람직하게는 GSG 스페이스가 상기 푸린 및 2A 서열 사이에 배치된다.
- [0186] 숙련된 기술자는 푸린이 특이적 인식 서열(정규적으로 R-X-R/K/X-R(서열번호 24))에서 전구체 단백질을 절단하고, 최종 R 후에 전구단백질을 절단하는(주로 골지 및 트랜스-골지 네트워크 내의) 분비 경로에 위치하는 편재성(ubiquitous) 칼슘-의존형 전구단백질 컨버타아제(convertase)임을 인식할 것이다. 따라서, 한 구현예에서, 상기 푸린 서열은 R-X-R/K/X-R이다. 그러나, 바람직하게는, 상기 푸린 서열은 최적화된 서열 RRRRRR(서열번호 25) GSG 서열이다. 바람직하게는, 상기 GSG 상기 푸린 서열의 3' 및 상기 2A 서열의 5'에 배치된다.
- [0187] 따라서, 바람직하게는, 상기 스페이스 서열은 NCBI 참조 서열: 진뱅크(GenBank): AAC97195.1에 의해 제공되고,

본 명세서에서 다음과 같은 서열번호 26으로 제공되는 푸린/T2A이다:

[0188] RRRRRRGSGEGRGSLTTCGDVEENPGP [서열번호 26]

[0189] 따라서, 바람직하게는, 상기 스페이서 서열은 서열번호 26에 나타난 것과 같은 아미노산 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0190] RNA 구조체 또는 레플리콘이 선천적 억제 단백질을 암호화하는 하나 이상의 서열을 포함하는 구현예에서, 상기 레플리콘은 선천적 억제제 단백질을 암호화하는 각각의 서열 사이, 또는 일부 IIP 사이에만 배치되는 링커 서열을 가질 수 있다.

[0191] 한 구현예에서, 상기 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열은 정지 코돈(stop codon)에 의해 분리되고, 이어서 하류 서열의 번역을 시작할 수 있는 내부 리보솜 도입 위치(IRES) 서열이 뒤따를 수 있다. 전형적인 IRES 서열은 뇌심근염 바이러스 또는 혈관 내피 성장 인자 및 타입 1 콜라겐-유도성 단백질(VCIP)의 IRES 서열과 같은 것을 포함하며, 본 기술분야의 기술자에게 알려져 있다. 따라서, 바람직하게는 상기 IRES 서열은 관심있는 적어도 하나의 펩티드 또는 단백질을 암호화하는 서열 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열 사이에 배치된다. 적어도 하나의 선천적 억제제 단백질을 암호화하는 복수의 서열이 사용된다면, 스페이서 서열은 공지된 절단 서열 및/또는 IRES 서열의 조합을 포함할 수 있다.

[0192] 다른 구현예에서, 상기 적어도 하나의 치료 생체분자 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열은 정지 코돈에 의해 분리되고, 이어서 하류 서열의 전사를 시작할 수 있는 제2 서브게놈 프로모터 서열이 뒤따를 수 있다.

[0193] 상기 RNA 구조체는 관심있는 적어도 하나의 펩티드 또는 단백질 및 적어도 하나의 선천적 억제 단백질을 암호화하는 서열의 5' 또는 3'에 배치된 적어도 하나의 비-구조 단백질(NSP)을 암호화할 수 있다. 바람직하게는, 상기 적어도 하나의 NSP를 암호화하는 서열은 관심있는 펩티드 또는 단백질 및 적어도 하나의 선천적 억제 단백질을 암호화하는 서열의 5'에 배치된다. 따라서, 바람직하게는 상기 적어도 하나의 NSP를 암호화하는 서열은 상기 RNA 구조체의 5' 말단에 배치된다.

[0194] 상기 RNA 구조체에 의해 암호화되는 적어도 하나의 비-구조 단백질은 RNA 폴리머라아제 nsP4일 수 있다. 바람직하게는, 상기 구조체는 nsP1, nsP2, nsP3 및 nsP4를 암호화한다. 숙련된 기술자는 nsP1은 바이러스 캡핑 효소 및 복제 복합체(RC)의 막 앵커(anchor)이지만, nsP2는 ns 폴리단백질 처리를 담당하는 RNA 헬리카아제(helicase) 및 프로테아제(protease)임을 이해할 것이다. nsP3는 몇 가지 숙주 단백질과 상호작용하고, 단백질 폴리- 및 모노-ADP-리보실화를 조정할 수 있으며, nsP4는 코어 바이러스 RNA-의존형 RNA 폴리머라아제이다.

[0195] 한 구현예에서, nsP1은 다음과 같은 서열번호 27로서 본 명세서에 제공된다:

[0196] MEKVHVDIEEDSPFLRALQRSFPQFEVEAKQVTDNDHANARAFSHLASKLIETEVDPSTILDIGSAPARMYSKHKYHCICPMRCAEDPDRLKYATKLLKNCKEITDKELDKMKELAAVMSDPDLETETMCLHDDDESCRYEGQVAVYQDVYAVDGPSTLYHQANKGVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVMERSRRGMSILRKKYLKPSNNVLFVSGSTIYHEKRDLLRSWHLPSVFLRGKQNYTCRCETIVSCDGYVVKRIASPGLYGKPSGYAATMHREGFLCKVTDLTNGERVSFPVCTYVPATLCDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFARWAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSYKRPDTQTIKVNDSDFHSFVLPRIGSNTLEIGLRTRIRKMLEEHKEPSPLITAEDVQEAACADEAKEVREAEELRAALPPLAADVEEPTLEADVLDMLQEAGA [서열번호 27]

[0197] 따라서, nsP1은 바람직하게는 실질적으로 서열번호 27에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

[0198] 한 구현예에서, nsP1은 다음과 같은 서열번호 28에 정의된 것과 같은 뉴클레오티드 서열에 의해 암호화된다:

[0199] ATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCATTCTCAGAGCTTTGACGCGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAACTGATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCGCCCGCAGAATGTATTCTAAGCACAAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAA AACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAAATGAAGGAGCTGGCCGCGTCATGAGCGACCCTGACCTGGAAGTGAAGACTATGTGCCTCCAC GACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGTTGACGGACCGACAAGTCTCTATACCAAGCCAATAAGGGAGTT AGAGTCGCCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTG TTAACGGCTCGTAACATAGGCCTATGACGCTCTGACGTTATGGAGCGGTACAGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTGAAACCATCCAACAAT GTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCCGCTGTATTTCACTTACGTGGCAAGCAAAATTAC

ACATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACG  
ATGCACCGCGAGGGATTCTTGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCAGCTACATTGTGTGAC  
CAAATGACTGGCATACTGGCAACAGATGTCAGTGGGACGACGCGCAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCACCGGTCGACCCAGAGA  
AACACCAATACCATGAAAAATTACCTTTTGGCCGTAGTGGCCAGGCATTGCTAGGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTA  
GGACTACGAGATAGACAGTTAGTCATGGGGTGTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATCATCAA  
GTGAACAGCGATTTCCTACTATTCGTGCTGCCAGGATAGGCAGTAACACATTGGAGATCGGGCTGAGAACAAAGATCAGGAAATGTTAGAGGAGCACAAG  
GAGCCGTACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGGCAGCCGATGAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTA  
CCACCTTTGGCAGCTGATGTTGAGGAGCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCC [서열번호 28]

[0200] 따라서, nsP1은 바람직하게는 실질적으로 서열번호 28에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0201] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 50으로 나타난 것과 같은 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0202] AUGGAGAAAGUUCACGUUGACAUUCGAGGAAGACAGCCAUUCCUCAGAGCUUUGCAGCGAGCUUCCCGAGUUUGAGGUAGAAGCCAAGCAGGUCACUGAU  
AAUGACCAUGCUAAUUGCCAGAGCGUUUUGCGAUCUGGCUUCAAACUGAUCGAAACGGAGGUGGACCCAUCCGACACGAUCCUUGACAUUGGAAGUGCGCC  
GCCCCGAGAAUGUAUUCUAGCACAAGUAUCUUGUAUCUGUCCGAGAGAGUGCGGAAGAUCCGGACAGAUUGUAUAGUAUGCAACUAAAGCUGAAGAAA  
AACUGUAAGGAAUAAACUGAUAGGAUUGGACAAGAAAUGAAGGAGCUGGCCGCCGUCUAGAGCGACCCUGACCUGGAAACUGAGACUUGUGCCUCCAC  
GACGACGAGUCGUGCGUACGAAGGGCAAGUCGUGUUUACCAGGAUGUAUACGCGGUGACGACCGACAAGUCUCUACCAAGCCAAUAAAGGGAGUU  
AGAGUCGCCUACUGGAUAGGCUUUGACACCACCCUUUUAUGUUUAAAGAACUUGGUGGAGCAUUAUCCAUACUACUACCAACUGGGCCGACGAAACCGUG  
UUAACGGCUCGUAACAUAGGCCUUAUGCAGCUCUGACGUUUGGAGCGGUCACGUAGAGGGAUGUCCAUUCUAGAAAGAAGUAUUUGAAACCAUCCAACAAU  
GUUCUUAUUCUGUUGGUCGACCAUCUACCACGAGAAGAGGGACUUAUCUGAGGAGCUGGCACCUGCCGUCUGUAUUUACCUUACGUGGCAAGCAAAUUAUAC  
ACAUGUCGGUGUGAGACUUAAGUUAUGUUGCGACGGUACGUCGUUAAAAGAAUAGCUAUCAGUCCAGGCCUGUAUGGGAAGCCUUCAGGCUAUGCUGCUACG  
AUGCACCGCGAGGGAUUCUGUGCUGCAAAGUGACAGACACAUUGAACGGGAGAGGGUCUCUUUCCCGUGGCACGUAUGUGCCAGCUACAUGUGUGAC  
CAAAUGACUGGCAUACUGGCAACAGAUUGCAGUGCGGACGACGCGCAAAACUGCUGGUUGGGCUCAACCAGCGUAUAGUCGUACACGGUCGACCCAGAGA  
AACACCAUACCAUGAAAAUUUACCUUUUGCCGUAGUGGCCAGGCAUUGCUAGGUGGCAAGGAUUAAGGAAGAUCAAGAAGAUAAAGGCCACUA  
GGACUACGAGAUAGACAGUUAGUCAUGGGGUGUUGUUGGCUUUUAGAAGGCACAAGAUAAUCUUAUUUAAGCGCCCGGAUACCCAAACCAUCAUCAA  
GUGAACAGCAUUAUCCACUUAUUGUGCUGCCAGGAUAGGCAGUAACAUUGGAGAUCCGGCUGAGAACAAAGAAUAGGAAAUUGUAGAGGAGCACAAG  
GAGCCGUCACCUCUUAUACCGCCGAGGACGUACAAGAAGCUAAGUGCGCAGCCGAUGAGGCUAAGGAGGUGCGUAAGCCGAGGAGUUGCGCGCAGCUCUA  
CCACCUUUGGCAGCUGAUGUUGAGGAGCCACUCUGGAAGCCGAUGUCGACUUGAUGUACAAGAGGCGUGGGCC [서열번호 50]

[0203] 한 구현예에서, nsP2는 다음과 같은 서열번호 29로서 본 명세서에 제공된다:

[0204] GSVETPRGLIKVTSYDGEDKIGSYAVLSPQAVLKSEKLSCHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESATIVYNERE FVNRY  
LHHIATHGGALNTDEEYKTVKPSHDEGYLYIDRKQCVKELVTGLTGLTGLVDPFHEFAYESLRTPAAPYQVPTIGVYVPGSGKSGIKSAVTKKD  
LVVSARKENCAEIRDVKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAGTLRALIAIRPKKAVLCGDPKQCGFFNMCLKVHFNHEICTQVPHK  
SISRRCTKS VTSVSTLFYDKKMRITNPKETKIVIDTTGSKPKQDDLILTCFRGWVKQLQIDYKGNEMTAAASQGLTRKGVYAVRYKVNENPLYAPTSEH  
VNVLLTRTEDRIVWKTLAGDPWIKTLTAKYPGNFTATIEEWQAEHDAIMRHLILERPDPDVFQNKANVCWAKALVPVLKTAGIDMTTEQWNTVDYFETDKAH  
SAEIVLNQLCVRFFGLDLSGLFSAPTVPISIRNNHWDNSPSNMYGLNKEVVRQLSRYPQLPRAVATGRVYDMNTGTLRNYDPRINLPVNNRRLPHALVL  
HHNEHPQSDFFSVSKLKGRTVLVVGKLSVPGKMVDWLSDRPEATFRARLDLGI PGDVPKYDIIFVNVRTPKYHHYQQCEDHAIKLSMLTKKACLHLNPG  
GTCVSI GYGADRASESII GAIARQKF SRVCKPKSSLEETEVLV FVFIGYDRKARTHSYKLSSTLTNIYTGSRLEHAGC [서열번호 29]

[0205] 따라서, nsP2는 바람직하게는 실질적으로 서열번호 29에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

[0206] 한 구현예에서, nsP2는 다음과 같은 서열번호 30에 정의된 것과 같은 뉴클레오타이드 서열에 의해 암호화된다:

[0207] GGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTCCGAGGCTGTACTCAAG  
AGTGAAAAATTATCTGTCATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAA  
GTAGTGGTGCCAGAGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTGAGTGAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAACAGGTAC  
CTGCACCATATTGCCACATAGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAACGTCAAGCCGAGCAGCAGCGCAATACCTGTACGACATC  
GACAGGAAACAGTGCCTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCAGAGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCTACGAGAGTCTGAGA  
ACACGACCAGCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCTGCCAGGATCAGGCAAGTCTGGCATCATTAAAGCGCAGTCACCAAAAAAGAT  
CTAGTGGTGAGCGCAAGAAAGAAAACGTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTC

TTGAATGGATGCAAACACCCGCTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAA  
AAGGCAGTGTCTGCGGGGATCCCAAACAGTGCAGTTTAAAAACATGATGTGCCTGAAAGTGCAATTTAAACCAGAGATTGTCACACAAGTCTTCCACAAA  
AGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTACGACAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTG  
ATTGACACTACCGGCAGTACAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAA  
ATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCCGTACAAGGTGAATGAAAATCCTCTGTACGCACCCACCTCAGAACAT  
GTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGAAAACTAGCCGGCGACCCATGGATAAAAACTGACTGCCAAGTACCCTGGGAATTC  
ACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAACGTTG  
TGTGGGCCAAGGCTTTAGTGCCGCTGTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTGAAACGGACAAAGCTCAC  
TCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTCTGCACCCACTGTTCCGTTATCCATTAGGAAT  
AATCACTGGGATAACTCCCCGTCGCTAACATGTACGGGTGAATAAAGAAGTGGTCCGTACGCTCTCTCGCAGGTACCCACAACCTGCCTCGGGCAGTTGCC  
ACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCTC  
CACCATAATGAACACCCACAGAGTGACTTTTCTTCATTGTCAGCAATTAAGGGCAGAACTGTCTGGTGGTGGGGAAAAGTTGTCCGTCACAGGCAAA  
ATGGTTGACTGGTTGTCAGACCGGCTGAGGCTACCTTCAGAGCTCGGTGGATTAGGCATCCAGGTGATGTGCCAAATATGACATAATATTTGTTAAT  
GTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGC  
GGAACCTGTGTACAGCATAGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTTGGTGTATAGCGCGCAGTTCAAGTTTCCCGGGTATGCAACCG  
AAATCCTCACTTGAAGAGACGGAAGTTCTGTTGTATTGTTGATGATCGCAAGGCCGTACGCACAATTTTACAAGCTTTCATCAACCTTGACCAAC  
ATTTATACAGTTCCAGACTCCACGAAGCCGATGT [서열번호 30]

[0208] 따라서, 바람직하게는 nsP2는 실질적으로 서열번호 30에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0209] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 51을 포함할 수 있다:

[0210] GGCUCAGUGGAGACACCUCUGGCUUGAUAAAGGUUACCAGCUACGAUGGCGAGGACAAGAUCCGCUUACGCUUGCUUUCUCCGAGGCUGUACUCAAG  
AGUGAAAAUUUUCUUGCAUCCACCCUCUGCUGAACAAGUCAUAGUGAUAAACACUCUGGCGGAAAAGGGCGUUAUUGCCGUGGAACCAUACCAUGGUAAA  
GUAGUGGUGCCAGAGGGACAUGCAAUACCCGUCCAGGACUUAAGCUCUGAGUGAAAGUGCCACCAUUGUGUACAACGAACGUGAGUUCGUAAACAGGUAC  
CUGCACCAUUAUUGCCACAUAGGAGGAGCGUGAACACUGAUGAAGAAUUAUACAAACUGUCAAGCCAGCGAGCAGCAGCGCAAUACCUUGUACGACAUC  
GACAGGAAACAGUGCGUCAAGAAAGAAUAGUCACUGGGCUAGGGCUCACAGGCGAGCUGGUGGAUCCUCCUUAUGAAUUCGCUACGAGAGUCUGAGA  
ACACGACCGAGCGCUCCUUAACCAAGUACCAACCAUAGGGGUGUAUGGCGUGCCAGGAUCAGGCAAGUCUGGCAUCAUUAAGCGCAGUCACCAAAAAAGAU  
CUAGUGGUGAGCGCAAGAAAGAAAACUGUGCAGAAUUAUAGGGACGUCAGAAAAUAGAAAGGGCUGGACGUCAAUUGCCAGAACUGUGGACUCAGUGCUC  
UUGAAUUGGAUGCAACACCCGUGAGAGCCUGUAUUAUGACGAAGCUUUGCUUGUCAUGCAGGUACUCUCAGAGCGCUCUAGCCAUUAUAGACCUAAA  
AAGGCAGUGCUCUGCGGGGAUCCAAACAGUGCGGUUUUUUAACAUGAUGGCCUGAAAGUGCAUUAUUAACACGAGAUUUGCACACAAGUCUCCACAAA  
AGCAUCUCUGCCGUGGACUAAAUCUGUGACUUCGGUGCUCUACCUUGUUUACGACAAAAAUGAGAACGAGCAUCCGAAAGAGACUAAAGAUUGUG  
AUUGACACUACCGGAGUACCAACCUAAGCAGGACGAUCUACUUCACUUGUUUCAGAGGGUGGGUGAAGCAGUUGCAAUAGAUUACAAGGCAACGAA  
AUAAUGACGGCAGCUGCCUCUACAGGGCUGACCCGUAAGGUGUGUAUGCCGUUCGUAACAAGGUGAAUAGAAAUCCUCUGACGCACCCACCUCAGAACAU  
GUGAACGUCCUACUGACCCGACGAGGACCGCAUCUGUGGAAACACUAGCCGGCGACCAUGGAUAAAAACACUGACUGCCAAGUACCCUGGGAAUUC  
ACUGCCACGAUAGAGGAGUGGCAAGCAGAGCAUGAUGCCAUCAUGAGGCACAUUCUGGAGAGACCGACCCUACCGACGUCUCCAGAAUAGGCAACGUG  
UGUUGGGCAAGGCUUAGUGCCGUGCUGAAGACCGUGGCAUAGACAUGACCACUGAACAAUGGAACACUGUGGAUUAUUUGAAACGGACAAAGCUCAC  
UCAGCAGAGAUAGAUUAGAACCAACUAGCGUGAGGUUCUUGGACUGAUCUGGACUCCGGUCUUAUUUCUGACCCACUGUCCGUUAUCCAUUAGGAU  
AAUCACUGGGAAUACUCCCGUGCCUUAACAUAGCAGGCGUAAUAAAGAGUGGUCGUCAGCUCUCGCAGGUACCCACAACUGCCUGGGCAGUUGCC  
ACUGGAAGAGUCUAGACAUGAACACUGGUACACUGCGCAUUAUGAUCCGCGCAUAAACCUAGUACCGUAAACAGAAAGACUGCCUACGCUUAGUCCUC  
CACC AUAAUGAACACCCACAGAGUGACUUUUCUUAUCGUCAGCAAUUGAAGGGCAGAACUGUCCUGGUGUGGGGAAAAGUUGUCCGUCCAGGCAAA  
AUGGUUGACUGGUUGACAGACCGGCCUGAGGCUACCUUCAGAGCUCGGCUGGAUUUAGGCAUCCAGGUGAUGUGCCAAAUAUGACAUAUUAUUGUUAU  
GUGAGGACCCAUUAUAAUACCAUACUACAGCAGUGUGAAGACCAUGCCAUUAAGCUUAGCAUGUUGACCAAGAAAGCUGUGCUGAUCUGAAUCCCGGC  
GGAACCUUGUGCAGCAUAGGUUAUGGUUACGUGACAGGGCCAGCGAAAGCAUCAUUGGUGCUAUAAGCGCGCAGUUAAGUUUCCCGGUAUGCAAACCG  
AAUCCUCACUUAAGAGACGGAAGUUCUGUUUAUUAUUGGUACGAUCGAAGGCCGUAACGCACAUAUUAAGCUUUAUUAACCUUGACCAAC  
AUUAUAACAGGUUCCAGACUCCACGAAGCCGAUGU [서열번호 51]

[0211] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 51로 나타난 것과 같은 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.

[0212] 한 구현예에서, nsP3은 다음과 같은 서열번호 31로서 본 명세서에 제공된다:

[0213] APSYHVVRGDIATATEGVI INAANSKGQPGGGVCGALYKFPESFDLQPIEVGKARLVKGAAKHI IHAVGPNFNKVSEVEGDKQLAEAYESI AKIVNDNNYK

SVAIPLSTGIFSGNKDRLTQSLNHLALTALDTTDADVAIYCRDKKWEMLKEAVARREAVEEICISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFS  
YLEGTFKHQAAKDIAEINAMWPVATEANEQVCMYILGESMSSIRSKCPVEESEASTPPSTLPCLCIHAMTPERVQRLKASRPEQITVCSFPLPKYRITGVQ  
KIQCSQPIILFSPKVPAYIHPKRYLVETPPVDETPEPSAENQSTEGTPEQPPLITEDETRTRTPPEPIIEEEEEEDSISLLSDGPTHQVLQVEADIHGPPSVSS  
SSWSIPHASDFDVSLSILDLEGASVTSGATSATNSYFASKMEFLARPVPAPRTVFRNPPHPAPRTRTPSLAPSACRSRTSLVSTPPGVNVRVITREELEA  
LTPSRTPSRVSRTSLVSNPPGVNVRVITREEFEAFVAQQQRFDA [서열번호 31]

[0214] 따라서, 바람직하게는 nsP3은 실질적으로 서열번호 31에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

[0215] 한 구현예에서, nsP3은 다음과 같은 서열번호 32에 정의된 것과 같은 뉴클레오티드 서열에 의해 암호화된다:

[0216] GCACCTCATATCATGTGGTGGCAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGC  
GGAGCGCTGTATAAGAAATCCCGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATTATGC  
GTAGGACCAAACTTCAACAAAGTTTCGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAG  
TCAGTAGCGATTCCACTGTTGCCACGGCATCTTTCCGGGAACAAAGATCGACTAACCAATCATTGAACATTTGCTGACAGCTTTAGACACCACTGAT  
GCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGAC  
TCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAACTTTCTCA  
TATTTGGAAGGGACCAAGTTTACCAGGCGGCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTAT  
ATCCTCGGAGAAAGCATGAGCAGTATTAGGTGCAATGCCCCGTCGAAGAGTCGAAGGCCTCCACACCCTAGCACGCTGCCTTGCTTGTGCATCCATGCC  
ATGACTCCAGAAAGAGTACAGCGCTAAAAGCCTCAGTCCAGAACAAATTACTGTGTGCTCATCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAG  
AAGATCCAATGTCTCCAGCCTATATTGTTCTACCGAAAGTGCTGCGTATATTCATCCAAGGAAGTATCTCGTGAAACACCACCGGTAGACGAGACTCCG  
GAGCCATCGGAGAGAACCAATCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATC  
GAAGAGGAAGAAGAGGATAGCATAAGTTTGCTGTCAGATGGCCCGACCCACAGGTGCTGCAAGTCGAGGCAGACATTACGGGCCGCCCTCTGTATCTAGC  
TCATCTGGTCCATTCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGGGAGCTAGCGTGACCAGCGGGCAACGTCAGCC  
GAGACTAACTCTTACTTCGAAAGAGTATGGAGTTTCTGGCGCGACCGGTGCTGCGCTCGAACAGTATTCAGGAACCTCCACATCCCGCTCCGCGCACA  
AGAACACCGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCGCCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCG  
CTTACCCCGTCACGACTCTAGCAGGTCGGTCTCGAGAACCAGCCTGGTCTCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCG  
TTCGTAGCACAACAATGACGGTTTGATGCGGGTGCA [서열번호 32]

[0217] 따라서, 바람직하게는 nsP3은 실질적으로 서열번호 32에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

[0218] 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 52를 포함할 수 있다:

[0219] GCACCCUCAUAUCAUGUGGUGCGAGGGGAUAUUGCCACGGCCACCGAAGGAGUGAUUAUAAUUGCUGCUAACAGCAAAGGACAACUGGCGGAGGGGUGUGC  
GGAGCGCUGUAUAAGAAAUUCCCGAAAGCUUGCAUUUACAGCCGAUCGAAGUAGGAAAAGCGCGACUGGUCAAAGGUGCAGCUAAACAUUAUCAUUGCC  
GUAGGACCAAAACUUAACAAAGUUCGAGGUGUAAGGUGACAAACAGUUGGCAGAGGCUUAUGAGUCCAUUCGUAAGAUAUGUCAACGAUAACAUAUACAAG  
UCAGUAGCGAUUCCACUGUUGUCCACCGGCAUCUUUCCGGGAACAAAGAUUCGACUAACCAUAUUGAACCAUUGUGACAGCUUAGACACCACUGAU  
GCAGAUUGAGCAUAUACUGCAGGGACAAGAAUUGGAAUAGACUCUCAAGGAAGCAGUGGCUAGGAGAGAAGCAGUGGAGGAGAUUGCAUAUCCGACGAC  
UCUUCAGUGACAGAACCGAUGCAGAGCUGUGAGGGUGCAUCCGAAGAGUUCUUUGGCUUGAAGGAAGGGCUACAGCACAAGCGAUUGGCAAAACUUCUCA  
UAUUUGGAAGGGACCAAGUUCACCAGGCGGCAAGGAUAUAGCAGAAUUAUUGCAUUGGCCCUGUGCAACGGAGGCCAAUGAGCAGGUAUGCAUGUAU  
AUCCUCGAGAAAGCAUGAGCAGUAUAGGUCGAAUUGCCCGUCGAAGAGUCGGAAGCCUCCACACCACCUAGCACGUGCCUUGCUUGUGCAUCCAGGCC  
AUGACUCCAGAAAGAUACAGCGCCUAAAAGCCUCACGUCCAGAACAAAUACUGUGUGCUAUCCUUUCCAUUGCCGAAGUAUAGAAUACUGUGUGCAG  
AAGAUCCAAUGUCCACCGCUUAUUGUUCACCGAAAGUGCCUGCGUAUUAUUAUCCAAAGGAAGUAUCUGUGGAAACACCACCGGUAGACGAGAUCCG  
GAGCAUCCGAGAGAACCAUCCACAGAGGGGACACCGAACAACCACCAUUAUACCGAGGAUGAGACCAGGACUAGAAGCCUGAGCCGAUCAUAUC  
GAAGAGGAAGAAGAGGAUAGCAUAAGUUUGCUGUCAGAUUGGCCCGACCCACAGGUGCUGCAAGUCGAGGCAGACAUCACGGGCCGCCUCUGUAUUCAGC  
UCAUCCUGGUCAUUCUCAUGCAUCCGACUUUGAUGUGGACAGUUUAUCCAUACUUGACACCCUGGAGGAGCUAGCGUGACCAGCGGGCAACGUCAGCC  
GAGACUAACUCUUACUUCGAAAGAGUAUGGAGUUUCUGGCGGACCGGUGCCUGCGCCUGAACAGUAUUCAGGAACCCUCCACAUCCCGCUCCGCGCACA  
AGAACACCGUCACUUGCACCAGCAGGGCCUGCUCGAGAACCAGCCUAGUUUCCACCCGCCAGGCGUGAAUAGGGUGAUACUAGAGAGGAGCUCGAGGCG  
CUUACCCCGUCACGACUCCUAGCAGGUCGGUCUGGAGAACCAGCCUGGUCUCCAACCCGCCAGGCGUAAAUAGGGUGAUUAAGAGAGGAGUUUGAGGCG  
UUCGUAGCACAACAACAUAGACGGUUUGAUGCGGGUGCA [서열번호 52]

[0220] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 52로 나타난 것과 같은 RNA 뉴클레오티드 서열 또는 그의 변이체 또는 단편을 포함한다. 한 구현예에서, nsP4는 다음과 같은 서열번호 33으로서 본 명세서에 제공된다:

YIFSSDTGGHGLQQKSVRQTVLSEVVLERTELEIISYAPRLDQKEEELLRKKLQLNPTPANRSRYQSRKVENMKAITARRILQGLGHYLKAEKGVECYRTLHP  
VPLYSSSVNRAFSSPKVAVEACNAMLKENFPTVASYCIPEYDAYLDMVDGASCCLDTASFCEPAKLRSFPKKHSYLEPTIRSAVPSAIQNTLQNVLAATKR  
NCNVTQMRELPLVDSAAFVNECFKKYACNNEYWETFKENPIRLTEENVVNYITKLKGPAAALFAKTHNLNMLQDIPMDRFVMDLKRDKVPTPGTKHTEERP  
KVQVIQAADPLATAYLCGIHRELVRRLNAVLPLNIHTLFDMSAEDFDAIEAEHFQPGDCVLETDIASFDKSEDDAMALTALMILEDLGVDAELLTLIEAAGF  
EISSIHLPTKTKFKFGAMMKSGMFLTLFVNTVINIVIASRVLRRERLTGSPCAAFIGDDNIVKGKSKDKLMADRCATWLNMEVKIIDAVVGEKAPYFCGGFIL  
CDSVTGTACRVADPLKRLFKLGKPLAADDEHDDRRRALHEESTRWNRVIGILSELCKAVESRYETVGTSIIVMAMTTLASSVKSFSYLRGAPITLYG [서  
열번호 33]

따라서, 바람직하게는 nsP4는 실질적으로 서열번호 33에 나타난 것과 같은 아미노산 서열, 또는 그의 생물학적 활성 변이체 또는 단편을 포함한다.

한 구현예에서, nsP4는 다음과 같은 서열번호 34에 정의된 것과 같은 뉴클레오티드 서열에 의해 암호화된다:

TACATCTTTTCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAACCGGTGCTATCCGAAGTGGTGTTGGAGAGGACCGAATTGGAGATT  
TCGTATGCCCCGCGCTCGACCAAGAAAAAGAAGATTACTACGCAAGAAATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAG  
GTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAAGTGGAGTGCTACCGAACCTGCATCCT  
GTTCCCTTTGTATTATCTAGTGTGAACCGTGCCCTTTTCAAGCCCCAAGGTCGAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAACTTTCCGACTGTGGCT  
TCTTACTGTATTATTCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCTGCAAGCTGCGCAGC  
TTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCCTTACGCGATCCAGAACACGCTCCAGAACGCTCTGGCAGCTGCCACAAAAAGA  
AATTGCAATGTCACGCAAATGAGAGAATTGCCCGTATTGGATTTCGGCGCCTTTAATGTGGAATGCTTCAAGAAATATGCGTGTAATAATGAATATTGGGAA  
ACGTTTAAAGAAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAATTAAGAGGACCAAAAGCTGCTGCTCTTTTTCGAAGACACAT  
AATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCC  
AAGGTACAGGTGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTATAGGAGATTAAATGCGTCTGCTTCCGAAC  
ATTCATACACTGTTTGATATGTGCGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAACCTGACATCGCGTCGTTT  
GATAAAAGTGAAGGACGACGCCATGGCTCTGACCGGTTAATGATTCTGGAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGC  
GAAATTTTCATCAATACATTTGCCCACTAAAATAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCTCACACTGTTTGTGAACACAGTCATTAAC  
ATTGTAATCGCAAGCAGAGTGTGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTATTGGAGATGACAATATCGTGAAAGGAGTCAAATCGGACAAA  
TTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGCGAGAAAGCGCCTATTCTGTGGAGGGTTTATTTTG  
TGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCTAAAAAGGCTGTTTAAAGCTTGCAAAACCTCTGGCAGCAGACGATGAACATGATGATGAC  
AGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTACAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAACCGTAGGAAT  
TCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACGCTACCTGAGAGGGGCCCTATAACTCTCTACGGC [서열번호 34]

따라서, 바람직하게는 nsP4는 실질적으로 서열번호 34에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편에 의해 암호화된다.

따라서, 상기 RNA 구조체는 다음과 같은 서열번호 53을 포함할 수 있다:

UACAUCUUUCCUCCGACACCGGUCAAGGGCAUUUACAACAAAAUCAGUAAGGCAAACGGUGCUAUCCGAAGUGGUGUUGGAGAGGACCGAAUUGGAGAUU

UCGUAUGCCCCGCGCCUCGACCAAGAAAAAGAAGAAUUAUACUACGCAAGAAAUUACAGUUAAAAUCCACACCUGCUAACAGAAGCAGAUACCAGUCCAGGAAG

GUUCCUUUGUAUUAUCUAAGUGGAGAACCGUGCCUUUUCAAGCCCCAAGGUCGAGUGGAAGCCUGUAACGCCAUGUUGAAAGAGAAACUUUCCGACUGUGGCU  
UCUUAACUGUAUUUAUCCAGAGUACGAUGCCUAUUUGGACAUGGUUAGCGGAGCUUCAUGCUGCUUAGACACUGCCAGUUUUUGCCCGCAAAGCUGCGCAGC  
UUUCCAAAGAAACACUCCUAUUUGGAACCCACAUAACGAUCGGCAGUGCCUUCAGCGAUCCAGAACACGCUCCAGAAACGUCCUGGCAGCUGCCACAAAAAGA  
AAUUGCAAUGUCACGCAAAUGAGAGAAUUGCCCGUAUUGGAUUCGGCGGCCUUUAUUGUGGAUUGCUUCAAGAAAUUUGCGUGUAAUAAUAAUUAUUGGGAA  
ACGUUUAAAGAAAACCCCAUCAGGCUUACUGAAGAAAACGUGGUAAAUAACAUUACCAAUUAAAAGGACCAAAAGCUGCUGCUCUUUUUGCGAAGACACAU  
AAUUGAAUAUGUUGCAGGACAUACCAAUGGACAGGUUGUAAUGGACUUAAGAGAGACGUGAAAGUGACUCCAGGAACAAAACAUACUUAAGAACCGGCC  
AAGGUACAGGUGAUCCAGGCGUCCGAUCCGCUAGCAACAGCGUAUCUGUGCGGAUCCACCGAGAGCUGGUUAGGAGAUUAAUUGCGGUCCUGCUUCCGAAC  
AUUCAUACACUGUUUGAUUUGUCGGCUGAAGACUUGAGCGCUAUUUAUAGCCGAGCACUUCACGCCUGGGGAUUGUGUUCUGGAAACUGACAU CGCGUCGUUU  
GAUAAAAGUGAGGACGACGCCAUGGCUCUGACCGCGUUAUUGAUUUCUGGAAGACUUAAGGUGUGGACGACAGCUGUUGACGCUGAUUGAGGCGGCCUUUCGGC  
GAAAUUCAUCAAAUACAUAUUGCCCACUAAAACUAAAUUUAAAUUCGGAGCCAUGAUGAAAUCUGGAAUGUCCUCACACUGUUGUGAACACAGUCAUUAAC  
AUUGUAAUCGCAAGCAGAGUGUUGAGAGAACGGCUAACCGGAUACCAUGUGCAGCAUUCAUUGGAGAUGACAAUUCUGGAAAGGAGUCAAAUUCGGACAAA  
UUAUUGGCAGACAGGUGCGCCACCUGGUUAAUAUGGAAGUCAAGAUUAUAGAUUGCUGUGGUGGGCGAGAAAGCGCCUUAUUUCUGUGGAGGGUUUAUUUUG  
UGUGACUCCGUGACCGGCACAGCGUGCCGUGUGGCAGACCCCCUAAAAGGCUGUUUAAGCUUGGCAAAACCUCUGGCAGCAGACGAUGAACAUGAUGAUGAC  
AGGAGAAGGGCAUUGCAUGAAGAGUCAACACGCGUGGAACCGAGUGGGUAUUCUUUCAGAGCUGUGCAAGGCAGUAGAUAUAGGUAUGAAACCGUAGGAACU

– 26 –

UCCAUCAUAGUUAUGGCCAUGACUACUCUAGCUAGCAGUGUUAUAUACUUCAGCUACCUGAGAGGGGCCCUAUAACUCUCUACGGC [서열번호 53]

- [0228] 따라서, 그러므로, 바람직하게는 상기 RNA 구조체는 실질적으로 서열번호 53으로 나타낸 것과 같은 RNA 뉴클레오타이드 서열 또는 그의 변이체 또는 단편을 포함한다.
- [0229] 바람직하게는, 숙주 세포에 존재하는 단백질과 함께, 본 발명의 RNA 구조체에 의해 암호화되는 비-구조 단백질을 관심있는 적어도 하나의 펩티드 또는 단백질 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열의 게놈 복제 및 전사를 위해 필요한 효소 복합체를 형성한다. 예를 들면, 하나 이상의 비-구조 단백질은 상기 구조체가 관심있는 적어도 하나의 펩티드 또는 단백질 및 적어도 하나의 선천적 억제제 단백질을 암호화하는 뉴클레오타이드 서열을 증폭하게 할 수 있는 폴리머라아제를 암호화할 수 있다.
- [0230] 상기 숙주 세포는 진핵생물 또는 원핵생물 숙주 세포일 수 있다. 바람직하게는, 상기 숙주 세포는 진핵생물 숙주 세포이다. 보다 바람직하게는, 상기 숙주 세포는 포유동물 숙주 세포이다.
- [0231] 상기 RNA 구조체는 적어도 하나의 비-구조 단백질의 5'에 배치된 프로모터를 추가로 포함할 수 있으며, 상기 프로모터는 적어도 하나의 비-구조 단백질을 암호화하는 서열에 작동가능하게 연결되고, 숙주 세포에서 적어도 하나의 비-구조 단백질을 발현할 수 있다.
- [0232] 바람직하게는, 상기 프로모터는 본 명세서에서 다음과 같은 서열번호 54로 나타낼 수 있는 5' UTR 보존된 서열 요소를 포함한다:
- [0233] AUGGGCGGCGCAUGAGAGAAGCCCAGACCAUUAACCUACCCAAA [서열번호 54]
- [0234] 따라서, 바람직하게는 상기 UTR은 적어도 하나의 비-구조 단백질의 5'에 배치되고, 실질적으로 서열번호 54에 나타낸 것과 같은 뉴클레오타이드 서열, 또는 그의 단편 또는 변이체를 포함한다.
- [0235] 바람직하게는, 상기 레플리콘은 폴리A 꼬리(polyA tail)를 포함한다. 바람직하게는, 상기 폴리A 꼬리는 상기 레플리콘의 3' 말단에 배치된다. 상기 레플리콘은 5' 캡을 추가로 포함할 수 있다. 본 발명의 문맥에서, 용어 "5'-캡"은 RNA 캡 구조를 닮은 5'-캡 유사체를 포함하며, 바람직하게는 생체내 및/또는 세포에서 여기에 부착될 경우 RNA를 안정화하거나, 및/또는 RNA의 번역을 향상시키는 능력을 갖도록 변형된다.
- [0236] 5'-캡을 갖는 RNA는 상기 5'-캡의 존재시에 DNA 주형의 시험관내 전사에 의해 달성될 수 있으며, 여기서 상기 5'-캡은 생성된 RNA 가닥 내로 공동-전사적으로 통합되거나, 상기 RNA는, 예를 들면, 시험관내 전사에 의해 생성될 수 있고, 상기 5' -캡은 캡핑 효소, 예를 들면, 백시니아 바이러스의 캡핑 효소를 이용하여 사후-전사적으로 RNA에 부착될 수 있다. 캡핑된 RNA에서, (캡핑된) RNA 분자의 제1 염기의 3' 위치는 포스포디에스테르 결합을 통해 RNA 분자의 후속 염기("제2 염기")의 5' 위치에 연결된다.
- [0237] 한 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 및 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열을 포함한다.
- [0238] 다른 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 프로모터, 적어도 하나의 비-구조 단백질을 암호화하는 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 및 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열을 포함한다.
- [0239] 또 다른 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 프로모터, 적어도 하나의 비-구조 단백질을 암호화하는 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열, 및 폴리A 꼬리를 포함한다.
- [0240] 다른 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 프로모터, 적어도 하나의 비-구조 단백질을 암호화하는 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열, 선택적으로 선천적 억제제 단백질을 암호화하는 각각의 적어도 하나의 서열 사이의 스페이서 서열, 및 폴리A 꼬리를 포함한다.
- [0241] 한 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 5' 캡, 프로모터, 적어도 하나의 비-구조 단백질을 암호화하는 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열, 선택적으로 선천적 억제제 단백질을 암호화하는 각각의 적어도 하나의 서열 사이의 스페이서 서열, 및 폴리A 꼬리를 포함한다.
- [0242] 한 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 5' 캡, 프로모터, 적어도 하나의 비-구조 단

백질을 암호화하는 서열, 스페이서, 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열 및 폴리A 꼬리를 포함한다.

[0243] 한 구현예에서, 상기 RNA 구조체는, 바람직하게는 5'에서 3'으로, 5' 캡, 프로모터, 적어도 하나의 비-구조 단백질을 암호화하는 서열, 스페이서, 선천적 억제제 단백질을 암호화하는 적어도 하나의 서열, 서브게놈 프로모터, 적어도 하나의 치료 생체분자를 암호화하는 서열, 스페이서 서열, 적어도 하나의 선천적 억제제 단백질을 암호화하는 서열, 선택적으로 적어도 하나의 선천적 억제제 단백질을 암호화하는 각각의 서열 사이의 스페이서 서열, 및 폴리A 꼬리를 포함한다.

[0244] 상기 PIV5의 V 단백질은 MDA5에 직접 결합하여 올리고머화를 방지하는 것으로 생각되지만, ORF4a는 dsRNA에 대한 PACT의 결합을 차단하는 것으로 생각된다. 본 발명자들은 상기 단백질을 암호화하는 유전자를 T2A 절단 부위에 의해 분리된 관심있는 유전자(GOI)(즉, 치료 생체분자) 다음에 saRNA 내로 통합시켜 선천적 감지 메커니즘을 유리하게 우회할 수 있는 구조체를 생성하였다. 발현시 내인성 프로테아제(T2A)에 의해 절단되는 단일 개방 해독 틀 내로 GOI 및 IIP를 짝을 이루는 것은 동일한 동역학으로 세팅된 비에서 동일한 세포 내에 이들의 발현을 최대화한다.

[0245] 따라서, 바람직하게는, 상기 RNA 구조체는, 5'에서 3'으로, 5' 캡, 51개 뉴클레오타이드의 보존된 서열 요소를 포함하는 프로모터, nsP1, nsP2, nsP3v, nsP4, 서브게놈 프로모터 26S, 치료 생체분자를 암호화하는 서열, T2A 스페이서 서열, PIV5 V 및/또는 MERS-CoV ORF4a를 암호화하는 서열 및 폴리A 꼬리를 포함한다.

[0246] 한 구현예에서, 따라서, 상기 RNA 구조체는 다음과 같은 서열번호 38을 포함하거나 이로 이루어질 수 있다:

[0247] AUGGGCGCGCAUGAGAGAAGCCAGACCAAUUACCCAAAAUGGAGAAAGUUCACGUUGACAUCGAGGAAGACAGCCAUUCCUCAGAGCUUUGCAGC  
GGAGCUUCCCGCAGUUGAGGUAGAAGCAAGCAGGUCACUGAUAAUGACCAUGCUAAUGCCAGAGCGUUUUCGAUCUUGGCUUCAAACUGAUCGAAACCG  
AGGUGGACCAUCCGACACGAUCCUUGACAUGGAAGUGCGCCCGCCGAGAAUGUAUUCUAAAGCACAAGUAUCAUUGUAUCUGUCCGAUGAGAUGUGCGG  
AAGAUCGGACAGAUUGUAUAGUAUGCAACUAAGCUGAAGAAAAACUGUAAGGAAUAACUGAUAGGAAUUGGACAAGAAAAUGAAGGAGCUGGCCGCCG  
UCAUGAGCGACCCUGACCUGGAAACUGAGACUAUGUGCCUCCACGACGACGAGUCGUGUCGUACGAAGGGCAAGUCGUGUUUACCAGGAUGUAUACGCGG  
UUGACGGACCGACAAGUCUCUAUACCAAGCCAAUAAAGGAGUUAAGAGUCGCCUACUGGAUAGGCUUUGACACCACCCUUUUAUGUUUAAGAACUUGGCUG  
GAGCAUAUCCAUAUACUUAACCAACUGGGCCGACGAAACCGUGUUAACGGCUCGUAAACUAGGCCUUAUGCAGCUCUGACGUUAUGGAGCGGUCACGUAGAG  
GGAUGUCCAUCUUAAGAAGAUAUUGAAACCAUCCAACAUAUGUUCUAUUCUGUUGGCUCGACCAUCUACCACGAGAAGAGGGACUUAUCAGAGGAGCU  
GGCACCUGCCGUCUGUAUUUACUUAACGUGGCAAGCAAAAUACACAUGUCGGUGUGAGACUAUAGUUAAGUUGCGACGGGUACGUCGUUAAAAAGAAUAGCUA  
UCAGUCCAGGCCUGUAUGGGAAGCCUUCAGGCUAUGCUGCUACGAUGCACCGCGAGGGAUUCUUGUGUGCAAAGUGACAGACACAUUGAACGGGGAGAGGG  
UCUCUUUCCCGUGUGCAGCUAUGUGCCAGCUACAUGUGGACCAAAUGACUGGCAUACUGGCAACAGAUUGCAGUGCGGACGACGCGCAAAACUGCUGG  
UUGGGCUCAACCAGCGUAUAGUCGUAACGGUCGACCCAGAGAAACACCAUAUACCAUGAAAAAUUACCUUUUGCCCGUAGUGGCCAGGCAUUGCUAGGU  
GGGCAAGGAUUAUAGGAAGAUCAAGAAGAUAAAGGCCACUAGGACUACGAGAUAGACAGUUAUGUAUGGGGUGUUGUUGGGCUUUUAGAAGGCACAAGA  
UAACAUCUAUUUAUAGCGCCCGGAUACCCAAACCAUACAUAAGGAGAACAGCGAUUCCACUCAUUCGUGCUGCCAGGAUAGGCAGUAACACAUUGGAGA  
UCGGGUGAGAACAAAGAUAGGAAAAUUGUUAAGGAGCACAAGGAGCGUCACCUCUAUUAACCGCGAGGACGUACAAGAAAGCUAAGUGCGCAGCCGAUG  
AGGCUAAGGAGGUGCGUAAGCCGAGGAGUUGCGCGCAGCUCUACCACCUUUGGCAGCUGAUUGUAGGAGGCCACUCUGGAAGCCGAGUGCAGUCUUAUGU  
UACAAGAGGUGGGCCGCGCUCAGUGGAGACACCUGUGGCUUGAUAAAGGUUACCAGCUACGAUGGCGAGGACAAGAUCGGCUCUUAACGUGUGCUUUCUC  
CGCAGGCUUACUACAAGAGUAAAAAUUUCUUGCAUCCACCCUCUCGUGAACAAGUUAUGAGUAUACACACUCUGGCCGAAAAGGGCGUUAUGCCGUGG  
AACCAUACCAUGGUAAAGUAGUGGUGCCAGAGGGACAUCAUACCCGUCAGGACUUAAGCUCUGAGUAAAGUGCCACCAUUGUGUACAACGAACGUG  
AGUUCGUAAACAGGUACCGCACCAUUAUGCCACACAUGGAGGAGCGUGAACACUGAUGAAGAAUUAUACAAAACUGUCAAGCCCAGCGAGCAGCAGCGG  
AAUACCUUACGACAUAGCAGGAAACAGUGCGUCAAGAAAGAAUAGUCACUGGGCUAGGGCUCACAGGCGAGCUGUGGAUCCUCCUCCAUGAAUUCG  
CCUACGAGAGUCUGAGAACGACGACCGCGUCCUUAACCAAGUACCAACCAUAGGGGUGUAUGGCGUGCCAGGAUCAGGCAAGUCUGGCAUUAUAAAGCG  
CAGUACCAAAAAAGAUCAUGUGGUGAGCGCAAGAAAGAAAACUGUGCAGAAAUUAUAAAGGACGUAAGAAAAUGAAAGGGCUGGACGUCAAUGCCAGAA  
CUGUGGACUCAGUCUCUUGAAUGAUGCAAAACACCCGUGAGAGACCCUGUAUUAUGACGAAGCUUUGCUUGUCAUGCAGGUACUCUACAGAGCGCUAUG  
CCAUAUAAGACCUAAAAAGGACAGUCUCUGCGGGAUCCAAACAGUGCGGUUUUUUAACAUGAUGUGCCUGAAAGUGCAUUUUAACACGAGAUUUGCA  
CACAAGUCUCCACAAAAGCAUCUCUGCCGUUGCACUAAAUCUGAGUUCGGUCGUCUACCCUUGUUUUAACGACAAAAAUGAGAACGACGAAUCCGA  
AAGAGACUAAGAUUGUGAUAGACAUACCGGAGUAACCAACCUAAGCAGGACGUAUCUAUUCACUUGUUUCAGAGGGUGGGUAGAGAGUUGCAAUAG  
AUUACAAAGGCAACGAAUUAUGACGCGAGCUGCCUCUAAGGGCUGACCCGUAAAGGUGUGUAUGCCGUUCGUAACAGGUGAAUAGAAAAUCCUCUGUACG  
CACCCACCUCAAGAUUGAAGCUUACUGACCCGACGAGGACCGCAUCGUGUGGAAAAACUAGCCGGCGACCCAUUGGAUAAAAACACUGACUGCCA  
AGUACCCUGGGAUUUACUGCCACGAUAGAGGAGUGGCAAGCAGAGCAUGAUGCCAUAUGAGGCACAUUUGGAGAGACCGGACCCUACCGACGUCUUC  
AGAAUAAAGGCAACGUGUGUGGCCAAGGCUUAGUGCCGGUGUGAAGACCGCUGGCAUAGACAUAGCCACUGAACAAUGAACACUGUGGAUUAUUUG  
AAACGGACAAGCUCACUCAGCAGAGAUAGUAUUGAACCAACUAUGCGUGAGGUUUUUGGACUCGUAUCUGGACUCCGGUUAUUUUGCACCCACUGUUC

CGUUAUCCAUAAGGAAUAAUCACUGGGUAACUCCCCGUCGCCUAACAUGUACGGGCGUAAUAAAGAAGUGGCCGUCAGCUCUCUCGAGGUACCCACAAC  
 UGCCUCGGGCGAGUUGCCACUGGAAGAGUCUAUGACAUGAACACUGGUACACUCGCGCAAUUAUGAUCCGCGCAUAAACCUAGUACCUGUAAACAGAAGACUGC  
 CUAUGCUUUAGUCCUCCACCAUAAUGAACCCACAGAGUACUUUUCUAUUCGUCAGCAAAUUGAAGGGCAGAACUGUCCUGGUGUCGGGGAAAAGU  
 UGUCCGUCACAGGCAAAUUGUAGCUGGUUGUCAGACCGGCCUGAGGCUACCUUCAGAGCUCGCGUGGAUUUAGGCAUCCAGGUGAUGUGCCCAAAUUG  
 ACAUAAUUAUUGUAAUGUGAGGACCCCAUUAUAAUACCAUACUAUCAGCAGUGUGAAGACCAUGCCAUUAAGCUUAGCAUGUUGACCAAGAAAGCUUGUC  
 UGCAUCUGAAUCCCGGGCGGAACCGUGUCAGCAUAGGUUAUGGUUACGUGACAGGGCCAGCGAAAGCAUCAUUGGUGCUAUAGCGCGGCAGUUAAGUUUU  
 CCCGGUUAUGCAAACCGAAAUCCUACUUGAAGAGACGGAAGUUCUGUUUGUAUUCAUUUGGUACGAUCGCAAGGCCGUAACGCACAAUUCUUAACAGCUUU  
 CAUCAACCUUGACCAACAUAUUUAACAGGUUCCAGACUCCACGAAGCCGAUGUGCACCUCUAUAUCAUGUGGUGCGAGGGGAUUAUUGCCACGGCCACCGAAG  
 GAGUGAUUAUAAUUGCUGCUAACAGCAAAGGACAACCUGGCGGAGGGGUGUGCGGAGCGCUGUAUAAAGAAUUCGCGAAAGCUUCGAUUUACAGCCGAUCG  
 AAGUAGGAAAAGCGCGACUGGUCAAAGGUGCAGCUAAACAUAUCAUUCAGCCGUAAGGACCAACUUAACAAGUUUCGAGGUGAAGGUGACAAACAGU  
 UGGCAGAGGCUUAUGAGUCCAUCGCUAAGAUUGUCAACGAUAAACAAUUAACAGCAGUAGCGAUUCCACUGUUGUCCACCGGCAUCUUUCCGGGAACAAAG  
 AUCGACUAACCCAAUCAUUGAACCAUUGCUGACAGCUUAGACACCACUGAUGCAGAUUAGCCAUUAUCUGCAGGGACAAGAAUUGGGAUAGACUCUCA  
 AGGAAGCAGUGGCUAGGAGAGAAGCAGUGGAGGAGAUUGCAUAUCCGACGACUUCUAGUGACAGAACCUGAUGCAGAGCUGGUGAGGGUGCAUCCGAAGA  
 GUUCUUUGGUGGAAGGAAGGGCUACAGCACAAGCGAUGGCAAAACUUCUCAUAUUGGAGGGACCAAGUUUACACAGGCGGCCAAGGAUUAAGCAGAAA  
 UUAUUGCCAUGUGGCCGUGUACAACGAGGCCAUGAGCAGGUUAUGCAUGUAUUCUUGGAGAAAGCAUGAGCAGUAUUAAGGUGCAAAUGGCCGUGCAAG  
 AGUCCGAAGCCUCCACACCACUAGCAGCUGCCUUGCUGCAUCCAUUGCAUCCAGAAAGAGUACAGCGCCUAAAAGCCUACGUCACGAGAACAAA  
 UUAUCUGUGCUCAUCCUUCCAUGGCCAAGUAUAGAUAUCACUGGUGUGCAGAAGAUCCAUUGCUCCAGCCUUAUUAUGUUCUACCGAAAGUGCCUGCGU  
 AUAUUAUCAAGGAAGUAUUCUGGAAACACCACCGGUAAGCAGAGACUCCGAGCCAUUGCGCAGAGAACC AUCCACAGAGGGGACACCUGAACAACCAC  
 CACUUAUAACCGAGGAUGAGACCAGGACUAGAACGCCUGAGCGCAUCAUCAUGAAGAGGAAGAAGAGGAUAGCAUAGUUUGCUGUCAGAUUGGCCGACCC  
 ACCAGGUGCUGAAGUCGAGGCAGACAUUCACGGGCCGCCUCUGUAUCUAGCUCAUCCUGGUCAUCCUCAUGCAUCCGACUUAUGUGGACAGUUUAU  
 CCAUACUUGACACCCUGGAGGGAGCUAGCGUGACCAGCGGGCAACGUCAGCCGAGACUAACUCUUAUCGCAAGAGUAUGGAGUUUCUGGCGCGACCCG  
 UGCCUGCGCCUCGAACAGUAUUCAGGAACCCUCCAUUCCGCUCCGCGCACAGAACACCGUCACUUGCACCAGCAGGGCCUGCUCGAGAACCAGCCUAG  
 UUUCCACCCCGCAGGCGUGAAUAGGGUGAUCACUAGAGAGGAGCUCGAGGCGCUUACCCGUCACGCACUCCUAGCAGGUGCGUCGAGAACCAGCCUGG  
 UCUCACACCCCGCAGGCGUAAUAGGGUGAUUACAAGAGAGGAGUUGAGGCGUUCGUAAGCACAACAACUAGCGUUUGAUGCGGGUGCAUACAUCUUUU  
 CCUCCGACACCGGUAAGGGCAUUUAACAACAAAAUCAGUAAGGCAACCGGUCUAUCCGAAGUGGUGUUGGAGAGGACCGAAUUGGAGAUUUCGUAUGCCC  
 CGCGCCUCGACCAAGAAAAAGAAGAUUACUACGCAAGAAAUACAGUUAUAAUCCACACCGUCUAACAGAAGCAGAUACCAGUCCAGGAAGGUGGAGAACA  
 UGAAAGCCAUAAACAGCUAGACGUAUUCUGCAAGGCCUAGGGCAUUAUUUGAAGGCAGAAAGGAAAGUGGAGUGCUACCGAACCCUGCAUCCUGUCCUUUGU  
 AUUCAUCUAGUGUGAACCGUGCCUUUAAGCCCAAGGUCGAGUGGAAGCCUGUAACGCCAUGUUGAAAGAGAACUUCGACUGGCUUCUUAUCUGUA  
 UUAUUCAGAGUACGAUGCCUAUUGGACAUUGGUGACGGAGCUUACUGCUGCUUAGACACUGCCAGUUUUGCCUGCAAAGCUGCGCAGCUUCCAAAGA  
 AACACUCCUAUUGGAACCCACAUAACGAUCGGCAGUGCCUUCAGCGAUCCAGAACACGCUCCAGAACGUCUCCGAGCUGCCACAAAAAGAAUUGCAAUG  
 UCACGCAAUAGAGAGAAUUGCCGUAUUGGAUUCGCGCGCCUUAAUGUGGAUUGCUUAAGAAAUUGCGUGUAAUAAUGAAUUAUGGGAACGUUAAAG  
 AAAACCCCAUCAGGCUUACUUAAGAAACGUGGUAUUUAUUAUACCAAAUUAUAAAGGACCAAAAGCUGCUCUUAUUGGCAAGACACAUAAUUGAAUA  
 UGUUGCAGGACAUACCAUUGGACAGGUUUGAAUGGACUUAAGAGAGACGUGAAAGUGACUCCAGGAACAAAACAUACUGAAGAACGGCCCAAGGUACAGG  
 UGAUCCAGGCGCCGUAUCCGUAAGCAACAGCGUAUCUGUGCGGAUCCACCGAGAGCUGGUUAGGAGAUUAAUUGCGGUCCUGCUUCCGAACAUUCAACAC  
 UGUUUUAUUGUGCGCUGAAGACUUGACGCUAUUAUAGCCGAGCUCUCCAGCCUGGGGAUUGUGUUCUGGAAACUGACAUCCGCGUGCUUUGAUAAGUG  
 AGGACGACGCAUUGGCUUGACCGCGUUAUUGAUUCUGGAAGACUUAAGGUGGAGCAGAGCUGUAGCGCUGAUUGAGGCGGCUUUGCGGCAAAUUAU  
 CAUAACAUUUGCCACUAAAACUAAAUUUAUUAUUCGGAGCCAUGAUGAAAUUGGAAUUGUCCUCACACUGUUGUGAACACAGUCAUUAACAUUGUAUUG  
 CAAGCAGAGUGUUGAGAGAACGGCUAACCGGAUACCAUGUGCAGCAUUCAUUGGAGAUACAUAUCGUGAAAGGAGUCAAUUCGGACAAAUUAUUGGCAG  
 ACAGGUGCGCCACCGUUGAAUUGGAAGUCAAGAUUAUAGAUGCUGUGGUGGGCGAGAAAGCGCCUUAUUCUGUGGAGGGUUUAUUUGUGUGACUCCG  
 UGACCGGCACAGCGUGCCGUGGCGAGACCCCUAAAAAGCGUGUUUAAAGCUUGGCAAAACCUUGGCAGCAGACGAUACAUGAUGACAGGAGAAGGG  
 CAUUGCAUGAAGAGUCAACACGUGGAACCGAGUGGUUAUUCUUCAGAGCUGUGCAAGGCAGUAGAAUACAGGUAUGAAACCGUAGGAACUCCAUAUAG  
 UUAUGGCCAUAGACUACUUGAGCAGUGUUAUUAUCAUUCAGCUACCUAGAGAGGGGCCCUUAUACUCUCUACGGCUAACCGUAUGGACUACGACAUAGU  
 CUAGUCCGCCAAGUCUAGCAUUGUAUGGCCACCAUGGAAGAUUGCCAAGAACAUAAGAAGGGCCUGCCCCAUUCUACCCCUUGGAAGUUGGAACAGCCGG  
 CGAGCAGCUGCACAAAGGCCAUGAAGAGUAACGCCUGGUGCCCGGCACAACGCCUUCACCGAUGCCCACUAGGAGGUGGACUACCUACGCCGAGUACU  
 CGAGAUGAGCGUGCGGCGUGGCCGAAGCUAUGAAGCGCUACGGCCUGAACCAACACCGGAUCGUCGUGCAGCGAGAACAGCCUGCAGUUCUUAUGCC  
 CGUGCUGGGCGCCUGUUUAUUGGAGUGGCGUGGCCCCUGCCAACGACAUCUACAACGAGCGGAGCUGCUGAACAGCAUGGGCAUCAGCCAGCCACCGU  
 GGUGUUCUGUGUCCAAAGAGGACUGCAGAAAAUCCUGAACGUGCAGAGAAGAGUGCCAUCAUCCAGAAAAUCAUCAUUGGACAGCAAGACCGACUACCA  
 GGGCUUCCAGAGCAUGUACACCUUCUGUACAGCCAUUCGCCCCUGGCUUACAGAGUACGACUUCUGGCCGAGAGCUUCGACCGGGACAAGACAAUCGC  
 CCUGAUCAUGAACAGCAGCGGACGACCGGACUCCUUAAGGCGUGGCCCCUGCCUCACAGAACUGCCUGCGUGCGGUUAGCCACGCCCGGACCCUACU  
 CGGAACAGAUCAUCCCGACACCGCAUCCUGAGCGUGGUGCCUUCACCACCGCUUCGGCAUGUUCACCACCCUGGGCUACCUGAUCUGCGGCUUCCG  
 GGUGGUGCUGAUGUACAGAUUCGAGGAAGAACUGUUCUGCGGAGCCUGCAGGACUACAAGAUCCAGAGCGCCUGCUGGUGCCUACCCUGUUCAGCUUCU

[0249] 다른 구현예에서, 상기 RNA 구조체는 다음과 같은 서열번호 39를 포함하거나 이로 이루어질 수 있다:

– 30 –

AGUACCCUGGAAUUCACUGCCACGAUAGAGGAGUGGCAAGCAGAGCAUGAUGCCAUCAGAGGCACAUUUGGAGAGACCGGACCCUACCGACGUCUCC  
 AGAAUAGGCAACGUGUGUGGGCCAGGCUUAGUGCCGGUGUGAAGACCGUGGCAUAGACAUGACCACUGAACAAUGGAACACUGUGGAUUUUUUG  
 AAACGGACAAGCUCACUCAGCAGAGAUAGUAUUGAACCAACUAGCGUGAGGUUCUUUGGACUCGAUCUGGACUCCGGUCUAAUUUCUGACCCACUGUUC  
 CGUUAUCCAUUAGGAUAAUACUGGGAAUACUCCCGUGCCUAAACUAGUACGGGUGAAUAAAGAAGUGGUCGUCAGCUCUUCGACAGGUACCCACAAC  
 UGCCUGGGCAGUUGCCACUGGAAGAGUCUAGACAUGAACACUGGUACACUGCGCAAUUAGAUCCGCGCAUAAACCUAGUACCGUAAACAGAACAGACUGC  
 CUCAGUCUUAGUCCUCCACCAUAAUGAACCCACAGAGUACUUUUCUACUUCGUCAGCAAAUUGAAGGGCAGAACUGUCCUGGUGUGCGGGAAAGU  
 UGUCCGUGCCAGGCAAAUGGUAGCUGGUUGUCAGACCGGCCUGAGGCUACCUUCAGAGCUCGGCUGGAUUUAGGCAUCCAGGUGAUGUGCCAAAUUG  
 ACAUAAUAAUUGUAAUGUGAGGACCCCAUAAUAAUACCAUACUACAGCAGUGAGAAGACCAUGCCAUAAGCUUAGCAUGUAGACCAAGAAAGCUUGUC  
 UGCAUCUGAAUCCCGGGCGAACCUGUGUCAGCAUAGGUUAUGGUUACGUCAGAGGGCCAGCGAAAGCAUCAUUGGUGCUAUAGCGGGCAGUUAAGUUU  
 CCCGGUAGUGCAAACCGAAAUCCUACUUGAAGAGACGGAAGUUCUGUUUGUAUUCUAGGGUACGAUCGCAAGGCCCGUACGCACAAUUCUACAAGCUU  
 CAUCAACCUUGACCAACAUUUUACAGGUUCCAGACUCCACGAAGCCGAUGUGCACCUCAUUAUAGUGGUGCGAGGGGAUUAUGCCACGGGCCACCGAAG  
 GAGUGAUUAAUAGCUGCUAACAGCAAAGGACAACCUGGCGGAGGGGUGUGCGGAGCGCUGUAUAGAAAUUCCGGAAAGCUUCGAUUUACAGCCGAUCG  
 AAGUAGGAAAGCGCAGCUGGUCAAAGGUGCAGCUAAACAUUACUAGCCGUGAGGACCAACUUAACAAGUUUCGAGGUUGAAGGUGACAACAGU  
 UGGCAGAGGCUUAGAGUCCAUUCGCUAAGAUUGUCAACGAUAAUUAAGUACAGUAGCGAUUCCACUGUUGUCCACCGGCAUCUUUCCGGGAACAAAG  
 AUCGACUAACCCAAUUAUGAACCAUUGCUGACAGCUUAGACACCACUGAUGCAGAUGAGCCAUAUACUGCAGGGACAAGAAUUGGAAUAGACUCUCA  
 AGGAAGCAGUGGCUAGGAGAGAAGCAGUGGAGGAGAUUGCAUUAUCCGACGACUUCUAGUGACAGAACCUGAUGCAGAGCUGGUGAGGGUGCAUCCGAAGA  
 GUUCUUUGGUGGAAGGAAGGGCUACAGCACAAGCGAUGGCAAAACUUCUACUUAUUGAAGGGACCAAGUUCACCAGGCGGCCAAGGAUUAAGCAGAAA  
 UUAAGGCCAUUGUGCCCGUUGCAACGGAGGCCAUGAGCAGGUUAUGCAUGUAUUAUCCUGGAGAAAGCAUGAGCAGUAUAGGUGCAAAUGCCCGUGCAAG  
 AGUGGGAAGCCUACACACCACUAGCAGCUGCCUUGCUGCAUCCAUUGCAUAGCAGAAAGAGUACAGCGCCUAAAGGCCUACGUCACGAGAACAAA  
 UUACUGUGUCUACCUUCCAUUGCCGAAGUAUAGAUAUCACUGGUGUGCAGAAGAUCCAUUGCUCCAGCCUUAUUGUUCUACCGAAAGUGCCUGCGU  
 AUAUUAUCCAAGGAAGUAUCUGGUGAAACACCACCGUAGACGAGACUCCGGAGCCAUCGGCAGAGAACCAUCCACAGAGGGGACCCUGAACACCCAC  
 CACUUAUAAACCGAGGAUGAGACCAGGACUAGAACGCCUGAGCCGAUCAUCAUGAAGAGGAAGAAGAGGAUAGCAUAGUUUGCUGUCAGAUUGCCCGACCC  
 ACCAGGUGCUGAAGUCGAGGCAGACAUUACGGGCGGCCUCUGUAUCUAGCUCUACUCCUGGUCAUCCUCAUGCAUCCGACUUAUGUGGACAGUUUAU  
 CCAUACUUGACACCCUGGAGGGAGCUGAGCUGACCGGGGCAACGUCAGCCGAGACUAAUCUUAUUCGCAAGAGUAUGGAGUUUCUGGCGCAGCCGG  
 UGCCUGCGCCUCGAACAGUAUUCAGGAACCCUCCAUUCCGCUCCGCGCACAGAACACCGUCACUUGCACCAGCAGGGCCUGCUGAGAACCAGCCUAG  
 UUUCCACCCCGCCAGGCGUGAAUAGGGUGAUCACUAGAGAGGAGCUCGAGGCGCUUACCCCGUCACGCACUCCUAGCAGGUCGGUCUCGAGAACCAGCCUGG  
 UCUCAACCCCGCCAGGCGUAAUAGGGUGAUUACAAGAGAGGAGUUUGAGGCGUUCGUGACACAACAACUAGCGGUUUGAUGCGGGUGCAUACAUCUUU  
 CCUCCGACACCGGUAAGGGCAUUUACAACAAAAUCAGUAAGGCAACCGGUCUUAUCCGAAGUGGUGUGGAGAGGACCGAAUUGGAGAUUUCGUAUGCCC  
 CGCGCCUGACCAAGAAAAAGAAAUUACUACGCAAGAAAUACAGUUAUAAUCCACACCGUCUAAAGAGCAGAUACAGUCCAGGAAGGUGGAGAACA  
 UGAAAGCCAUAAACAGCUAGACGUUAUUCGCAAGGCCUAGGGCAUUAUUGAAGGCAAGGAAAGUGGAGUGCUACCGAACCCUGCAUCCUGUCCUUUGU  
 AUUCAUCUAGUGUAGCCGUGCCUUUACAAGCCCCAAGGUCGAGUGGAAGCCUGUAACGCCAUGUUGAAGAGAACUUAUCCGACUGUGGCUUCUACUGUA  
 UUAUCCAGAGUACGAUGCCUUAUUGGACAUUGGUGACGGAGCUUACUGCUGCUUAGACACUGCCAGUUUUGCCUGCAAAGCUGCGCAGCUUCCAAAGA  
 AACACUCCUUAUUGGAACCCACAUAACGAUCGGCAGUGCCUUCAGCGAUCCAGAACACGCUCCAGAACGUCUGGCGAGUGCCACAAAAGAAUUGCAAUG  
 UCACGCAAUAGAGAGAAUUGCCGUAUUGGAUUCGCGCGCCUUAAUGUGGAUUGCUUAAGAAAUUGCGUGUAUAAUAGAAUUAUGGGAACGUAUAAAG  
 AAAACCCCAUCAGGCUUACUGAAGAAACGUGGUAUUUACAUAUCCAAAUUAAAGGACCAAAAGCUGCUGCUCUUUUGCAAGACACAUAAUUGAAUA  
 UGUUGCAGGACAUACCAUUGGACAGGUUUGAAUUGGACUUAAGAGAGACGUGAAAGUACUCCAGGAACAAAACAUACUAGAAGACGGCCCAAGGUACAGG  
 UGAUCCAGGUGCCGAUCCGCUAGCAACAGCGUAUCUGUGCGGAUCCACCGAGAGCUGGUUAGGAGAUUAAUUGCGGUCCUGCUUCCGAACAUCAUACAC  
 UGUUUGAUUAGUGCGGUGAAGACUUGAGCGCUAAUUAAGCCGAGCACUCCAGCCUGGGGAUUGUGUUCUGGAAACUGACAUCGCGUGCUUUGAUAAGUG  
 AGGACGACGCCAUGGCUUGACCCGCUAAUUGAUUUGGAAGACUAGGUGUGGACGAGCUGUUGACGCGUAUUGAGGCGGCUUCCGCGCAAAUUCU  
 CAUAUUAUUGCCCAUAAACUAAAUUAAUUAUCCGAGGCAUGAUGAAUUCUGGAUUGUCCUACACUGUUGUGAACACAGUCAUUAACAUUGUAUUGC  
 CAAGCAGAGUGUUGAGAGAACGGCUAACCGGAUACCAUGUGCAGCAUUCUUGGAGAUACAUAUCUGGAAAGGAGUCAAUCCGACAAAUUAAUGGCAG  
 ACAGGUGCGCACCGUGUAGAUUGGAAGUCAAGAUUAUAGAUGCUGUGGUGGGCGAGAAAGCGCCUUAUUCUGUGGAGGGUUUAUUUGUGUGACUCCG  
 UGACCGGCACAGCGUGCCGUGUGGACAGCCCCUAAAAAGGCGUUAUAGCUUGGCAACCUUGGCAGCAGACGAUGAACAUAGUAGACAGGAGAAGGG  
 CAUUGCAUGAAGAGUACAACGUGGAACCGAGUGGGUAUUCUUCAGAGCUGUGCAAGGCAUGAAGUAACAGGUAUGAAACCGUAGGAACUCCAUAUAG  
 UUAUGGCCAUAGACUACUAGCUGAGCAGUGUUAUUAUACUACGCUACCUAGAGAGGGGCCCUUAUACUCUCUACGGCUAACCGUAUGGACUACGACAUAGU  
 CUAGUCCGCCAAGUCUAGCAUAGUAGGCCACCAUGGAAGAUUGCAAGAACAUAAGAAGGGCCUGCCCCAUUUAUCCCGUGGAAGUAGGAACAGCCGG  
 CGAGCAGCUGCACAGGCCAUGAAGAGUACGCCUGGUGCCCGGCACAACGCCUUCACCGAUGCCCAUUCGAGGUGGACAUACCUACGCCGAGUACU  
 CGAGAUAGCGUGCGGUGGCCGAGCUAUGAAGCGCUACGCCUGAACCAACACCGGAUCGUCGUGGAGCGAGAACAGCCUGCAGUUCUUAUGCC  
 CGUGCUGGGCGCCUGUUUAUUGGAGUGGUGUGGCCCGGCCAACGACAUUCUACAGCGCGAGCUGCUGAACAGCAUGGGCAUCAGCCAGCCACCGU  
 GGUGUUCGUGUCCAAAGAGGACUGCAGAAAAUCCUGAACGUGCAGAGAAGGUGCCCAUUAUCCAGAAAAUCAUCAUAGGACAGCAAGACCGACUACCA  
 GGGCUUCCAGAGCAUGUACACCUUCUGUAGCCAGCAUCUGCCCCUGGCUUACAGAGUACGACUUCUGGCCGAGAGCUUCGACCGGACAAGACAUCG

[0254] ATGGGCGGCGCATGAGAACAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTACGTTGACATCGAGGAAGACAGCCATTCTCTCAGAGCTTTGACGCGGAGCTTCCCCGAGTTTGAGGTAGAAGCCAAGCAGGTCAGTGATAATGACCATGCTAATGCCAGAGCGTTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCGCCCGCAGAATGTATTCTAAGCACAAGTATCATTGTATCTGTCCGATGAGATGTGCGG AAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAACTGTAAGGAAAAATACTGATAAGGAATTGGACAAGAAAAATGAAGGAGCTGGCCGCCG TCATGAGCGACCCGTGACCTGGAACTGAGACTATGTGCCTCCACGACGACGAGTCGTGTCTGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGG TTGACGGACCGACAAGTCTCTATCACCAGCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAAGAACTGGCTG GAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCTGACGTTATGGAGCGGTCACGTAGAG GGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTGAGGAGCT GGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAAATTACACATGTTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTA TCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGTCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGG TCTCTTTTCCCGTGTGCACGTATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAAGTCGGACGACGCGCAAAAACGTCTGG TTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACACCAATACCATGAAAAATTACCTTTTGGCCGTAGTGGCCAGGCATTTGCTAGGT GGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGA TAACATCTATTATAAGCGCCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTACTATTCTGTCTGCCAGGATAGGCAGTAACACATTGGAGA TCGGGCTGAGAAACAAGAATCAGGAAAAATGTTAGAGGAGCACAAAGGAGCCGTACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCAGCCGATG AGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCACTCTGGAAGCCGATGTCGACTTGATGT TACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTC CGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGG AACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAGGACTTTCAGCTCTGAGTGAAGTGCCACCATTGTGTACAACGAACGTG AGTTTCGTAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAACCTGTCAAGCCAGCGAGCAGCAGCGG AATACCTGTACGACATCGACAGGAAACAGTGGCTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGATCCTCCCTTCCATGAATTCC CCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTTAAAGCG CAGTCACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATATAAGGGACGTCAAGAAAAATGAAAGGGCTGGACGTCAATGCCAGAA CTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCTGTATATTGACGAAGCTTTTGCTTGTGTCATGCAGGTACTCTCAGAGCGCTCATAG CCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAACAGTGCGGTTTTTTTAAACATGATGTGCCTGAAAGTGCATTTTAAACCACGAGATTTGCA CACAAGTCTTCCACAAAAGCATCTCTCGCGTTCGACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGA AAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTATTCTCACTTGTTCAGAGGGTGGGTGAAGCAGTTGCAAATAG ATTACAAAAGGCAACGAAATATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCCGTACAAGGTGAATGAAAAATCCTCTGTACG

CACCCACCTCAGAACATGTGAACGTCTACTGACCCGCACGGAGGACCGCATCGTGTGAAAACTAGCCGGCGACCCATGGATAAAAACTGACTGCCA  
 AGTACCTGGGAATTTACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGACCGGACCCTACCGACGTCTTCC  
 AGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTG  
 AAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCACTATGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTCTGCACCCACTGTTT  
 CGTTATCCATTAGGAATAACTACTGGGATACTCCCGTCGCCTAACATGTACGGGTGAATAAAGAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAAC  
 TGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAACAGAAAGACTGC  
 CTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTGCTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGAAAAAGT  
 TGTCCGTCCAGGCAAAATGGTTGACTGGTTGTGACACCGGCTGAGGCTACCTTCAGAGCTCGGCTGGATTTAGGCATCCAGGTGATGTGCCAAATATG  
 ACATAATATTTGTTAATGTGAGGACCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGCTTGT  
 TGCATCTGAATCCCGGGGAACCTGTGTGACATAGGTTATGGTTACGTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGACGTTCAAGTTTT  
 CCCGGTATGCAAAACGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTGTATTTCATTGGGTACGATCGCAAGGCCGTACGCACAATCTTACAAGCTTT  
 CATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGATGTGCACCTCATATCATGTGGTGCAGGGGATATTGCCACGGCCACCGAAG  
 GAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATCCCGAAAGCTTCGATTACAGCCGATCG  
 AAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATTTCGCGTAGGACCAAACTTCAACAAAGTTTCGAGGTTGAAGGTGACAAACAGT  
 TGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTCCGGGAACAAAG  
 ATCGACTAACCAATCATTGAACCATTTGTGTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAATGACTCTCA  
 AGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGA  
 GTTCTTTGGTGGAAGGAAGGGTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGACCAAGTTTCACAGGCGGCCAAGGATATAGCAGAAA  
 TTAATGCCATGTGGCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCTCGGAGAAAGCATGAGCAGTATTAGTTCGAAATGCCCGTCGAAG  
 AGTCGGAAGCCTCCACACCACCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCTAAAAGCCTCACGTCCAGAACAAA  
 TTACTGTGTGCTCATCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGAGAAGATCCAAATGCTCCAGCCTATATTGTTCTACCGAAAGTGCTGCGT  
 ATATTTCATCAAGGAAGTATCTCGTGGAACACCACCGGTAGACGAGACTCCGAGCCATCGGCAGAGAACCAATCCACAGAGGGGACACCTGAACAACCAC  
 CACTTATAACCGAGGATGAGACCAGGACTAGAACGCTGAGCCGATCATCATGAAGAGGAAGAAGAGGATAGCATAAGTTTGTGTGATGAGGCGGACCC  
 ACCAGGTGTGCAAGTCGAGGCAGACATTCACGGGCGCCCTCTGTATCTAGCTCATCTGGTCCATTCTCATGCATCCGACTTTGATGTGGACAGTTTAT  
 CCATACTTGACACCTGGAGGGAGTACGCTGACGAGCGGCAACGTCAGCCGAGACTAACTCTTACTTCGCAAAAGAGTATGGAGTTTCTGGCGGACCGG  
 TGCCTGCGCTCGAACAGTATTACAGGAACCTCCACATCCGCTCCGCGCACAAGAACACCGTCACTTGACCCAGCAGGGCCTGCTCGAGAACCAGCCTAG  
 TTTCCACCCCGCCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGGTCCGGTCTCGAGAACCAGCCTGG  
 TCTCCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAAGCGTTCTGTAGCACAACAACATGACGGTTTGTGCGGGTGATACATCTTTT  
 CCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAACCGTGCTATCCGAAGTGGTGTGGAGAGGACCGAATTGGAGATTTCTGTATGCC  
 CGCGCTCGACCAAGAAAAAGAAATTAACGCAAGAAATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACAGTCCAGGAAGGTGGAGAACA  
 TGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCTGCATCTGTCTCTTGT  
 ATTCATCTAGTGTGAACCGTGCTTTTCAAGCCCCAAGTCGAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAATTTCCGACTGTGGCTTCTTACTGTA  
 TTATTCCAGAGTACGATGCTATTGAGCATGGTTGACGGAGCTTATGCTGCTTAGACACTGCCAGTTTTCGCCCTGCAAGCTGCGCAGCTTTCCAAAGA  
 AACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCTTCAGCGATCCAGAACACGCTCCAGAACGCTCCTGGCAGCTGCCACAAAAAGAAATTGCAATG  
 TCACGCAATGAGAGAATTGCCCGTATTGGATTGCGCGGCTTTAATGTGGAATGCTTCAAGAAATATGCGTGTAAATGAATATTGGGAAACGTTTAAAG  
 AAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGAGCAAAAAGCTGCTGCTCTTTTTCGCAAGACACATAATTGAAATA  
 TGTTCAGGACATACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGAGCTGAAAGTGAAGTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG  
 TGATCCAGGTGCGGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAAATGCGGTCTGCTTCCGAACATTCATACAC  
 TGTTTGATATGTGCGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGATTGTGTTCTGGAAGTACATCGCGCTGTTGATAAAAGTG  
 AGGACGACGCCATGGCTCTGACCGGTTAATGATTCTGGAAGACTTAGGTGTGGACGAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTTCAT  
 CAATACATTTGCCCACTAAAATAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCTCACTGTTTGTGAACACAGTCATTAACATTGTAATCG  
 CAAGCAGAGTGTTGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTTCATTGGAGATGACAATATCGTGAAAGGAGTCAAAATCGGACAAATTAATGGCAG  
 ACAGGTGCGCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCTTATTTCTGTGGAGGGTTATTTTGTGTGACTCCG  
 TGACCGGCACAGCGTGCCGTGTGGCAGACCCCTAAAAAGGCTGTTTAAAGCTTGGCAAACTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGG  
 CATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGAAGGCAGTAGAATCAAGGTATGAACCGTAGGAACCTCCATCATAG  
 TTATGGCCATGACTACTAGTACGAGTGTAAATCATTCAGTACCTGAGAGGGGCCCTATAACTCTCTACGGCTAACCTGAATGGACTACGACATAGT  
 CTAGTCCGCCAAGTCTAGCATATGTATGGCCACCATGGAAGATGCCAAGAATCAAGAAGGGCCCTGCCCCATTCTACCCCTGGAAGATGGAACAGCCGG  
 CGAGCAGCTGCACAAGGCCATGAAGAGATACGCCCTGGTGCCCGGCACAATCGCCTTCACCGATGCCACATCGAGGTGGACATCACCTACGCCGAGTACTT  
 CGAGATGAGCGTGCGGTGGCCGAAGCTATGAAGCGCTACGGCTGAACCAACACCGGATCGTCGTGTGACGCGAGAACAGCCTGCAGTTCCTCATGCC  
 CGTGCTGGGCGCCCTGTTTATCGGAGTGGCTGTGGCCCTGCCAACGACATCTACAACGAGCGGAGCTGCTGAACAGCATGGGCATCAGCCAGCCACCGT  
 GGTGTTGCTGTCCAAGAAGGGACTGCAGAAAATCCTGAACGTGCAGAAGAAGCTGCCCATCATCCAGAAAATCATCATGACAGCAAGACCGACTACCA

GGGCTTCCAGAGCATGTACACCTTCGTGACCAGCCATCTGCCCCCTGGCTTCAACGAGTACGACTTCGTGCCCCGAGAGCTTCGACCGGACAAGACAATCGC  
CCTGATCATGAACAGCAGCGGCAGCACCAGGACTGCCTAAAGGCGTGCCCTGCCTCACAGAACTGCCTGCGTGCGGTTTAGCCACGCCCCGGGACCTATCTT  
CGGCAACCAGATCATCCCCGACACCGCCATCCTGAGCGTGGTGCCTTCCACCACGGCTTCGGCATGTTACACCCTGGGTACCTGATCTGCGGCTTCG  
GGTGGTGCTGATGTACAGATTCGAGGAAGAACTGTTCTGCGGAGCCTGCAGACTACAAGATCCAGAGCGCCCTGCTGGTGCCTACCCTGTTACGCTTCTT  
TGCCAAGAGCACCCTGATCGATAAGTACGACCTGAGCAACCTGCACGAGATCGCCTTGGCGGAGCCCCCTGTCTAAAGAAGTGGGAGAGGCCGTGGCCAA  
GCGGTTCATCTGCCTGGCATCAGACAGGGCTATGGCCTGACCGAGACAACAGCGCCATTCGTATCACCCCCGAGGGCGACGATAAGCCTGGCGCCGTGGG  
AAAGGTGGTGCCATTCTTCGAGGCCAAGGTGGTGGACCTGGACACCGCAAGACACTGGGCGTGAACCAGAGGGGCGAACTGTGTGTGCGGGGACCTATGAT  
CATGAGCGGTACGTGAACAACCCGAGGCCACCAACGCGCTGATTGACAAGGATGGCTGGCTGCACAGCGGCACATTGCTACTGGGACGAGGACGAGCA  
CTTCTTCATCGTGGACCGGTGAAGTCCCTGATCAAGTACAAGGGCTACCAGGTGGCCCCAGCCGAGCTGGAATCTATCTGCTGCAGCACCCCAACATCTT  
CGATGCCGCGTGGCAGGACTGCCCGATGATGATGCTGGCGAAGTGCACGCGCTGTGGTGGTGTGGAACACGGAAGAACCATGACCGAGAAAGAAATCGT  
GGACTACGTGGCCAGCCAAGTGACACCGCCAAGAACTGAGAGGCGCGTGGTGTGTTGTGGACGAGGTGCCAAAGGGCTGACAGGCAAGCTGGACGCCCCG  
GAAGATCAGAGAGATCCTGATTAAGGCCAAGAAAGGCGCAAGATCGCCGTGGATCGGAGAAAGAGAGGCTCTGGCGAAGGCAGAGGACGCTGCTTACATG  
TGGCGAGCTGGAAGAGAACCCCGGACCTATGGACCTACCAGCTGAGCTTCAGCCCCGACGAGATCAACAAGCTGATCGAGACAGGCTGAACACCGTGGA  
ATACTTACCAGCCAGCAAGTGACCGGCACAAAGCAGCCTGGGCAAGAACAATTCCTCCAGGCGTGACCGGCTGCTGACAAATGCTGCCGAGGCCAAGAT  
CCAAGAGAGCACAACCACCAGAAGGGCTCTGTTGGAGGCGGAGCCAAGCCTAAGAAGCCAGACCTAAGATCGCCATCGTGGCCCGGACGATAAGACAGT  
GCCTGGCAAGCCCATTCCTAATCCTCTGCTGGGCTCGACAGCACCCCTAGCACACAGACAGTGTGGATCTGAGCGGCAAGACACTGCCTAGCGGCAGCTA  
TAAGGCGGTGAAGCTGGCCAAGTTCGGCAAAGAAAACCTGATGACCCGTTTCATCGAGGAACCCAGAGAGAACCCTATCGCCACCAGCTCTCCATCGACTT  
CAAGAGAGGCAGAGACACCGCGGCTTCCACAGAAGAGAGTACAGCATTGGCTGGGTGGGAGATGAAGTGAAGTGACCGAGTGGTGAACCCAGCTGCAG  
CCCTATTACAGCCGCGCTAGAAGATTGAGTGCACCTGTACCAGTGTCTGTGACCTGTAGCGAGTGCAGCGGGACACATGATGAGCGGCCGGAATTG  
GCAAGCTGCTTACATAGAACTCGCGGCGATTGGCATGCCGCTTAAAAATTTTATTTTATTTTCTTTTCCGAATCGGATTTGTTTTAAATATTC  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAA [서열번호 40]

[0255] 따라서, 바람직하게는 상기 핵산 서열은 실질적으로 서열번호 40에 나타난 것과 같은 뉴클레오티드 서열, 또는  
그의 단편 또는 변이체를 포함한다.

[0256] 한 구현예에서, 상기 핵산 서열은 다음과 같은 서열번호 41을 포함하거나 이로 이루어질 수 있다:

[0257] ATGGGCGGCGCATGAGAGAAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATCCTCAGAGCTTGCAGC  
GGAGCTTCCCGAGTTTGAGGTAGAAGCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAGCGTTTCGCATCTGGCTTCAAACTGATCGAAACGG  
AGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCGAGAAATGTATTCTAAGCACAAGTATCATTGTATCTGTCGATGAGATGTGCGG  
AAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAAATGAAGGAGCTGGCCCGG  
TCATGAGCGACCCCTGACCTGGAAGTGAAGTATGTGCTCCACGACGACGAGTGTGCTGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGG  
TTGACGGACCGACAAGTCTCTATACCAAGCCAATAAGGGAGTTAGAGTCGCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAAGAACTGGCTG  
GAGCATATCCATCATACTCTACCAACTGGGCGGACGAAACCGTGTAAACGGCTCGTAACATAGGCTATGCAGCTTGACGTTATGGAGCGGTACGTAAG  
GGATGTCCATTCTTAGAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGGACTTACTGAGGAGCT  
GGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTGCGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAGAATAGCTA  
TCAGTCCAGGCTGTATGGGAAGCCTCAGGCTATGCTGTACGATGCACCGCGAGGGATTTCTGTGCTGCAAGTGCAGACACATTGAACGGGGAGAGGG  
TCTCTTTTCCGTGTGACGATGTGTCAGCTACATTGTGTGACCAATGACTGGCATACTGGCAACAGATGTGATGCGGACGACGCGCAAAACTGCTGG  
TTGGGCTCAACCAGCGTATAGTCGTCAACGTCGCACCCAGAGAAACACCAATACCATGAAAAATTACCTTTTGCCCCGTAGTGGCCAGGCATTGCTAGGT  
GGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAGGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGTGTGTTGGGCTTTTAGAAGGCACAAGA  
TAACATCTATTTATAAGCGCCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTACTATTCGTGCTGCCAGGATAGGCAGTAACACATTGGAGA  
TCGGGCTGAGAACAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCAGCCGATG  
AGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCACTCTGGAAGCCGATGTCGACTTGTGT  
TACAAGAGGTGGGGCCGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTC  
CGCAGGCTGTAAGAGTGAAAAATTATCTTGCATCCACCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGG  
AACCATAACATGGTAAAGTAGTGGTGCAGAGGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATTTGTGTACAACGAACGTG  
AGTTCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAAGTGTCAAGCCAGCGAGCAGCAGCGG  
AATACCTGTACGACATCGACAGGAAACAGTGCCTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCG  
CCTACGAGAGTCTGAGAACACGACGACCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAGCG  
CAGTACCAAAAAAGATCTAGTGGTGAAGCGCAAGAAAGAAACTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAGGGCTGGACGTCAATGCCAGAA  
CTGTGGAAGTCACTGCTTGAATGGATGCAAACACCCCGTAGAGACCTGTATATTGACGAAGCTTTTGTGCTGATGAGGTAAGTCTCAGAGCGCTCATAG  
CCATTATAAGACCTAAAAAGGCAGTGTCTGCGGGGATCCCAACAGTGCAGTGTGCTGAAAGTGCAATTTAAACACGAGATTTGCA

CACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTACGACAAAAAATGAGAACGACGAATCCGA  
AAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAG  
ATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAACTCTGTACG  
CACCCACCTCAGAACATGTGAACGTCTACTGACCCGCACGGAGGACCGCATCGTGTGAAAAACACTAGCCGGCGACCCATGGATAAAAACTGACTGCCA  
AGTACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGACCGGACCTACCGACGTCTTCC  
AGAATAAGGCAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACATGGAACACTGTGGATTATTTTG  
AAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTCTGCACCCACTGTTT  
CGTTATCCATTAGGAATAACTCTGGGATAACTCCCGTCGCCTAACATGTACGGGTGAATAAAGAAGTGGTCCGTGAGTCTCTCGCAGGTACCCACAAC  
TGCCTCGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGC  
CTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTGTCAGCAAATTGAAGGGCAGAACTGTCTGGTGGTGGGGGAAAAGT  
TGTCCGTCCAGGCAAAATGGTTGACTGGTTGTGACACCGGCTGAGGCTACCTTCAGAGCTCGGCTGGATTAGGCATCCAGGTGATGTGCCAAATATG  
ACATAATATTTGTTAATGTGAGGACCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGCTTGT  
TGCATCTGAATCCCGGGCAACCTGTGTGACATAGGTTATGGTTACGTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGCAGTTCAAGTTTT  
CCCGGTATGCAAAACGAAATCTCACTGAAGAGACGGAAGTTCTGTTGTATTTCATTGGGTACGATCGCAAGGCCCGTACGCACAATTCTTACAAGCTTT  
CATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGATGTGCACCTCATATCATGTGGTGGCAGGGGATATTGCCACGGCCACCGAAG  
GAGTGATTATAAATGTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATCCCGGAAAGCTTCGATTACAGCCGATCG  
AAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATTTCATGCGGTAGGACCAAACTTCAACAAAGTTTCGGAGGTGAAGGTGACAAACAGT  
TGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTCCGGGAACAAAG  
ATCGACTAACCAATCATTGAACATTTGTGTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAATGACTCTCA  
AGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGA  
GTTCTTTGGTGGAAAGGGTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGACCAAGTTTCACCGGCGGCCAAGGATATAGCAGAAA  
TTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCTCGGAGAAAGCATGAGCAGTATTAGTTCGAAATGCCCGTCGAAG  
AGTCCGAAGCCTCCACACCCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCAGAACAAA  
TTACTGTGTCTCATCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAGAAGATCCATGTCTCCAGCCTATATTGTTCTACCGAAAGTGCTGCGT  
ATATTTCATCAAGGAAGTATCTCGTGGAAACACCACCGGTAGACGAGACTCCGAGCCATCGGCAGAGAACCAATCCACAGAGGGGACACCTGAACAACCAC  
CACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTTGTCTGATGCCCCGACCC  
ACCAGGTGCTGCAAGTCGAGGACGACATTCACGGGCGCCCTCTGTATCTAGCTCATCTGGTCCATTCTCATGATCCGACTTTGATGTGGACAGTTTAT  
CCATACTTGACACCTGGAGGGAGTACGCTGACAGCGGGCAACGTGAGCCGAGACTAACTCTTACTTCGCAAGAGTATGGAGTTTCTGGCGCGACCGG  
TGCCTGCGCCTGCAACAGTATTACAGAACCCCTCCACATCCCGCTCCGCGCACAAGAACCCGTCACTTGACCCAGCAGGGCTGCTCGAGAACCAGCCTAG  
TTTCCACCCCGCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGGTGCGTCTCGAGAACCAGCCTGG  
TCTCCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAAGCGTTTCGTAGCACAACAACATGACGGTTTGATGCGGGTGCATACATCTTTT  
CCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAACCGTGCTATCCGAAGTGGTGTGGAGAGGACCGAATTGGAGATTTCGTATGCC  
CGCGCTCGACCAAGAAAAGAAAGTACTACGCAAGAAATACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACAGTCCAGGAAGGTGGAGAACA  
TGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCTGCATCTGTCTCTTGT  
ATTCTATAGTGTGAACCGTGCTTTTCAAGCCCAAGGTGCGAGTGAAGCCTGTAAAGCCATGTTGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTA  
TTATTCCAGAGTACGATGCTATTGAGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTGTCCCTGCAAGCTGCGCAGCTTTCCAAAGA  
AACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGAACGCTCTGGCAGCTGCCACAAAAGAAATTGCAATG  
TCACGCAATGAGAGAATTGCCCGATTGGATTGCGCGCCTTAAATGTGGAATGCTTCAAGAAATATGCGTGTAAATGAATATTGGGAAACGTTTAAAG  
AAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGAGCAAAAAGCTGTGCTCTTTTGGCAAGACACATAATTTGAATA  
TGTTGCAGGACATACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGAGCTGAAAGTACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGG  
TGATCCAGGTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAAATGCGGTCTGCTTCCGAACATTCATACAC  
TGTTTGATATGTGCGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGATTGTGTTCTGGAAACTGACATCGCGTCTGTTGATAAAAGTG  
AGGACGACGCCATGGCTCTGACCGGTTAATGATTCTGGAAGACTTAGGTGTGGACGACAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTTCAT  
CAATACATTTGCCCACTAAAACATAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCAGCTGTTTGTGAACACAGTCATTAACATTGTAATCG  
CAAGCAGAGTGTTGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGAAAGGAGTCAATTCGGACAAATTAATGGCAG  
ACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTATTTTGTGTGACTCCG  
TGACCGGCACAGCGTGCCGTGTGGCAGACCCCTAAAAAGGCTGTTTAAAGCTTGGCAAACTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGG  
CATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGAAGGCAGTAGAATCAAGGTATGAACCGTAGGAACCTCCATCATAG  
TTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACGTACCTGAGAGGGGCCCTATAACTCTCTACGGCTAACCTGAATGGACTACGACATAGT  
CTAGTCCGCCAAGTCTAGCATATGTATGGCCACCATGGAAGATGCCAAGAACATCAAGAAGGGCCCTGCCCCATTCTACCCCTGGAAGATGGAACAGCCGG  
CGAGCAGTGCACAAGGCCATGAAGAGATACGCCCTGGTGCCCGCACAAATCGCCTTACCGATGCCACATCGAGGTGGACATCACTACGCCGAGTACTT

CGAGATGAGCGTGGCGGTGGCCGAAGCTATGAAGCGCTACGGCCTGAACACCAACCACGGATCGTCGTGTGCAGCGAGAACAGCCTGCAGTTCTTCATGCC  
CGTGCTGGGCGCCCTGTTTATCGGAGTGGCTGTGGCCCTGCCAACGACATCTACAACGAGCGGAGCTGCTGAACAGCATGGGCATCAGCCAGCCACCGT  
GGTGTTCGTGTCCAAGAAGGGACTGCAGAAAATCCTGAACGTGCAGAAGAAGCTGCCCATCATCCAGAAAATCATCATCATGGACAGCAAGACCGACTACCA  
GGGCTTCAGAGCATGTACACCTTCGTGACCAGCCATCTGCCCTGGCTTCAACGAGTACGACTTCGTGCCGAGAGCTTCGACCCGGGACAAGACAATCGC  
CCTGATCATGAACAGCAGCGGACGACCCGACTGCCTAAAGCGGTGGCCCTGCCTCAGAAAATGCTGCGTGGCGGTTAGCCACGCCCGGAGCCCTATCTT  
CGGAACAGATCATCCCGACACCGCCATCTGAGCGTGGTGCCTTTCCACCACGGCTTCGGCATGTTACCAACCTGGGTACCTGATCTGCGGCTCCG  
GGTGGTGCTGATGTACAGATTCGAGGAAGAACTGTTCTGCGGAGCTGCAGGACTACAAGATCCAGAGCGCCCTGCTGGTGCTACCTGTTTACGCTTCTT  
TGCCAAGAGCACCCTGATCGATAAGTACGACCTGAGCAACCTGCACGAGATCGCTTGGCGGAGCCCCCTGTCTAAAGAAGTGGGAGAGGCCGTGGCCAA  
GCGGTTCCATCTGCCTGGCATCAGACAGGGCTATGGCCTGACCGAGACAACAGCGCCATTCTGATCACCCCGAGGGCGACGATAAGCCTGGCGCCGTGGG  
AAAGTGGTGCCATTCTTCGAGGCCAAGGTGGTGGACCTGGACACCGCAAGACACTGGGCGTGAACAGAGGGGCGAACTGTGTGTGCGGGGACCTATGAT  
CATGAGCGGTACGTGAACAACCCGAGGCCACCAACGCTGATTGACAAGGATGGCTGGCTGCACAGCGGCGACATTGCCTACTGGGACGAGGACGAGCA  
CTTCTTCATCGTGGACCGGTGAAGTCCCTGATCAAGTACAAGGGCTACCAGGTGGCCCCAGCCGAGCTGGAATCTATCTGCTGCAGACCCCAACATCTT  
CGATGCCGCGTGGCAGGACTGCCCAGTATGATGTGCGGAAGTGCACGCGCTGTGGTGGTGCTGGAACACGGAAGACCATGACCGAGAAAGAAATCGT  
GGACTACGTGGCAGCCAAGTGACACCGCCAAGAACTGAGAGCGCGGTGGTGTGTGGACGAGGTGCCAAAGGGCTGACAGCAAGCTGGACGCGCCG  
GAAGATCAGAGAGATCCTGATTAAGGCCAAGAAAGCGGCAAGATCGCCGTGGATCGGAGAAAGAGAGGCTCTGGCGAAGGCAGAGGACGCTGCTTACATG  
TGGCGACGTGGAAGAAACCCCGACCTATGGACTACGTGTCCCTGTGTAACAGATTGTCAGAAAGTACCTGAACAGCCCCCTACACACCTGTGTGATCAT  
CCCCAAGCCTACCGCAAGTACACACCTCTCGTGGGCACATCTCTGCACCCCGTGTGTGGAATTGCCAGCTGAGCTTGGCCGGCTACACCGAGTCTGCCGT  
GAACAGCACAAAGGCCCTGGCCAAACAGGACGCGCTCAGAGAATTGCCTGGCTGTGCACAAGGATGGCGGCATCCCTGATGGCTGTAGCTGTACCTGAG  
ACACAGCAGCTGTTCGCCCAGAGCGAGGAAGAGGAATCCTTCAGAACTGATGAGCGGCGCGAATTGGCAAGCTGCTTACATAGAACTCGCGCGATTGG  
CATGCCGCTTAAAAATTTATTTATTTTCTTTTCTTTTCCGAATCGGATTTGTGTTTAAATATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA [서열번호 41]

[0258] 따라서, 바람직하게는 상기 핵산 서열은 실질적으로 서열번호 41에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 단편 또는 변이체를 포함한다.

[0259] 제3 측면에서, 상기 제2 측면에 따른 핵산 서열을 포함하는 발현 카세트(cassette)가 제공된다.

[0260] 본 발명의 핵산 서열은 바람직하게는 재조합 벡터, 예를 들면 상기 RNA 구조체를 생산할 수 있도록 관심있는 숙주 세포 내로 운반하기 위한 재조합 벡터 내에 포함된다.

[0261] 따라서, 제4 측면에서, 상기 제3 측면에 따른 발현 카세트를 포함하는 재조합 벡터가 제공된다.

[0262] 한 구현예에서, 상기 벡터는 상기과 같은 서열번호 35의 핵산 서열을 포함할 수 있으며, 여기서 "GOI"는 치료 생체분자를 암호화하는 서열의 위치를 나타낸다:

[0263] CGCCAGCAACGCGAGCTCTAATACGACTACTATAGATGGGCGGCGCATGAGAGAAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACA  
TCGAGGAAGACAGCCATTCTCAGAGCTTTGCAGCGGAGCTTCCCGAGTTTGGAGTAGAAGCCAAGCAGGTCAGTATAATGACCATGCTAATGCCAGAG  
CGTTTTCGCATCTGGCTTCAAACTGATCGAAACGAGGTGGACCATCCGACACGATCCTTGACATTGGAAGTGGCCCCGCCGAGAATGTATTCTAAGC  
ACAAGTATCATTTGATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAACTGTAAGGAAATAACTGATA  
AGGAATTGGACAAGAAAATGAAGGAGCTGGCCGCCGTCATGAGCGACCTGACCTGGAAGTGAAGTATGTGCTCCACGACGACGAGTCTGTGCTGCTACG  
AAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGTTGACGACCGACAAGTCTCTATACCAAGCCAATAAGGGAGTTAGAGTCGCTACTGGATAGGCT  
TTGACACACCCCTTTTATGTTTAAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTGTAAACGGCTCGTAACATAGGCC  
TATGCAGCTCTGACGTTATGGAGCGGTACGTAGAGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTGTGTTGGCTCGA  
CCATCTACCACGAGAAGAGGACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTCTACTTACGTGGCAAGCAAAATTACACATGTCGGTGTGAGACTATAG  
TTAGTTGCGACGGGTACGTCTTAAAGAATAGCTATCAGTCCAGGCCGTATGGAAGCCCTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTGTG  
GCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCAGTATGTGCCAGCTACATTGTGTGACCAATGACTGGCATACTGGCAA  
CAGATGTGAGTGGGACGACGCGCAAACTGCTGGTTGGGCTCAACCAGCTATAGTCGTAACGGTCGCACCCAGAGAAACCAATACCATGAAAAATT  
ACCTTTTGCCCGTAGTGGCCAGGCATTGTAGGTGGGCAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTAG  
TCATGGGGTGTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTCAT  
TCGTGCTGCCAGGATAGGCAGTAACACATTGGAGATCGGGCTGAGAAACAAGATCAGGAAAAATGTTAGAGGAGCACAAAGGAGCCGTCACCTCTCATTACCG  
CCGAGGACGTACAAGAAGCTAAGTGCAGCGGATGAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTG  
AGGAGCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGTGGGGCGGCTCAGTGGAGACCTCGTGGCTTGATAAAGGTTACCAGCTACGATG  
GCGAGGACAAGATCGGCTTTACGCTGTGCTTTCTCCGAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCTCTCGCTGAACAAGTCATAGTGA  
TAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGACATGCAATACCCGTCCAGGACTTTCAAGCTC

TGAGTGAAGTGCCACCATTGTGTACAACGAGTGTGTTGTAACAGGTACCTGCACCATTATGCCACACATGGAGGAGCGCTGAACACTGATGAAGAAT  
ATTACAAAAGTGTCAAGCCAGCGAGCAGCAGCGGAATACCTGTACGACATCGACAGGAAACAGTGCCTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCA  
CAGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCG  
TGCCAGGATCAGGCAAGTCTGGCATCATTAAAGCGCAGTCACCAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAGTGTGCAGAAATATAAGGGACG  
TCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTGAATGGATGCAAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTT  
TTGCTTGTATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAACAGTGCCTTTTAAACATGA  
TGTGCCTGAAAGTGCATTTTAACCACGAGATTGACACAAAGTCTCCACAAAAGCATCTCTCGCCGTGCACTAAATCTGTGACTTCGGTCGTCTCAACCT  
TGTTTTACGACAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAACCTAAGCAGGACGATCTCATTCTCA  
CTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATG  
CCGTTCCGTTACAAGTGAAATGAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGAAAAACAC  
TAGCCGGCGACCCATGGATAAAAAACTGACTGCCAAGTACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGAAGCAGAGCATGATGCCATCATGAGGC  
ACATCTTGGAGAGACCGACCTACCGACGTCTCCAGAATAAGGCAACGTTGTGTGGGCCAAGGCTTGTAGTCCGGTGCTGAAGACCGTGGCATAGACA  
TGACCACTGAACAATGGAACACTGTGGATTATTTGAAACGACAAAGCTCACTCAGCAGAGATAGTATTGAACCACTATGCGTGAGGTTCTTTGGACTCG  
ATCTGGACTCCGGTCTATTTCTGCACCCACTGTTCCGTATCCATTAGGAATAATCACTGGGATAAATCCCGTCGCCTAACATGTACGGGTGAATAAAG  
AAGTGGTCCGTACGTCTCTCGCAGGTACCCACAACCTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATC  
CGGCATAAACCTAGTACCTGTAACAGAGACTGCCTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTCTGTCAGCAAAAT  
TGAAGGGCAGAACTGTCTGGTGGTCCGGGAAAAGTTGTCCGTCCAGGCAAAATGGTTGACTGGTTGTGACACCGGCCTGAGGCTACCTCAGAGCTCGGC  
TGGATTTAGGCATCCAGGTGATGTGCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATG  
CCATTAAGCTTAGCATGTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGCGGAACCTGTGTGACATAGGTTATGGTTACGCTGACAGGGCCAGCGAAA  
GCATCATTGGTGCTATAGCGCGCAGTTCAAGTTTCCCGGTATGCAACCCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTGTATTTCATTGGGTACG  
ATCGCAAGGCCGTACGCACAATTCTTACAAGCTTTTCATCAACCTTGACCAACATTATACAGGTTCCAGACTCCACGAAGCCGATGTGCACCCCTCATATC  
ATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATA  
AGAAATTCGCGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGGACTGGTCAAAGGTGCAGTAAACATATCATTTCATGCCGTAGGACCAAACT  
TCAACAAAGTTTCGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAAACAATTACAAGTCAGTAGCGATTTC  
CACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCA  
TATACTGCAGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCAGTGACAG  
AACCTGATGCAGAGCTGGTGAGGGTGATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGA  
CCAAGTTTACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAA  
GCATGAGCAGTATTAGTTCGAAATGCCCCGTGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCCTGCTTGTGTCATCCATGCCATGACTCCAGAAA  
GAGTACAGCGCTAAAAGCCTCAGTCCAGAAACAATTACTGTGTGCTATCCTTTCCATTGCGGAAGTATAGAATCACTGGTGTGAGAAGATCCAATGCT  
CCCAGCTATATTGTTCTCACCGAAAGTGCTGCGTATATTTCATCAAGGAAGTATCTCGTGAAACACCACCGGTAGACGAGACTCCGAGCCATCGGCAG  
AGAACCAATCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATATCGAAGAGGAAGAAG  
AGGATAGCATAAGTTGTGTGATGAGGCGGACCCACAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAGTCTATCTGTTCCA  
TTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCTGGAGGAGCTAGCGTGACCAGCGGGCAACGTCAGCCGAGACTAACTCTT  
ACTTCGCAAGAGATATGGAGTTTCTGGCGGACCGGTGCCTGCGCTCGAACAGTATTACGAACCTCCACATCCCGCTCCGCGCACAAGAACCCTGAC  
TTGCAACCCAGCAGGGCTGCTCGAGAACCAGCTAGTTTCCACCCCGCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTGAGGCGCTTACCCCGTCAC  
GCACTCCTAGCAGTCCGTCTCGAGAACCAGCTGGTCTCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCTGTAGCACAAC  
AACAATGACGGTTTGTATGCGGTGCATATCTTTTCTCCGACACCGGTCAAGGGCATTACAACAAAATCAGTAAGGCAACCGGTGCTATCCGAAGTGG  
TGTGGAGAGGACCGAATTGGAGATTTCGTATGCCCGCGCCTCGACCAAGAAAAAGAAATTACTACGCAAGAAATTACAGTTAAATCCACACCTGCTA  
ACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAG  
TGGAGTGCTACCGAACCTGCATCTGTTCTTTGATTATCTAGTGTGAACCGTGCTTTTCAAGCCCAAGGTGCGAGTGGAAGCCTGTAACGCCATGT  
TGAAAGAGAATTTCCGACTGTGGCTTCTTACTGTATTATCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTACGTGCTTAGACACTGCCA  
GTTTTTGCCCTGCAAGGTGCGCAGCTTTCCAAAGAAACCTCCTATTTGGAACCCACAATACGATCGGCAGTGCTTCAGCGATCCAGAACACGCTCCAGA  
ACGTCCTGGCAGTGCCACAAAAGAAATTTGCAATGTACGCAAAATGAGAGAATTGCCCGTATTGGATTGCGGCGCTTTAATGTGGAATGCTTCAAGAAAT  
ATGCGTGAATAATGAATATTGGGAAACGTTTAAAGAAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGACCAAAAG  
CTGCTGCTCTTTTTCGAAGACACATAATTTGAATATGTTGACGACATACCAATGGACAGGTTTGAATGGACTTAAAGAGAGACGTGAAAGTACTCCAG  
GAACAAAACATACTGAAGAACGGCCCAAGGTACAGGTGATCCAGGTGCGGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGA  
GATTAATGCGGTCTGCTTCCGAACATTATACACTGTTGATATGTCGGTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTG  
TTCTGAAACTGACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTGGAAGACTTAGGTGTGGACGACAGCTGT  
TGACGCTGATTGAGGCGGCTTTCGGGCAAAATTCATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCA  
CACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTGAGAGAACGGCTAACCGATCACCATGTGCAGCATTCATTGGAGATGACAATA

TCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGC  
CTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGGCGTGGCAGACCCCTAAAAAGGCTGTTAAGCTTGGCAAACCTCTGG  
CAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGCGCAGTAG  
AATCAAGGTATGAAACCGTAGGAACCTCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACGTACCTGAGAGGGGCCCTATAA  
CTCTCTACGGCTAACCTGAATGGACTACGACATAGTCTAGTCCGCCAAGTCTAGCATATGGCCACC - - - - - GOI  
- - - - -  
TGATGAGCGGCCGGAATTGGCAAGCTGTCTACATAGAACTCGCGGCGATTGGCATGCCGCCTAAAAATTTTATTTTATTTTCTTTTCTTTTCCGAATCG  
GATTTTGTTTTAAATATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGCGTCGAGGGAATTAATCTTGAAGACGAAAGGGCCAGGTGGCAC  
TTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCGTATAAATGCTTCAATA  
ATATTGAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTATTCCCTTTTTTGCGGCATTTTGCCTTCTGTTTTGCTCAGCCAGAAACGCT  
GGTGAAGTAAAAGATGCTGAAGATCAGTTGGGTGCAGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCCGAAGA  
ACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGTATGTGGCGCGTATTATCCGTGTTGACGCCGGCAAGAGCAACTCGGTGCGCGCATACACTATTC  
TCAGAACTGACTTGGTTGAGTACTACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAA  
CACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCCTTTTTTGACAAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGA  
ACCGGAGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTACTTAC  
TCTAGCTTCCCGCAACAATTAAGACTGGATGGAGGCGGATAAAGTTGACAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA  
ATCTGGAGCCGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGAGTCAGGC  
AACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGTGACACCAAGTTACTCATATATACTTTAGATTGA  
TTTAAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC  
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCTAATCTGCTGCTTGCAAAACAAAAAACACCGCTACCAGCGGTGGTTG  
TTTGCCGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAACTACTGCTCTTAGTGAGCCGTAGTTAGGCCA  
CCACTTCAAGAACTCTGTAGACCGCTACATACCTCGCTCTGTAATCCTGTTACCAAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCAGGTTGGA  
CTCAAGACGATAGTTACCGGATAAGGCGCAGCGTGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA  
CCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGAGGGTCGGAACAGAGAGCGCACGAGGGA  
GCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTGTGCGGTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCT  
ATGGAATA [서열번호 35]

[0264] 따라서, 바람직하게는 상기 벡터는 실질적으로 서열번호 35에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0265] 벡터가 MERS-CoV ORF4a를 포함하는 RNA 구조체를 암호화하는 핵산 서열을 포함하는 구현예에서, 상기 벡터는 다음과 같은 서열번호 36의 핵산 서열을 포함할 수 있으며, 여기서 "GOI"는 치료 생체분자를 암호화하는 서열의 위치를 나타낸다:

[0266] CGCCAGCAACGCGAGCTCTAATACGACTACTATAGATGGGCGGCGCATGAGAGAAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACA  
TCGAGGAAGACAGCCATTCTCAGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAG  
CGTTTTCGCATCTGGCTTCAAACTGATCGAAACGAGGTGGACCCATCCGACAGATCCTTGACATTGGAAGTGCGCCCGCCGAGAATGTATTCTAAGC  
ACAAGTATCATTTGATCTGTCCGATGAGATGTGCGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAACTGTAAAGAAAACTGATA  
AGGAATTGGACAAGAAAAATGAAGGAGCTGGCCGCCGTCATGAGCGACCTGACCTGGAACTGAGACTATGTGCTCCACGACGACGAGTCGTGCTGCTACG  
AAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGTTGACGGACCGACAAGTCTCTATACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCT  
TTGACACCACCCCTTTTATGTTTAAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTGTAAACGGCTCGTAACATAGGCC  
TATGCAGCTCTGACGTTATGGAGCGGTACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTGTGTTGGCTCGA  
CCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTGCGGTGTGAGACTATAG  
TTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCGTGTATGGGAAGCCCTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGT  
GCTGCAAAAGTGACAGACACATTGAACGGGGAGAGGGTCTTTTTCCCGTGTGACGATATGTGCCAGCTACATTGTGTGACCAATGACTGGCATACTGGCAA  
CAGATGTCAGTGCAGACGACGCGCAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACACCAATACCATGAAAAATT  
ACCTTTTGCCCGTAGTGGCCAGGCATTGTGATAGTGGGCAAGGAATATAAGGAAGATCAAGAAGATGAAGGCCACTAGGACTACGAGATAGACAGTTAG  
TCATGGGGTGTGTTGGGCTTTTGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTCAT  
TCGTGCTGCCAGGATAGGCAGTAACACATTGGAGATCGGGCTGAGAACAAGATCAGGAAATGTTAGAGGAGCACAAAGGAGCCGTCACCTCTCATTACCG  
CCGAGGACGTACAAGAAGCTAAGTGCAGCCGATGAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTTACCACCTTTGGCAGCTGATGTTG  
AGGAGCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATG  
GCGAGGACAAGATCGGCTTTACGCTGTGCTTCTCCGAGGCTGACTCAAGAGTGAAAAATTATCTTGCATCCACCTCTCGCTGAACAAGTCATAGTGA  
TAACACACTCTGGCCGAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCGTCCAGGACTTTCAAGCTC

TGAGTGAAAGTGCCACCATTGTGTACAACGACGTGAGTTTCGTAACAGGTACCTGCACCATTATGCCACACATGGAGGAGCGCTGAACACTGATGAAGAAT  
ATTACAAAAGTGTCAAGCCAGCGAGCAGCAGCGGAATACCTGTACGACATCGACAGGAAACAGTGCCTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCA  
CAGGCGAGCTGGTGGATCCTCCCTTCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCG  
TGCCAGGATCAGGCAAGTCTGGCATCATTAAGCGCAGTCACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAGTGTGCAGAAATATAAGGGACG  
TCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTT  
TTGCTTGTATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAACAGTGCCTTTTAAACATGA  
TGTGCCTGAAAGTGCATTTTAACACGAGATTGACACAAAGTCTCCACAAAAGCATCTCTCGCCGTGCACTAAATCTGTGACTTCGGTCGTCTCAACCT  
TGTTTTACGACAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAACCTAAGCAGGACGATCTCATTCTCA  
CTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATG  
CCGTTCCGTTACAAGTGAAATGAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGAAAAACAC  
TAGCCGGCGACCCATGGATAAAAAACACTGACTGCCAAGTACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGAAGCAGAGCATGATGCCATCATGAGGC  
ACATCTTGGAGAGACCGACCTACCGACGTCTCCAGAATAAGGCAACGTGTGTTGGGCCAAGGCTTGTAGTCCGGTGCTGAAGACCGTGGCATAGACA  
TGACCACTGAACAATGGAACACTGTGGATTATTTGAAACGACAAAGCTCACTCAGCAGAGATAGTATTGAACCACTATGCGTGAGGTTCTTTGGACTCG  
ATCTGGACTCCGGTCTATTTCTGCACCCACTGTTCCGTATCCATTAGGAATAATCACTGGGATAAATCCCGTCGCCTAACATGTACGGGTGAATAAAG  
AAGTGGTCCGTACGTCTCTCGCAGGTACCCACAACCTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATC  
CGGCATAAACCTAGTACCTGTAACAGAGACTGCCTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTCTGCAGCAAAAT  
TGAAGGGCAGAACTGTCTGGTGGTCCGGGAAAAGTTGTCCGTCCAGGCAGAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTCAGAGCTCGGC  
TGGATTTAGGCATCCAGGTGATGTGCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTAAGACCATG  
CCATTAAGCTTAGCATGTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGCGGAACCTGTGTGAGCATAGGTTATGGTTACGCTGACAGGGCCAGCGAAA  
GCATCATTGGTGCTATAGCGCGCAGTTCAAGTTTCCCGGTATGCAACCCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTGTATTTCATTGGGTACG  
ATCGCAAGGCGGTACGCACAATTCTTACAAGCTTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCCTCATATC  
ATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATA  
AGAAATTCGCGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGGACTGGTCAAAGGTGCAGCTAAACATATCATTTCATGCCGTAGGACCAAACT  
TCAACAAAGTTTCGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAAACAATTACAAGTCAGTAGCGATTTC  
CACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCA  
TATACTGCAGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCAGTGACAG  
AACCTGATGCAGAGCTGGTGAGGGTGATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGA  
CCAAGTTTACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAA  
GCATGAGCAGTATTAGTTCGAAATGCCCCGTGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCCTGCTTGTGTCATCCATGCCATGACTCCAGAAA  
GAGTACAGCGCCTAAAAGCCTCAGTCCAGAAACAATTACTGTGTGCTCATCTTTCCATTGCGGAAGTATAGAATCACTGGTGTGAGAGATCAATGCT  
CCCAGCTATATTGTTCTCACCGAAAGTGCTGCGTATATTTCATCAAGGAAGTATCTCGTGAAACACCACCGGTAGACGAGACTCCGAGCCATCGGCAG  
AGAACCAATCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATATCGAAGAGGAAGAAG  
AGGATAGCATAAGTTGTGTGATGAGGCGGACCCACAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAGTCTCATCTGTTCCA  
TTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCTGGAGGAGCTAGCGTGACCAGCGGGCAACGTCAGCCGAGACTAACTCTT  
ACTTCGCAAGAGATATGGAGTTTCTGGCGCGACCGGTGCCTGCGCTCGAACAGTATTACGAACCCCTCCACATCCCGCTCCGCGCACAAGAACCCTGAC  
TTGCAACCCAGCAGGGCTGCTCGAGAACAGCCTAGTTTCCACCCCGCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTGAGGCGCTTACCCCGTCAC  
GCACTCCTAGCAGGTGCTGCTCGAGAACAGCCTGGTCTCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCTGTAGCACAAC  
AACAATGACGGTTTGTATGCGGTGTCATATCTTTTCTCCGACACCGGTCAAGGGCATTACAACAAAATCAGTAAGGCAACCGGTGCTATCCGAAGTGG  
TGTGGAGAGGACCGAATTGGAGATTTCGTATGCCCGCGCCTCGACCAAGAAAAAGAAATTACTACGCAAGAAATTACAGTTAAATCCACACCTGCTA  
ACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAAGAAAAG  
TGGAGTGCTACCGAACCTGCATCTGTTCTTTGATTATCTAGTGTGAACCGTGCTTTTCAAGCCCAAGGTGCGAGTGGAAGCCTGTAACGCCATGT  
TGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTATCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTACGTGCTTAGACACTGCCA  
GTTTTTGCCCTGCAAGGTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCTTCAGCGATCCAGAACACGCTCCAGA  
ACGTCCTGGCAGTGCCACAAAAGAAATTTGCAATGTACGCAAAATGAGAGAATTGCCCGTATTGGATTGCGGCGCTTTAATGTGGAATGCTTCAAGAAAT  
ATGCGGTGAATAATGAATATTGGGAAACGTTTAAAGAAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGACCAAAAG  
CTGCTGCTCTTTTTCGAAGACACATAATTTGAATATGTTGACGACATACCAATGGACAGGTTTGAATGGACTTAAAGAGAGACGTGAAAGTACTCCAG  
GAACAAAACATACTGAAGAACGCCCCAAGGTACAGGTGATCCAGGTGCGGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGA  
GATTAATGCGGTCTGCTTCCGAACATTCATACACTGTTGATATGTCGGTGGAAGCTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTG  
TTCTGAAACTGACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTGGAAGACTTAGGTGTGGACGACAGCTGT  
TGACGCTGATTGAGGCGGCTTTGCGGCAAAATTCATCAATACATTTGCCCACTAAACTAAATTTAAATTCGAGGCCATGATGAAATCTGGAATGTTCTCA  
CACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATA

TCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGC  
CTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCTGTGGCAGACCCCTAAAAAGGCTGTTAAGCTTGGCAAACCTCTGG  
CAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG  
AATCAAGGTATGAAACCGTAGGAACCTCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTACGTACCTGAGAGGGGCCCTATAA  
CTCTCTACGGCTAACCTGAATGGACTACGACATAGTCTAGTCCGCCAAGTCTAGCATATGGCCACC - - - - - GOI  
CGGAGAAAGAGAGGCTCTGGCGAAGGCAGAGGCAGCCTGCTTACATGTGGCGACGTGGAAGAGAACCCCGGACCTATGGACTACGTGTCCCTGTGAACCA  
ATTTGGCAGAAGTACCTGAACAGCCCTACACCACCTGTCTGTACATCCCCAAGCCTACCGCCAAGTACACACCTCTCGTGGGCACATCTCTGCACCCCGTG  
CTGTGGAATTGCCAGCTGAGCTTTGCCGGCTACACCGAGTCTGCCGTGAACAGCACAAAGGCCCTGGCCAAACAGGACGCCGCTCAGAGAATTGCCTGGCTG  
CTGCACAAGGATGGCGGCATCCCTGATGGCTGTAGCCTGTACCTGAGACACAGCAGCCTGTTGCGCCAGAGCGAGGAAGAGGAATCCTTCAGCAACTGATGA  
GCGGCCCGCAATTGGCAAGCTGCTTACATAGAACTCGCGCGATTTGGCATGCGCCCTAAAAATTTTATTTTATTTTCTTTTCTTTTCCGAATCGGATTTT  
GTTTTTAATATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGCGTCGAGGGGAATTAATCTTGAAGACGAAAGGGCCAGGTGGCACTTTTCG  
GGGAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTG  
AAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCTTATTCCTTTTTCGCGCATTTTGCCTTCCTGTTTTGCTCACCAGAAACGCTGGTGAA  
AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAATGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTCGCCCCGAAGAAGCTTT  
TCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAA  
TGACTTGGTTGAGTACTACCACTACAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCACTGCTGCCATAACCATGAGTGATAACACTGC  
GGCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTCGACAACATGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGA  
GCTGAATGAAGCCATACCAACGACGAGCGTGACACCAGTGCCTGTAGCAATGGCAACACGTTGCGCAAACTATTAACCTGGCAACTACTTACTCTAGC  
TTCCCGCAACAATTAATAGACTGGATGGAGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTCCGGCTGGCTGGTTATTGCTGATAAATCTGG  
AGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGCAGGGGAGTCAGGCAACTAT  
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTTGGTAACGTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAA  
ACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTAACGTGAGTTTTCGTTCCTAGAGCGTCAGACCC  
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTCGCGCGTAATCTGTGCTTGCACAAAAAACACCGCTACCAGCGTGGTTTGTGTTGCC  
GGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTACGACAGCGCAGATACCAAACTACTGCTCTTAGTGAGCGGTAGTTAGGCCACCACTT  
CAAGAACTCTGTAGACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAG  
ACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGTTCTGTGCACACAGCCAGCTTGGAGCGAACGACCTACACGAACTGAGATACCTACA  
GCGTGAGCATTGAGAAGCGCCACGCTTCCCGAAGGAGAAAGCGGACAGGTATCCCGTAAGCGGCAGGGTTCGAACAGGAGAGCGACGAGGGAGCTTCC  
AGGGGAAACGCCTGGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAA  
AAA [서열번호 36]

[0267] 따라서, 바람직하게는 상기 벡터는 실질적으로 서열번호 36에 나타난 것과 같은 뉴클레오티드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0268] 벡터가 PIV5를 포함하는 RNA 구조체를 암호화하는 핵산 서열을 포함하는 구현예에서, 상기 벡터는 다음과 같은 서열번호 37의 핵산 서열을 포함할 수 있으며, 여기서 "GOI"는 치료 생체분자를 암호화하는 서열의 위치를 나타낸다:

[0269] CGCCAGCAACGCGAGCTCTAATACGACTCACTATAGATGGGCGGCGCATGAGAGAAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGACA  
TCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAG  
CGTTTTTCGCATCTGGCTTCAAACTGATCGAAACGAGGTGGACCCATCCGACAGATCCTTGACATTGGAAGTGCGCCCGCCGAGAATGTATTCTAAGC  
ACAAGTATCATTTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAACTGTAAAGAAATACTGATA  
AGGAATTGGACAAGAAAAATGAAGGAGCTGGCCGCCGTCATGAGCGACCTGACCTGGAACTGAGACTATGTGCCTCCACGACGACGAGTCGTGCTGCTACG  
AAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAGCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCT  
TTGACACCACCCCTTTTATGTTTAAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCC  
TATGCAGCTCTGACGTTATGGAGCGGTACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTCTGTTGGCTCGA  
CCATCTACCACGAGAAGAGGACTTACTGAGGAGCTGGCACCTGCCGCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGGTGTGAGACTATAG  
TTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGT  
GCTGCAAAGTGACAGACATTTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCAGTATGTGCCAGCTACATTGTGTGACCAATGACTGGCATACTGGCAA  
CAGATGTCAGTGGGACGACGCGCAAAACTGCTGGTTGGGCTCAACACGCTATAGTCGTAACGGTCGCACCCAGAGAAACACCAATACCATGAAAAATT  
ACCTTTTGCCCGTAGTGGCCAGGCATTTGCTAGGTGGGCAAGGAATATAAGGAAGATCAAGAAGATGAAGGCCACTAGGACTACGAGATAGACAGTTAG  
TCATGGGGTGTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTCAT  
TCGTGCTGCCAGGATAGGCAGTAACACATTGGAGATCGGGCTGAGAACAAGATCAGGAAAATGTTAGAGGAGCACAGGAGCGCTACCTCTCATTACCG

CCGAGGACGTACAAGAAGCTAAGTGCAGCCGATGAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTG  
AGGAGCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGAGACACCTCGTGCTTGATAAAGGTTACCAGCTACGATG  
GCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTCCGAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAGTGA  
TAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCAGGACTTTCAAGCTC  
TGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCTGTAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAAT  
ATTACAAAAGTGTCAAGCCAGCGAGCAGCAGCGGAATACCTGTACGACATCGACAGGAAACAGTGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCA  
CAGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCG  
TGCCAGGATCAGGCAAGTCTGGCATCATTAAGGCGCAGTCACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAAGTGTGCAGAAATTATAAGGGACG  
TCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCGTCAGAGACCTGTATATTGACGAAGCTT  
TTGCTTGTATGCAGGACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAACAGTGCGGTTTTTTTAAACATGA  
TGTGCCTGAAAGTGCAATTTAAACCACGAGATTGTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTCTCAACCT  
TGTTTTACGACAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAACCTAAGCAGGACGATCTCATTCTCA  
CTTGTTCAGAGGGTGGGTGAAGCAGTTGCAAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGGTGACCCGTAAGGTGTGTATG  
CCGTTCCGTACAAGGTGAATGAAAATCCTCTGTACGACCCACCTCAGAACATGTGAACGCTCTACTGACCCGCACGAGGACCGCATCGTGTGAAAAACAC  
TAGCCGGCGACCCATGGATAAAACACTGACTGCCAAGTACCCTGGGAATTTACTGCCACGATAGAGGAGTGGAAGCAGAGCATGATGCCATCATGAGGC  
ACATCTTGGAGAGACCGGACCTACCGACGTCTTCCAGAATAAGGCAACGTTGTGTGGGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACA  
TGACCACTGAACAATGGAACACTGTGGATTATTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCG  
ATCTGGACTCCGCTCTATTTCTGCACCCACTGTTCGTTATCCATTAGGAATAATCACTGGGATAACTCCCGTCGCTAACATGTACGGGTGAATAAAG  
AAGTGGTCCGTACGCTCTCTCGCAGGTACCCACAACCTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATC  
CGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTACTTTCTTCTATTCGTCAGCAAT  
TGAAGGGCAGAACTGTCTGTGGTGGGGAAAAAGTTGTCCGTCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCTGAGGCTACCTTCAGAGCTCGGC  
TGGATTTAGGCATCCAGGTGATGTGCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATG  
CCATTAAGCTTAGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGCGGAACCTGTGTCAGCATAGGTTATGGTTACGCTGACAGGGCCAGCGAAA  
GCATCATTGGTGCTATAGCGCGGAGTTCAAGTTTTCCCGGTATGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTGTATTCTATGGGTACG  
ATCGCAAGGCCCGTACGCACAATTCTTACAAGCTTTCATCAACCTTGACCAACATTATACAGGTTCCAGACTCCACGAAGCCGAGTGTGACCCCTCATATC  
ATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATA  
AGAAATTCGGAAGGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGGACTGGTCAAAGGTGCAGCTAAACATATCATTATGCCGTAGGACCAAACT  
TCAACAAAGTTTCGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATT  
CACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCAATCATTGAACATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCA  
TATACTGCAGGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCACTGACAG  
AACCTGATGCAGAGCTGGTGAGGTTGCATCCGAAGAGTTCTTTGGCTGGAAGGAAGGCTACAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGA  
CCAAGTTTACCAGGCGGCAAGGATATAGCAGAAATTAATGCCATGTGGCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAA  
GCATGAGCAGTATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCTTGCTTGTGATCCATGCCATGACTCCAGAAA  
GAGTACAGCGCTAAAAGCCTCAGTCCAGAACAATTAAGTGTGTCTCATCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGAGAAGATCCAATGCT  
CCCAGCTATATTGTTCTACCGAAAGTGCTGCGTATATTATCCAAGGAAGTATCTGTTGAAACACCAACCGGTAGACGAGACTCCGAGGCATCGGCAG  
AGAACCAATCCACAGAGGGGACACCTGAACAACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCTGAGCCGATCATCATGAAGAGGAAGAAG  
AGGATAGCATAAGTTTGTGTGATGAGGCGGACCCACAGGTGCTGCAAGTCGAGGCAGACATTACGGGCCGCCCTCTGTATCTAGCTCATCTGTTCCA  
TTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCTGGAGGGAGCTAGCGTGACCAGCGGGCAACGTACGCGGAGACTAACTCTT  
ACTTCGAAAAGAGTATGGAGTTTCTGGCGGACCGGTGCGCTCGAACAGTATTACAGGAACCTCCACATCCCGCTCCGCGCACAAGAACACCGTCAC  
TTGACCCAGCAGGGCTGCTCGAGAACCAGCTAGTTTCCACCCCGCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCGTCAC  
GCACTCTAGCAGGTCCGTCTCGAGAACCAGCTGGTCTCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCTGTAGCACAAC  
AACAATGACGGTTTGATGCGGTGCATACATCTTTTCTCCGACACCGGTCAAGGCAATTTACAACAAAAATCAGTAAGGCAACCGTGCTATCCGAAGTGG  
TGTTGGAGAGGACCGAATTGGAGATTTGATATGCCCCGCGCTCGACCAAGAAAAAGAAATTAAGTACGCAAGAAATTACAGTTAAATCCACACCTGCTA  
ACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCTAGGGCATTATTGAAGGCAGAGGAAAAAG  
TGGAGTGCTACCGAACCCCTGCATCCTGTTCCCTTTGATTCTAGTGTGAACCGTGCCCTTTTCAAGCCCAAGGTGCGAGTGAAGCCTGTAACGCCATGT  
TGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTATCCAGAGTACGATGCTATTTGGACATGGTTGACGGAGCTTCATGTGCTTAGACACTGCCA  
GTTTTTGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGA  
ACGTCCTGGCAGCTGCCACAAAAAGAAATGCAATGTACGCAAAATGAGAGAATTGCCCGTATTGGATTGCGCGGCTTTAATGTGGAATGCTTCAAGAAAT  
ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGACCAAAAG  
CTGCTGCTCTTTTTCGAAGACACATAATTTGAATATGTTGAGGACATACCAATGGACAGGTTTGAATGGACTTAAAGAGAGACGTGAAAGTGACTCCAG  
GAACAAAACATACTGAAGAACGCCCCAAGGTACAGGTGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGA

GATTAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTG  
TTCTGGAAAGTACATCGCGTCGTTTGATAAAAGTGAGGACGACCCATGGCTCTGACCGCGTTAATGATTCTGGAAGACTTAGGTGTGGACGAGAGCTGT  
TGACGCTGATTGAGGCGGCTTTCGGCGAAATTTTCATCAATACATTTGCCCACTAAACTAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCA  
CACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTGAGAGAACGGCTAACCGGATCACCATGTGACGATTCATTGGAGATGACAATA  
TCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGC  
CTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCTAAAAAGGCTGTTTAAGCTTGGCAAACCTCTGG  
CAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAG  
AATCAAGGTATGAAACCGTAGGAATTCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTAAATCATTACGCTACCTGAGAGGGGCCCTATAA  
CTCTCTACGGCTAACCTGAATGGACTACGACATAGTCTAGTCCGCCAAGTCTAGCATATGGCCACC - - - - - GOI  
- - - - -  
CGGAGAAAGAGAGGCTCTGGCGAAGGCAGAGGCAGCCTGCTTACATGTGGCGACGTGGAAGAGAACCCCGGACCTATGGACCTACCGACCTGAGCTTACG  
CCCGACGAGATCAACAAGCTGATCGAGACAGGCCTGAACACCGTGAATACTTCACCAGCCAGCAAGTGACCGGCACAAGCAGCCTGGGCAAGAACAAT  
CCTCCAGGCGTGACCGGCTGTGACAAATGCTGCCGAGGCCAAGATCCAAGAGAGCACCAACCACCAGAAGGGCTCTGTTGGAGGCGGAGCCAAGCCTAAG  
AAGCCCAGACCTAAGATCGCCATCGTGCCCGGACGATAAGACAGTGCCTGGCAAGCCCATTCCTAATCCTCTGCTGGGCCCTGACAGCACCCCTAGCACA  
CAGACAGTGTGGATCTGAGCGGCAAGACACTGCCTAGCGGCAGCTATAAGGGCGTGAAGCTGGCCAAGTTTCGGCAAAGAAAACCTGATGACCCGGTTCATC  
GAGGAACCCAGAGAGAACCCTATCGCCACCAGCTCTCCCATCGACTTCAAGAGAGGCAGAGACACCGGCGGCTTCCACAGAAGAGAGTACAGCATTGGCTGG  
GTCGGAGATGAAGTGAAAGTGACCGAGTGGTGCAACCCAGCTGCAGCCCTATTACAGCCGCGCTAGAAGATTGAGTGCACCTGTCACCAAGTGTCTGTG  
ACCTGTAGCGAGTGCAGCGGGACACATGATGAGCGGCCGGAATTGGCAAGCTGCTTACATAAGAACTCGCGGCGATTGGCATGCCGCTTAAAAATTTTAT  
TTTATTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAATATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGCGTCGAGGGGAATTAATT  
CTTGAAGACGAAAGGGCCAGGTGGCACTTTTCGGGAAATGTGCGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAG  
ACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTGGCGCATTTTGCT  
TCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAA  
GATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCGGGCAAGA  
GCAACTCGGTGCGCCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG  
CAGTGTGCCATAACCATGAGTGATAACACTGCGGCCAATTACTTCTGACAACGATCGAGGACCGAAGGAGCTAACCGCTTTTTTGACAACATGGGGGA  
TCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCG  
CAAATATTAAGTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTGCAGGACCATTCTGCGCTCGGCCCT  
TCCGGCTGGCTGGTTTATGTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGT  
AGTTATCTACACGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGTACAGCA  
AGTTTACTCATATATACTTTAGATTGATTTAACTTTCATTTTTAATTTAAAGGATCTAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAAATCCCTTA  
ACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAACAAA  
AAAACACCGCTACCAGCGGTGGTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAACTAGT  
CCTTCTAGTGTAGCGTAGTTAGGCCACCACTTCAAGAATCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGTGCCAGTGG  
CGATAAGTGTGCTTACCGGGTTGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGTTGCTGCACACAGCCAGCTTGA  
GCGAACGACCTACACGAACTGAGATACCTACACGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG  
GGTCGGAACAGGAGAGCGCACGAGGAGCTTCCAGGGGAAACGCTGTTATCTTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTT  
GTGATGCTCGTCAGGGGGCGGAGCCTATGGAATAA [서열번호 37]

[0270] 따라서, 바람직하게는 상기 벡터는 실질적으로 서열번호 37에 나타난 것과 같은 뉴클레오타이드 서열, 또는 그의 변이체 또는 단편을 포함한다.

[0271] 본 발명의 saRNA 구조체는 주형으로서 도 7 또는 도 8에 나타나 있는 DNA 플라스미드를 이용해 만들어질 수 있다. 이후, RNA 사본(copy)은 T7 폴리머라아제와 같은 폴리머라아제를 이용한 시험관내 전사에 의해 만들어질 수 있고, 상기 T7 프로모터는 도 7 또는 도 8에서 플라스미드 지도(map)에서 saRNA의 상류에 나타나 있다. 따라서, 제1 측면의 saRNA 구조체는 주형으로서 도 7 또는 도 8에 나타나 있는 서열번호 35 내지 서열번호 37 중 어느 하나에 나타난 것과 같은 핵산 서열을 갖는 DNA 플라스미드, 또는 그의 변이체 또는 단편을 이용해 만들어질 수 있다. 물론, 다른 RNA 폴리머라아제, 예를 들면 SP6 또는 T3 폴리머라아제가 T7 폴리머라아제 대신에 사용될 수 있음이 인식될 것이며, 이 경우 상기 saRNA 구조체는 대신에 SP6 또는 T3 프로모터를 포함할 수 있다.

[0272] 제1 측면의 RNA 구조체를 암호화하는 제4 측면의 벡터는 예를 들면 플라스미드, 코스미드(cosmid) 또는 파지(phage)이거나, 및/또는 바이러스 벡터일 수 있다. 이러한 재조합 벡터는 세포를 뉴클레오타이드 서열로 형질전환하기 위한 본 발명의 운반 시스템에서 매우 유용하다. 상기 뉴클레오타이드 서열은 바람직하게는 DNA 서열일 수

있고, 상기 DNA 서열이 제1 측면의 RNA 구조체를 형성하는 RNA 서열을 암호화한다.

- [0273] 제1 측면의 RNA 구조체를 암호화하는 재조합 벡터는 또한 다른 기능적 요소를 포함할 수 있다. 예를 들면, 재조합 벡터는 숙주 세포에서 벡터의 도입시 전이유전자 발현을 시작하기 위한 적합한 프로모터를 포함하는 다양한 다른 기능적 요소를 추가로 포함할 수 있다. 예를 들어, 상기 벡터는 바람직하게는 숙주 세포, 예컨대 박테리아 세포의 핵에서 자발적으로 복제할 수 있다. 이 경우, DNA 복제를 유도 또는 조절하는 요소가 상기 재조합 벡터에 필요할 수 있다. 대안적으로, 상기 재조합 벡터는 숙주 세포의 게놈 내로 통합되도록 디자인될 수 있다. 이 경우, (예컨대 상동성 재조합에 의한) 표적화된 통합에 유리한 DNA 서열이 예측된다. 적합한 프로모터는 예로서 SV40 프로모터, CMV, EF1a, PGK, 바이러스 긴 말단 반복뿐만 아니라 유도가능한 프로모터, 예컨대 테트라사이클린 유도가능한 시스템을 포함할 수 있다. 상기 카세트 또는 벡터는 또한 종결자(terminator), 예컨대 베타 글로빈, SV40 폴리아데닐화 서열 또는 합성 폴리아데닐화 서열을 포함할 수 있다. 상기 재조합 벡터는 또한 필요시 핵산의 발현을 제어하기 위한 프로모터 또는 조절자(regulator) 또는 인핸서(enhancer)를 포함할 수 있다.
- [0274] 상기 벡터는 또한 형질감염 또는 형질전환된 세포의 선택을 가능하게 하고, 이중성 DNA가 통합된 벡터를 갖고 있는 세포의 선택을 가능하게 하는 클로닝 공정에서 선택가능한 마커로서 사용될 수 있는 유전자를 코딩하는 DNA를 포함할 수 있다. 예를 들면, 암피실린, 네오마이신, 푸로마이신 또는 클로람페니콜 저항성이 예측된다. 도 7 및 도 8에 나타나 있는 벡터는 박테리아에서 플라스미드를 선별하기에 유용한 암피실린 저항성 마커를 포함한다. 대안적으로, 상기 선택가능한 마커 유전자는 전이유전자(들)를 함유하는 벡터와 동시에 사용될 수 있는 상이한 벡터에 있을 수 있다. 상기 카세트 또는 벡터는 또한 뉴클레오티드 서열의 발현을 조절하거나, 발현된 폴리펩티드를 숙주 세포의 소정 부분으로 표적화하는데 관여되는 DNA를 포함할 수 있다.
- [0275] 정제된 벡터는 적합한 수단, 예컨대 직접 식작용 섭취(endocytotic uptake)에 의해 숙주 세포 내로 직접적으로 삽입될 수 있다. 상기 벡터는 형질감염, 감염, 전기천공, 미세주사, 세포 융합, 원형질체 융합 또는 탄도적 충격(ballistic bombardment)에 의해 숙주 세포(예컨대 진핵생물 또는 원핵생물 세포) 내로 직접 도입될 수 있다. 대안적으로, 본 발명의 벡터는 입자 총을 이용해 숙주 세포 내로 직접 도입될 수 있다.
- [0276] 상기 핵산 분자는 (필수적인 것은 아니지만) 숙주 세포의 DNA에 통합되게 되는 것일 수 있다. 미분화된 세포는 인정적으로 형질전환되어 유전적으로 변형된 딸세포의 생산을 유도할 수 있다(이 경우, 예컨대 특정 전사 인자 또는 유전자 활성화제를 이용한 대상체에서의 발현의 조절이 필요할 수 있다). 대안적으로, 상기 운반 시스템은 분화된 세포의 불안정하거나 일시적인 형질전환을 선호하도록 디자인될 수 있다. 이 경우일 때는 발현의 조절이 덜 중요할 수 있는데, 그 이유는 형질전환된 세포가 죽거나 단백질의 발현을 중단할 때 상기 DNA 분자의 발현이 중단될 것이기 때문이다.
- [0277] 대안적으로, 상기 운반 시스템은 벡터에 통합되지 않고 숙주 세포에 핵산 분자를 제공할 수 있다. 예를 들어, 상기 핵산 분자는 리포솜 또는 바이러스 입자 내에 통합될 수 있다. 대안적으로, "네이키드(naked)" 핵산 분자가 적합한 수단, 예컨대 직접 식작용 섭취에 의해 숙주 세포 내로 삽입될 수 있다.
- [0278] 제5 측면에서, 제1 측면의 RNA 구조체, 제2 측면의 핵산 서열, 제3 측면의 발현 카세트 또는 제4 측면의 벡터, 및 약학적으로 허용가능한 비히클(vehicle)을 포함하는 약학적 조성물이 제공된다.
- [0279] 제6 측면에서, 제5 측면에 따른 약학적 조성물을 제조하기 위한 공정이 제공되며, 상기 방법은 제1 측면의 RNA 구조체, 제2 측면의 핵산 서열, 제3 측면의 발현 카세트 또는 제4 측면의 벡터를 약학적으로 허용가능한 비히클과 접촉시키는 단계를 포함한다.
- [0280] 제7 측면에서, 제1 측면의 RNA 구조체의 제조 방법이 제공되며, 상기 방법은
- [0281] a) i) 숙주 세포 내로 제4 측면의 벡터를 도입하는 단계; 및
- [0282] ii) 제1 측면의 RNA 구조체의 생산을 야기하기 위한 조건 하에 상기 숙주 세포를 배양하는 단계를 포함하거나; 또는
- [0283] b) 제4 측면에 따른 벡터로부터 RNA 구조체를 전사하는 단계를 포함한다.
- [0284] 단계 a)의 숙주 세포는 진핵생물 또는 원핵생물 숙주 세포일 수 있다. 바람직하게는, 상기 숙주 세포는 진핵생물 숙주 세포이다. 보다 바람직하게는, 상기 숙주 세포는 포유동물 숙주 세포, 예컨대 인간 배아 신장 293 세포 또는 중국 햄스터 난소(CHO) 세포이다. 단계 (b)는 시험관내 또는 생체내에서, 바람직하게는 시험관내에서 수행될 수 있다.

- [0285] 시험관내 전사의 적합한 방법은 본 기술분야에 잘 알려져 있으며, 본 기술분야의 기술자에게 알려져 있을 것이다. 예를 들면, [Molecular Cloning, A Laboratory Manual, 2nd edition. (1989) editor C Nolan, Cold Spring Harbor Laboratory Press]에 기술되어 있다.
- [0286] 제1 측면의 RNA 레플리콘은 치료범용으로 특히 적합하다.
- [0287] 본 발명자들은 제1 측면의 RNA 구조체가 치료법에서 생체내에서 사용하기 위해 시험관내 전사에 의해 생성될 것으로 예측했지만, 본 기술분야의 숙련자는 상기 RNA 구조체는 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터를 대상체에게 생체내 운반함으로써 치료법을 위해 대상체에서 생체내에서 생성될 수 있음을 인식할 것이다.
- [0288] 따라서, 제8 측면에 따르면, 약제로서 또는 치료법에서 사용하기 위한 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물이 제공된다.
- [0289] 본 발명의 제9 측면에서, 원생동물, 진균, 박테리아 또는 바이러스 감염의 예방, 개선 또는 치료에 사용하기 위한 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물이 제공된다.
- [0290] 상기 원생동물, 진균, 박테리아 또는 바이러스 감염은 제1 측면에서 정의된 것과 같은 원생동물, 진균, 박테리아 또는 바이러스의 감염일 수 있다.
- [0291] 본 발명의 제10 측면에서, 암의 예방, 개선 또는 치료에 사용하기 위한 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물이 제공된다.
- [0292] 상기 암은 제1 측면에서 정의된 것일 수 있다.
- [0293] 본 발명의 제11 측면에서, 원생동물, 진균, 박테리아 또는 바이러스 감염을 치료하기 위한 방법이 제공되며, 상기 방법은 이를 필요로 하는 대상체에게, 치료적 유효량의 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물을 투여하는 단계를 포함한다.
- [0294] 치료되어야 하는 원생동물, 진균, 박테리아 또는 바이러스 감염은 제1 측면에서 정의된 것과 같은 원생동물, 진균, 박테리아 또는 바이러스의 감염일 수 있다.
- [0295] 본 발명의 제12 측면에서, 암을 치료하기 위한 방법이 제공되며, 상기 방법은 이를 필요로 하는 대상체에게 치료적 유효량의 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물을 투여하는 단계를 포함한다.
- [0296] 치료되어야 하는 암은 제1 측면에서 정의된 것일 수 있다.
- [0297] 본 명세서에 기술된 RNA 구조체는 바이러스 감염 및 암에 대해 대상체를 백신화하는 효과적인 수단을 제공한다.
- [0298] 따라서, 본 발명의 제13 측면에서, 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물을 포함하는 백신이 제공된다.
- [0299] 바람직하게는, 상기 백신은 적합한 어주번트(adjuvant)를 포함한다.
- [0300] 상기 어주번트는 제1 측면의 RNA 구조체 서열, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물에 암호화되는 암호화된 분자 어주번트, 또는 운반 제형 내로 통합된 어주번트일 수 있다.
- [0301] 상기 암호화된 분자 어주번트는 사이토카인, 예를 들면 IL-12, GM-CSF, IL-2, IFN- $\gamma$ , 또는 이펙터(effector) 단백질, 예컨대 CD40L, Flt-3 또는 미생물 단백질, 예컨대 플라젤린 또는 콜레라 독소 B를 암호화할 수 있다.
- [0302] 운반 제형 내로 통합된 어주번트는 박테리아 지질펩티드, 지질단백질 및 리포테이코산; 마이코박테리아 지질글리칸; 효모 자이모산, 포린, 지질폴리사카라이드, 지질 A, 모노포스포릴 지질 A(MPL), 플라젤린, CpG DNA, 해모조인, 사포닌(Quil-A, QS-21, 토마틴, ISCOM, ISCOMATRIX<sup>TM</sup>), 스쿠알렌계 에멀전, 폴리머, 예컨대 PEI, 카르복실, 지질 나노입자 및 박테리아 독소(CT, LT)로 이루어진 군으로부터 선택될 수 있다.

- [0303] 본 발명의 제14 측면에서, 대상체에서 면역 반응을 자극하는데 사용하기 위한 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물이 제공된다.
- [0304] 상기 면역 반응은 제1 측면에서 정의된 항원에 다른 원생동물, 박테리아, 바이러스, 진균 또는 암에 대해 자극될 수 있다.
- [0305] 제15 측면에서, 줄기 세포 치료법에서 사용하기 위한 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물이 제공된다.
- [0306] 줄기 세포 치료법은 체세포를 줄기 세포 특징을 갖는 세포로 재프로그래밍하는 것에 관한 것일 수 있다.
- [0307] 체세포는 제1 측면에서 정의된 것과 같은 줄기 세포 특징을 갖는 세포로 체세포를 재프로그래밍 하는 것을 향상시킬 수 있는 하나 이상의 단백질을 운반함으로써 재프로그래밍될 수 있다.
- [0308] 제16 측면에 따르면, 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물을 세포로 운반하는 단계를 포함하는 생체의 또는 시험관 내에서 세포를 변형시키는 방법이 제공된다.
- [0309] 바람직하게는, 상기 방법은 생체외에서 수행된다.
- [0310] 상기 세포는 진핵생물 또는 원핵생물 세포일 수 있다. 바람직하게는, 상기 세포는 진핵생물 세포이다. 보다 바람직하게는, 상기 세포는 포유동물 숙주 세포이다. 가장 바람직하게는, 상기 세포는 인간 세포이다.
- [0311] 바람직하게는, 상기 변형된 세포는 세포-치료법 적응증을 위해 적합하다.
- [0312] 제17 측면에서, 제16 측면의 방법으로부터 수득되거나 이것에 의해 수득가능한 변형된 세포가 제공된다.
- [0313] 제18 측면에서, 치료법, 선택적으로 세포 치료법에서 사용하기 위한 제17 측면의 변형된 세포가 제공된다.
- [0314] 제1 측면에 따른 RNA 구조체, 제2 측면에 따른 핵산, 제3 측면에 따른 발현 카세트, 제4 측면에 따른 벡터 또는 제5 측면에 따른 약학적 조성물(본 명세서에서 활성제로 알려짐)이 질환을 치료, 개선, 또는 예방하거나 백신화하기 위한 단일치료법(즉, 상기 활성제의 사용)으로서 사용될 수 있는 약제에서 사용될 수 있음이 인식될 것이다. 대안적으로, 본 발명에 따른 활성제는 질환의 치료, 개선, 또는 예방을 위해 알려진 치료법에 부가하여, 또는 이와 조합하여 사용될 수 있다.
- [0315] 본 발명의 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 특히 조성물이 사용되어야 하는 방식에 따라 다수의 상이한 형태를 갖는 조성물에 조합될 수 있다. 따라서, 예를 들면, 상기 조성물은 분말, 정제, 캡슐, 액체, 연고, 크림, 겔, 히드로겔, 에어로졸, 스프레이, 미셀 용액, 경피 패치, 리포솜 현탁액, 폴리플렉스, 에멀전, (표면에 RNA를 갖거나 캡슐화된) 지질 나노입자의 형태 또는 치료 또는 백신화를 필요로 하는 인간 또는 동물에게 투여될 수 있는 임의의 다른 적합한 형태일 수 있다. 본 발명에 따른 약제의 비히클은 이것이 제공되는 대상체에 의해 잘-용인되는 것이어야 함이 인식될 것이다.
- [0316] 본 발명의 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 또한 느린-또는 지연된-방출 장치 내에 통합될 수 있다. 이러한 장치는, 예를 들면, 피부 위 또는 아래에 삽입될 수 있고, 상기 약제는 수 주 또는 심지어 수 개월에 걸쳐 방출될 수 있다. 상기 장치는 치료 부위와 적어도 인접하게 위치될 수 있다. 이러한 장치는 상기 유전자 구조체 또는 재조합 벡터를 이용한 장기간의 치료가 필요하고 정상적으로 빈번한 투여(예컨대 적어도 매일 주사)가 필요할 때 특히 유리할 수 있다.
- [0317] 그러나, 바람직한 구현예에서, 본 발명에 따른 약제는 혈류, 근육, 피부 내로, 또는 치료를 필요로 하는 부위 내로 직접 주사함으로써 대상체에게 투여될 수 있다. 가장 바람직하게는, 상기 RNA 구조체를 포함하는 약제는 근육 내로 주사된다. 주사는 정맥내(볼루스 또는 수액) 또는 피하(볼루스 또는 수액), 또는 진피내(볼루스 또는 수액), 또는 근육내(볼루스 또는 수액)일 수 있다.
- [0318] 필요한 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물의 양은 그 생물학적 활성 및 생체이용성에 의해 결정되고, 차례로 상기 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물의 투여 방식, 물리화학적 특성 및 단일치료법 또는 조합 치료법으로 사용되는지 여부에 의존할 것임이 인식될 것이다. 투여의 빈도는 또한 치료되는 대상체 내에서 상기 활성제의 반감기에 의해 영향을 받을 것이다. 투여되어야 하는 최적의 복용량은 본 기술분야의 기술자에 의해 결정될 수 있고, 사용되는 특정 RNA 구조체, 핵산 서열, 발현

카세트, 벡터 또는 약학적 조성물, 약학적 조성물의 강도, 투여 방식, 및 바이러스 감염의 타입 및 진행에 따라 변할 것이다. 대상체의 연령, 중량, 성별, 식이, 및 투여 시간을 포함하는 치료되는 특정 대상체에 의존하는 부가적인 인자는 복용량을 조정할 필요성을 야기할 것이다.

- [0319] 일반적으로, 0.001  $\mu\text{g}/\text{kg}$ 의 체중 내지 10  $\text{mg}/\text{kg}$ 의 체중, 또는 0.01  $\mu\text{g}/\text{kg}$ 의 체중 내지 1  $\text{mg}/\text{kg}$ 의 체중의 본 발명의 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물의 일일 용량이 사용되는 활성제에 따라 질환을 치료, 개선, 또는 예방하기 위해 사용될 수 있다.
- [0320] 일일 용량은 단일 투여(예컨대 단일 일일 주사 또는 코 스프레이의 흡입)로서 제공될 수 있다. 대안적으로, 상기 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 하루 동안에 2회 이상의 투여가 필요할 수 있다. 예로서, 상기 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 0.07  $\mu\text{g}$  내지 700  $\text{mg}$ (즉, 70  $\text{kg}$ 의 체중으로 추정함)의 2회(또는 치료되는 질환의 증세에 따라 그 이상)의 일일 용량으로 투여될 수 있다. 치료를 받는 환자는 기상시 제1 용량을, 이후 저녁(2회 용량 요법이라면)에 또는 이후에 3 또는 4시간 간격으로 제2 용량을 취할 수 있다. 대안적으로, 느린 방출 장치는 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물을 반복된 용량을 투여할 필요 없이 환자에게 최적의 용량을 제공하기 위해 사용될 수 있다.
- [0321] 그러나, 바람직하게는, 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 매주 용량, 보다 바람직하게는 격주 용량으로 제공될 수 있다.
- [0322] 제약 산업에 의해 종래 도입되는 것들과 같은 공지된 절차(예컨대 생체내 실험, 임상 시험, 등)는 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트 또는 벡터의 구체적인 제형 및 정확한 치료 요법(예컨대, 제제의 일일 용량 및 투여 빈도)을 형성하기 위해 사용될 수 있다.
- [0323] "대상체"는 척추동물, 포유동물, 또는 가축일 수 있다. 따라서, 본 발명에 따른 조성물 및 약제는 임의의 포유동물, 예를 들면 가축(예컨대, 말), 애완동물을 치료하기 *dnlgo* 사용될 수 있거나, 다른 수의과 적용분야에 사용될 수 있다. 그러나, 가장 바람직하게는, 상기 대상체는 인간이다.
- [0324] 상기 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물의 "치료적 유효량"은 대상체에게 투여될 때 임의의 해당 질환을 개선, 예방 또는 치료하기 위해 필요한 전술한 양인 임의의 양이다.
- [0325] 예를 들면, 본 발명의 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 약 0.0001  $\text{mg}$  내지 약 800  $\text{mg}$ , 바람직하게는 약 0.001  $\text{mg}$  내지 약 500  $\text{mg}$ 으로 사용될 수 있다. 상기 레플리콘, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물의 양은 약 0.01  $\text{mg}$  내지 약 250  $\text{mg}$ , 가장 바람직하게는 약 0.01  $\text{mg}$  내지 약 1  $\text{mg}$ 의 양인 것이 바람직하다. 바람직하게는, 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 1-200  $\mu\text{g}$ 의 용량으로 투여된다.
- [0326] 본 명세서에서 나타낼 때, "약학적으로 허용가능한 비히클"은 약학적 조성물을 제형화하는데 유용하다고 본 기술분야의 기술자에게 알려진 임의의 공지된 화합물 또는 공지된 화합물들의 조합이다.
- [0327] 한 구현예에서, 상기 약학적으로 허용가능한 비히클은 고체일 수 있고, 상기 조성물은 분말 또는 정제의 형태일 수 있다. 고체인 약학적으로 허용가능한 비히클은 향미제, 윤활제, 가용화제, 현탁화제, 염료, 충전제, 활택제, 압착 보조제, 불활성 결합제, 감미제, 보존제, 염료, 코팅물, 또는 정제-붕괴제로서도 작용할 수 있는 하나 이상의 물질을 포함할 수 있다. 상기 비히클은 또한 캡슐화 물질일 수 있다. 분말에서, 상기 비히클은 본 발명에 따른 세분된 활성제와 혼합되는 세분된 고체이다. 정제에서, 상기 활성제(예컨대, 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물)는 필요한 압착 특성을 갖는 비히클과 적합한 비율로 혼합되고 원하는 형태 및 크기로 압착될 수 있다. 상기 분말 및 정제는 바람직하게는 상기 활성제를 최대 99%까지 함유한다. 적합한 고체 비히클은, 예를 들면 인산칼슘, 스테아르산마그네슘, 탈크, 당, 락토오스, 덱스트린, 전분, 젤라틴, 셀룰로오스, 폴리비닐피롤리돈, 저융점 왁스 및 이온 교환 수지를 포함한다. 다른 구현예에서, 상기 약학적 비히클은 겔일 수 있고, 상기 조성물은 크림 등의 형태일 수 있다.
- [0328] 그러나, 상기 약학적 비히클은 액체일 수 있고, 상기 약학적 조성물은 용액의 형태이다. 액체 비히클은 용액, 현탁액, 에멀전, 시럽, 엘릭시르 및 가압 조성물의 제조에 사용된다. 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 약학적으로 허용가능한 액체 비히클, 예컨대 물, 유기 용매, 이들의 혼합물 또는 약학적으로 허용가능한 오일 또는 지방에 용해 또는 현탁될 수 있다. 상기 액체 비히클은 다른 적합한 약학적 첨가제, 예컨대 가용화제, 에멀전화제, 버퍼, 보존제, 감미제, 향미제, 현탁화제, 증점제, 착색제, 점도 조절제, 안정화제 또는 삼투-조절제를 함유할 수 있다. 경구 및 비경구 투여를 위한 액체 비히클의 적합한

예는 (상기와 같은 첨가제, 예컨대 셀룰로오스 유도체, 바람직하게는 나트륨 카르복시메틸 셀룰로오스 용액을 부분적으로 함유하는) 물, (1가 알코올 및 다가 알코올, 예컨대 글리콜을 포함하는) 알코올 및 그의 유도체, 및 오일(예컨대 분획화된 코코넛 오일 및 아라키스 오일)을 포함한다. 비경구 투여의 경우, 상기 비히클은 또한 오일성 에스테르, 예컨대 에틸 올레에이트 및 이소프로필 미리스테이트일 수 있다. 멸균 액체 비히클은 비경구 투여용 멸균 액체 형태 조성물에서 유용하다. 가압 조성물을 위한 액체 비히클은 할로겐화 탄화수소 또는 다른 약학적으로 허용가능한 추진제(propellant)일 수 있다.

[0329] 멸균 용액 또는 현탁액인 액체 약학적 조성물은, 예를 들면, 피하, 진피내, 척추강내, 경막외, 복강내, 정맥내, 특히 근육내 주사에 의해 이용될 수 있다. 본 발명의 핵산 서열, 또는 발현 카세트는 멸균수, 식염수, 또는 다른 적절한 멸균 주사가능한 매체를 이용하여 투여 시점에 용해 또는 현탁될 수 있는 멸균 고체 조성물로서 제조될 수 있다.

[0330] 본 발명의 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 다른 용질 또는 현탁화제(예를 들면, 용액이 등장압이 되게 하기에 충분한 식염수 또는 글루코오스), 담즙 염, 아카시아, 젤라틴, 소르비탄 모노올레이트, 폴리소르베이트 80(소르비톨의 올레에이트 에스테르 및 에틸렌 옥사이드로 공중합된 그의 무수물) 등을 함유하는 멸균 용액 또는 현탁액의 형태로 경구 투여될 수 있다. 본 발명에 따른 RNA 구조체, 핵산 서열, 발현 카세트, 벡터 또는 약학적 조성물은 또한 액체 또는 고체 조성물 형태로 경구 투여될 수 있다. 경구 투여용으로 적합한 조성물은 알약, 캡슐, 과립, 정제, 및 분말과 같은 고체 형태, 및 용액, 시럽, 엘릭시르, 및 현탁액과 같은 액체 형태를 포함한다. 비경구 투여를 위해 유용한 형태는 멸균 용액, 에멀전, 및 현탁액을 포함한다.

[0331] 본 발명은 그의 변이체 또는 단편을 포함하여 실질적으로 본 명세서에 나타난 임의의 서열의 아미노산 또는 핵산 서열을 포함하는 임의의 핵산 또는 펩티드 또는 그의 변이체, 유도체 또는 유사체로 확장됨이 인식될 것이다. 용어 "실질적으로 아미노산/뉴클레오타이드/펩티드 서열", "변이체" 및 "단편"은 본 명세서에 나타난 임의의 하나의 서열의 아미노산/뉴클레오타이드/펩티드 서열과 적어도 40% 서열 동일성(identity), 예를 들면 서열 번호 1 내지 서열번호 55 등으로 확인된 서열과 40% 동일성을 갖는 서열일 수 있다.

[0332] 나타난 임의의 서열과 65% 이상, 보다 바람직하게는 70% 이상, 보다 더 바람직하게는 75% 이상의 서열 동일성, 보다 더 바람직하게는 80% 이상의 서열 동일성을 갖는 아미노산/폴리뉴클레오타이드/폴리펩티드 서열이 또한 예측된다. 바람직하게는, 상기 아미노산/폴리뉴클레오타이드/폴리펩티드 서열은 나타난 임의의 서열과 적어도 85% 동일성, 보다 바람직하게는 적어도 90% 동일성, 보다 더 바람직하게는 적어도 92% 동일성, 보다 더 바람직하게는 적어도 95% 동일성, 보다 더 바람직하게는 적어도 97% 동일성, 보다 더 바람직하게는 적어도 98% 동일성, 가장 바람직하게는 본 명세서에 나타난 임의의 서열과 적어도 99% 동일성을 갖는다.

[0333] 숙련된 기술자는 2개의 아미노산/폴리뉴클레오타이드/폴리펩티드 서열 사이에 동일성 백분율을 계산하는 방법을 인식할 것이다. 2개의 아미노산/폴리뉴클레오타이드/폴리펩티드 서열 사이의 동일성 백분율을 계산하기 위하여, 상기 2개의 서열의 정렬이 먼저 준비되어야 하고, 이어서 서열 동일성 값을 계산한다. 2개의 서열에 대한 동일성 백분율은 다음에 따라 상이한 값을 취할 수 있다: (i) 서열을 정렬하기 위해 사용된 방법, 예를 들면, ClustalW, BLAST, FASTA, 스미스-워터맨(상이한 프로그램에서 실행됨), 또는 3D 비교로부터의 구조적 정렬; 및 (ii) 정렬 방법에 의해 사용된 파라미터, 예를 들면, 국소 대 전체 정렬, 사용된 쌍별-점수 매트릭스(예컨대 BLOSUM62, PAM250, Gonnet 등), 및 갭-페널티, 예컨대 기능적 형태 및 상수.

[0334] 정렬을 만든 후, 2개의 서열 사이에 동일성 백분율을 계산하는 많은 다양한 방식이 있다. 예를 들면, 동일성의 수를 다음에 의해 나눌 수 있다: (i) 가장 짧은 서열의 길이; (ii) 정렬의 길이; (iii) 서열의 평균 길이; (iv) 비-갭 위치의 수; 또는 (v) 돌출부(overhang)를 배제한 등가물 위치의 수. 아울러, 동일성 백분율은 또한 매우 길이 의존성임이 인식될 것이다. 따라서, 서열의 쌍이 짧을수록 더 높은 서열 동일성이 우연히 일어날 것으로 예상할 수 있다.

[0335] 따라서, 단백질 또는 DNA 서열의 정확한 정렬은 복잡한 공정임이 인식될 것이다. 대중적인 다중 정렬 프로그램 ClustalW(Thompson *et al.*, 1994, Nucleic Acids Research, 22, 4673-4680; Thompson *et al.*, 1997, Nucleic Acids Research, 24, 4876-4882)는 본 발명에 따른 단백질 또는 DNA의 다중 정렬을 생성하기 위한 바람직한 방법이다. ClustalW에 대한 적합한 파라미터는 다음과 같을 수 있다: DNA 정렬의 경우: 갭 개방 페널티 = 15.0, 갭 연장 페널티 = 6.66, 및 매트릭스 = Identity. 단백질 정렬의 경우: 갭 개방 페널티 = 10.0, 갭 연장 페널티 = 0.2, 및 매트릭스 = Gonnet. DNA 및 단백질 정렬의 경우: ENDGAP = -1, 및 GAPDIST = 4. 본 기술분야의 기술자는 최적의 서열 정렬을 위해서는 상기 및 다른 파라미터들을 변화시킬 필요가 있을 수 있음을 인식할 것이다.

[0336] 바람직하게는, 2개의 아미노산/폴리뉴클레오티드/폴리펩티드 서열 사이의 동일성 백분율의 계산은 이후 이러한 정렬로부터  $(N/T) \times 100$ 으로서 계산될 수 있으며, 여기서 N은 서열이 동일한 잔기를 공유하는 위치의 수이고, T는 갭을 포함하고 돌출부를 포함하거나 배제한 비교된 위치의 총 수이다. 바람직하게는, 돌출부는 상기 계산에 포함된다. 따라서, 2개의 서열 사이의 동일성 백분율을 계산하기 위한 가장 바람직한 방법 (i) 예를 들면, 상기에 나타난 것과 같은 적합한 세트의 파라미터를 이용하여 ClustalW 프로그램을 이용하여 서열 정렬을 준비하는 단계; 및 (ii) N 및 T의 값을 다음의 식: 서열 동일성 =  $(N/T) \times 100$  내로 삽입하는 단계를 포함한다.

[0337] 유사한 서열을 확인하기 위한 대안적인 방법은 본 기술분야의 기술자에게 알려질 것이다. 예를 들면, 실질적으로 유사한 뉴클레오티드 서열은 엄격한 조건 하에서 DNA 서열 또는 그의 상보체(complement)와 교잡하는 서열에 의해 암호화될 것이다. 엄격한 조건이란 것은, 본 발명자들은 뉴클레오티드가 대략 45°C에서  $3 \times$  염화나트륨/시트르산나트륨(SSC)과, 이어서 대략 20-65°C에서  $0.2 \times$  SSC/0.1% SDS에서 적어도 1회의 세척 후 필터-결합된 DNA 또는 RNA에 혼성화하는 것을 의미한다. 대안적으로, 실질적으로 유사한 폴리펩티드는, 예를 들면, 서열번호 1 내지 서열번호 57에 나타난 서열과 적어도 1개이지만, 5, 10, 20, 50 또는 100개 미만의 아미노산이 상이할 수 있다.

[0338] 유전자 코드의 축퇴(degeneracy)로 인하여, 본 명세서에 기술된 임의의 핵산 서열은 이것에 의해 암호화된 단백질의 서열에 실질적으로 영향을 미치지 않으면서 다양해지거나 변화되어 그의 기능적 변이체를 제공할 수 있음이 명확하다. 적합한 뉴클레오티드 변이체는 서열 내에서 동일한 아미노산을 암호화하여 침묵적(동의적) 변화를 생산하는 상이한 코돈의 치환에 의해 변경된 서열을 갖는 것들이다. 다른 적합한 변이체는 상동성 뉴클레오티드 서열을 갖지만 치환되는 아미노산과 유사한 생물물리학적 특성의 측체를 갖는 아미노산을 암호화하는 상이한 코돈의 치환에 의해 변경되어 보존된 변화를 생산하는 서열의 전부 또는 일부를 포함하는 것들이다. 예를 들면, 작은 비극성의 소수성 아미노산은 글리신, 알라닌, 루이신, 이소루이신, 발린, 프롤린, 및 메티오닌을 포함한다. 큰 비극성의 소수성 아미노산은 페닐알라닌, 트립토판 및 티로신을 포함한다. 극성의 중성 아미노산은 세린, 트레오닌, 시스테인, 아스파라긴 및 글루타민을 포함한다. 양을 대전된(염기성) 아미노산은 리신, 아르기닌 및 히스티딘을 포함한다. 음으로 대전된(산성) 아미노산은 아스파르트산 및 글루탐산을 포함한다. 따라서, 어느 아미노산이 유사한 생물물리학적 특성을 갖는 아미노산으로 교체될 수 있음이 인식될 것이고, 숙련된 기술자는 상기 아미노산을 암호화하는 뉴클레오티드 서열을 알 것이다.

[0339] (임의의 부속 청구항, 요약 및 도면을 포함하는) 본 명세서에 기술된 모든 특성들 및/또는 여기에 개시된 임의의 방법 또는 공정의 모든 단계들은, 이러한 특성 및/또는 단계의 적어도 일부가 상호 배타적인 조합을 제외하고는, 임의의 조합으로 임의의 상기 측면들과 조합될 수 있다.

### 도면의 간단한 설명

[0340] 본 발명을 보다 잘 이해하고, 본 발명의 구현예가 효과를 내기 위해 수행될 수 있는 방식을 보이기 위하여, 이제 실시예로서 부속하는 도면에 대한 참조가 행해질 것이다:

도 1은 베네주엘라 말 뇌염 바이러스(VEEV) 백본에 기반하는 자가-증폭 RNA 레플리콘 또는 구조체의 한 구현예의 개략도를 보여준다. 소위 '스틸티콘' 백터는 비-구조 단백질(NSP1-4), 및 GOI(관심있는 유전자)의 상류 또는 하류에 있는 선천적 억제성 단백질(IIP)을 암호화하는 saRNA 레플리콘/구조체이다.

도 2는 계놈 복제가 센서 분자인 MDA5 및 PACT에 의해 인식되는 dsRNA를 야기함을 보여준다. 상기 센서는 전사 인자(NF- $\kappa$ B, IRF-3, -7), RNA 증폭 및 단백질 발현을 직접 억제하는 제한 인자(점선)의 활성화를 포함하는 하류의 캐스케이드를 활성화하기 위한 신호를 전달한다. PIV-V 또는 ORF4a의 발현은 MDA5 및 PACT에 의한 dsRNA의 선천적 인식을 차단하여 합성 RNA로부터의 레플리콘 RNA 증폭 및 단백질 발현을 제한하는 하류 캐스케이드의 활성화를 방지한다.

도 3은 a) 시험관내에서 IIP-암호화VEEV 레플리콘의 스크리닝을 보여준다. 세포를 리포터 단백질로서 루시퍼라아제를 함유하는 2개 배치의 RNA로 형질감염시키고, 24시간 후에 단백질 발현에 대해 평가하였다. HeLa 및 MRC5는 HEK와 비교하여 보다 온전한 IFN 발현 경로를 갖는 것으로 알려져 있다. b) 야생형에 비해 발현에 있어서의 배수 변화로서 도 3a를 나타낸 데이터이다.

도 4는 리포터 단백질로서 반딧불이 루시퍼라아제를 이용하여 BALB/c 및 BL/6 마우스에서 IIP-암호화 레플리콘(#6 및 #8)의 스크리닝을 보여준다. BL/6 마우스는 BALB/c보다 IFN을 더 발현하는 것으로 알려져 있고, 따라서 IIP는 3일에 더 높은 루시퍼라아제를 발현하고, >14일 지속되는 더 긴 발현을 할 수 있다.

도 5는 리포터 단백질로서 가우시아(gaussia) 루시페라아제를 이용하여 BALB/c 및 BL/6 마우스에서 IIP-암호화 레플리콘(#6 및 #8)의 스크리닝을 보여준다. BL/6 마우스는 BALB/c보다 IFN을 더 발현하는 것으로 알려져 있고, 따라서 IIP는 가용성 리포터 단백질을 더 길게 발현할 수 있고, >14일 지속된다.

도 6은 GOI를 포함하는 RNA 구조체를 암호화하는 발현 벡터의 한 구현예의 구조체 지도를 보여준다.

도 7은 GOI - MERS-CoV ORF4a를 포함하는 RNA 구조체를 암호화하는 발현 벡터의 한 구현예의 구조체 지도를 보여준다.

도 8은 GOI - PIV5를 포함하는 RNA 구조체를 암호화하는 발현 벡터의 한 구현예의 구조체 지도를 보여준다.

도 9는 본 발명의 구조체, 및 야생형 및 IIP VEEV 레플리콘으로부터의 시험관내 단백질 발현의 개략적인 도면을 보여준다. a) 야생형 및 시스(cis)-암호화 IIP VEEV 레플리콘의 개략도. b) 자가-중복 RNA의 선천적 감지의 개략도. c) 상대적 광 단위(RLU)로서 측정되는 HEK 293T.17, HeLa 및 MRC5 세포에서의 반딧불이 루시페라아제 saRNA의 시험관내 형질감염. 막대는 n=3에 대하여 야생형 VEEV 대조군에 대해 표준화된 평균 배수 변화±표준 편차를 나타낸다.

도 10은 C57BL6/J 마우스에서 WT 및 MERS CoV-2 레플리콘의 용량 적정을 보여준다. 단백질 발현은 0.2, 2 또는 20  $\mu$ g의 RNA를 근육내 주사한 후 7일 및 10일에 정량되었다. 각각의 점은 단일 마우스를 나타내고, 막대는 n=10으로 평균±SEM을 나타낸다. \*는 다중 비교로 크루스칼-월리스(Kruskal-Wallis) 테스트를 이용해 평가될 때  $p<0.05$ 의 유의성을 시사한다.

도 11은 C57BL6/J 마우스에서 JAK 억제제 록솔리티닙(ruxolitinib)을 이용한 WT 및 MERS CoV\_2 레플리콘의 공동-제형화를 보여준다. 단백질 발현은 100  $\mu$ g의 록솔리티닙과 함께 5  $\mu$ g의 RNA를 근육내 주사한 후 4, 7, 10 및 14일(각각, a, b, c, 및 d)에 정량되었다. 각각의 점은 단일 마우스 다리를 나타내고, 막대는 n=10을 평균±SEM을 나타낸다.

도 12는 인간 피부 외식편(explant)에서 eGFP ± MERS CoV\_2 RNA(MERS-CoV ORF4a에 해당함)(0.2, 2 또는 20  $\mu$ g)(a,b) 또는 ± 록솔리티닙(0.1, 1, 10 또는 100  $\mu$ g)(c,d)의 단백질 발현을 보여준다. eGFP 발현 세포의 수(%GFP+ 세포)(a,c) 및 세포 당 총 단백질 발현(GFP 중앙 형광 강도(MFI))(b,d)를 주사 후 72시간에 정량하였다. 각각의 점은 n=3으로 평균±SEM을 나타낸다. \*은 다중 비교로 크루스칼-월리스 테스트를 이용해 평가될 때  $p<0.05$ 의 유의성을 시사한다.

도 13은 토끼에서 RABV ± MERS-CoV\_2(MERS-CoV ORF4a에 해당함)의 면역원성을 보여준다. a) n=5로 0 및 4주에 20  $\mu$ g의 프라임(prime) 및 부스트(boost)로 근육내 면역화 이후 RABV 항원-특이적 IgG 항체 역가. b) n=5로 위형화된 RABV 바이러스에 대한 중화 IC50으로서, 회색 점선은 검출 한계를 나타낸다. \*는 다중 비교로 크루스칼-월리스 테스트를 이용해 평가될 때  $p<0.05$ 의 유의성을 시사한다.

도 14는 상대적 광 단위(RLU)로서 측정되는 HEK 293T.17, HeLa 및 MRC5 세포에서의 반딧불이 루시페라아제 saRNA의 시험관내 형질감염을 보여준다. 막대는 n=3에 대하여 평균±표준 편차 대조군을 나타낸다. 'a' 및 'b'는 2개의 별도로 준비된 RNA의 배치(batch)를 시사한다.

도 15는 상대적 광 단위(RLU)로서 측정되는 a) 마우스(MEF), b) 토끼(RK13), c) 비-인간 영장류(LLC) 및 d) 인간(MRC5) 세포에서의 WT fLuc, MERS-CoV\_2 ORF4a 및 PIV-5 RNA의 시험관내 형질감염을 보여준다. 막대는 n=3에 대하여 평균±표준 편차 대조군을 나타낸다.

도 16은 BALB/c 및 C57BL6/J 마우스에서 세포내(반딧불이 루시페라아제) 및 분비된(가우시아 루시페라아제) 단백질의 생체내 루시페라아제 발현을 보여준다. 단백질 발현은 생체내 영상화 시스템(IVIS)을 이용하여 근육(a,c,e) 또는 혈청(b,d,f)에서 3일(a,b) 7일(c,d) 및 14일(e,f)에 정량되었다. 각각의 점은 단일 마우스를 나타내고, 막대는 fLuc의 경우 n=10 및 gLuc의 경우 n=5로 평균±SEM을 나타낸다.

도 17은 0.2  $\mu$ g의 eGFP RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅(gating)으로 오버레이(overlay)한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터(unsupervised cluster)의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 18은 2  $\mu$ g의 eGFP RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 19는 20  $\mu\text{g}$ 의 eGFP RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 20은 0.2  $\mu\text{g}$ 의 eGFP-PIV-5 RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 21은 2  $\mu\text{g}$ 의 eGFP-PIV-5 RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 22는 20  $\mu\text{g}$ 의 eGFP-PIV-5 RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 23은 0.2  $\mu\text{g}$ 의 eGFP-MERS-CoV\_2 RNA(MERS-CoV ORF4a에 해당함)로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 24는 2  $\mu\text{g}$ 의 eGFP-MERS-CoV\_2 RNA로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 25는 20  $\mu\text{g}$ 의 eGFP-MERS-CoV\_2 RNA(MERS-CoV ORF4a에 해당함)로 처리된 인간 피부 외식편에 대해 살아있는 세포(회색), eGFP+ 세포에 대한 게이팅으로 오버레이한 것(녹색), 표현형에 의해 분리된 것(청색)의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 플롯을 보여준다.

도 26은 유세포 분석에 의해 결정될 때 eGFP  $\pm$  독소 제형의 진피내(ID) 주사 후 인간 피부 외식편에 존재하는 세포 및 GFP+ 세포의 표현형적 동일성을 보여준다. a)  $n=3$ 으로 2  $\mu\text{g}$ 의 eGFP-암호화 saRNA  $\pm$  0.1, 1, 10 또는 100  $\mu\text{g}$ 의 독소로 처리된 인간 피부 외식편으로부터 추출된 총 세포 및 외식편 유래의 GFP-발현 피부 세포의 개체군(population)에서 세포의 동일성. b) GFP를 발현하는 각각의 표현형의 세포의 백분율. 세포는 다음의 항체를 이용해 확인되었다: 상피 세포(CD45-), 섬유아세포(CD90+), NK 세포(CD56+), 백혈구(CD45+), 랑게르한스 세포(CD1a+), 단핵구(CD14+), 수지상 세포(CD11c+), T 세포(CD3+), 및 B 세포(CD19+).

도 27은 마우스 및 래트에서 RABV  $\pm$  MERS-CoV\_2(MERS-CoV ORF4a에 해당함)의 면역원성을 보여준다. a)  $n=5$ 로 0 및 4주에 1  $\mu\text{g}$ 의 프라임 및 부스트로 마우스의 근육내 면역화 이후 RABV 항원-특이적 IgG 항체 역가. b)  $n=5$ 로 위형화된 RABV 바이러스에 대한 마우스의 중화 IC50으로서, 회색 점선은 검출 한계를 나타낸다.

a)  $n=5$ 로 0 및 4주에 20  $\mu\text{g}$ 의 프라임 및 부스트로 래트의 근육내 면역화 이후 RABV 항원-특이적 IgG 항체 역가. b)  $n=5$ 로 위형화된 RABV 바이러스에 대한 래트의 중화 IC50으로서, 회색 점선은 검출 한계를 나타낸다.

도 28은 saRNA 감지에 대한 PIV-5 V 및 MERS-CoV ORF4a의 제안된 메커니즘의 개략도를 보여준다.

도 29는 SARS-CoV-2 당단백질만을 암호화하는 saRNA(즉, IIP 없음)와 비교할 때 선천적 억제제 단백질인 MERS-ORF4a과 함께 공동-발현시에 HeLa 세포에서 본 발명의 한 구현예에 따른 saRNA로부터 SARS-CoV-2 당단백질의 증가된 발현을 보여주는 중앙 형광 강도(MFI) 데이터를 보여준다.

## 발명을 실시하기 위한 구체적인 내용

### 실시예

본 발명자들은 saRNA의 선천적 인식을 억제하는 것으로 알려져 있는 바이러스 유래의 시스-암호화 단백질이 선천적 감지를 약화시키고 saRNA 백신의 단백질 발현 및 면역원성 모두를 향상시킬 것임을 가정하였다. 따라서, 본 발명자들은 선천적 억제 단백질(IIPs) 및 관심있는 유전자를 함유하는 다양한 RNA 레플리콘을 디자인 및 테스트한 후, 상기 레플리콘이 (관심있는 유전자에 의해 암호화된) 세포내 및 분비된 단백질 발현 모두를 향상시키는지 여부를 특징분석하였다.

### 재료 및 방법

[0344] 선천적 억제성 단백질(IIP) 레플리콘의 클로닝

[0345] 반딧불이 루시페라아제, 가우시아 루시페라아제, 향상된 녹색 형광 단백질(eGFP), 광견병 당단백질(RABV) 및 베네주엘라 말 뇌염 유래의 레플리카아제(replicase)를 암호화하는 자가-증폭 RNA를 이전에 기술된 것과 같이 플라스미드 벡터 내로 클로닝하였다(53). 인터페론 억제 단백질의 라이브러리를 T2A 절단 부위(진뱅크 접근 번호 AAC97195.1)를 갖는 관심있는 유전자(fLuc, gLuc, eGFP 또는 RABV)의 일부로서 상기 벡터 벡본 내로 클로닝하였다. 상기 인터페론 억제 단백질은 다음의 진뱅크 접근 번호에서 발견될 수 있다: HSV-2 Us1(Z86099.2), HSV-1 Us1(AW069381.1), HSV-1 Us11(YP\_009137147.1), OV20.0L(AF053969.1), BVDV Npro(AIE38066.1), PIV-5 V(YP\_138513.1), MERS-CoV M(AHC74104.1), MERS-CoV ORF4a(AHC74090.1), 란가트 바이러스 NS5(AF253420) 및 인플루엔자 바이러스 NS1(DQ508893.1). 마우스에서의 연구를 위하여, N100D 돌연변이를 갖는 PIV-5 V 단백질을 사용하였다(45).

[0346] saRNA의 시험관내 전사

[0347] 시험관내 전사를 이용해 자가-증폭 RNA를 생산하였다. 플라스미드 DNA(pDNA)를 대장균(*E. coli*)(New England Biolabs, UK) 내로 형질전환시켰고, 100 µg/ml의 카르베니실린(Sigma Aldrich, UK)을 갖는 100 ml의 루리아 브로스(LB)에서 배양하였다. 이어서, 상기 pDNA를 Plasmid Plus MaxiPrep 키트(QIAGEN, UK)를 이용해 분리하였고, pDNA의 최종 농도를 NanoDrop One(ThermoFisher, UK)에서 측정하였다. 37°C에서 3시간 동안 MluI를 이용해 pDNA를 선형화하였다. 1 µg의 선형화된 pDNA 주형을 이용하여 mMachine™ T7 Transcription(Invitrogen, UK)에서 시험관내 형질감염을 위한 RNA를 준비하였고, 제조사의 프로토콜을 이용하여 MEGAClear™ Transcription Clean-Up Kit(Invitrogen, UK)을 이용해 정제하였다. 이전에 기술된 것과 같이 생체외 및 생체내 실험을 위한 RNA를 준비하였다(2). 캡핑되지 않은 RNA 전사체를 제조사의 프로토콜을 이용하여 37°C에서 2시간 동안 MEGAScript™ T7 Transcription 반응(Invitrogen, UK)을 이용해 1 µg의 선형화된 pDNA 주형을 이용하여 생산하였다. 이후, -20°C에서 밤새 LiCl 침전시킴으로써 전사체를 정제하였고, 4°C에서 20분 동안 14,000 RPM으로 원심분리하여 RNA를 펠렛화하였고, 70% EtOH로 1회 세정하였고, 4°C에서 5분 동안 14,000 RPM으로 다시 원심분리하였고, UltraPure H<sub>2</sub>O(Ambion, UK)에 재현탁시켰다. 제조사의 프로토콜을 이용하여 37°C에서 2시간 동안 ScriptCap™ Cap 1 캡핑 시스템 키트(CellScript, WI, USA)를 이용하여 정제된 전사체를 캡핑하였다. 이후, 상기 기술된 것과 같이 LiCl 침전을 이용해 캡핑된 전사체를 마지막으로 정제하였고, RNA 보관 버퍼(10 mM HEPES, 0.1 mM EDTA 및 100 mg/ml 트레할로오스)에 재현탁시켰고, 추가로 사용할 때까지 -80°C에 보관하였다.

[0348] saRNA 제형

[0349] 단백질 발현 실험을 위한 fLuc, gLuc 및 eGFP saRNA는 이전에 기술된 것과 같은 적정 방법을 이용하여 100 kDa pABOL로 복합체화하였다(2). 생체내 면역원성 실험을 위한 RABV saRNA는 8 kDa pABOL로 복합체화하였다. 간략하게, RNA 및 pABOL를 HEPES 버퍼(20 mM HEPES, H<sub>2</sub>O 중의 5 wt.% 글루코오스, pH 7.4)에 희석하였고, 10 ml/분의 유속에서 4:1의 부피 비(RNA 대 폴리머)로 NanoAssemblr 벤치탑 제형 단위(Precision Nanosystems, Inc., Vancouver, Canada)에서 조합하였다. 폴리머 대 saRNA의 최종 비는 45:1(w/w)였다. 폴리플렉스를 신중하게 제조하였고, 제조 1시간 내에 사용하였다. 공동-제형화를 위하여, 표시된 용량으로 록솔리티닙(록소, Selleck Chemicals, UK)을 폴리플렉스에 직접 첨가하였다.

[0350] 시험관내 형질감염

[0351] HEK293T.17 세포(ATCC, USA), HeLa 세포(ATCC, USA), MRC5 세포(ATCC, USA), 마우스 배아 섬유아세포(MEF) 세포(SigmaAldrich, UK), RK13 토끼 신장 세포(Public Health England, UK) 및 LLC-MK2 히말라야 원숭이 신장 세포(ATCC, USA)에서 형질감염을 수행하였다. 세포를 10%(v/v) 소 태아 혈청(FBS), 5 mg/ml L-글루타민 및 5 mg/ml 페니실린/스트렙토마이신(ThermoFisher, UK)(HEK, HeLa, MEF cells)을 함유하는 완전 둘베코 변형 이글 배지(cDMEM)(Gibco, Thermo Fisher, UK), 10%(v/v) 소 태아 혈청(FBS), 5 mg/ml L-글루타민 및 5 mg/ml 페니실린/스트렙토마이신(ThermoFisher, UK)(MRC5, RK13 cells)을 갖는 완전 변형 이글 배지(cMEM) 또는 1% 말 혈청(Gibco, ThermoFisher, UK)(LLC 세포)을 갖는 완전 배지199(cM199, SigmaAldrich, UK)에서 배양하였다. 형질감염 24시간 전에 투명한 96-웰 플레이트에서 웰 당 50,000 세포의 밀도로 세포를 플레이트팅하였다. 이후, 배양 배지를 완전히 제거하였고, 50 µl의 미리 데운 형질감염 배지(DMEM + 5 mg/ml L-글루타민, MEM + 5 mg/ml L-글루타민 또는 M199)로 교체하였다. 이후, 100 µl의 폴리플렉스 용액(100 ng의 saRNA를 함유함)을 각각의 웰에 첨가하였고, 4시간 동안 인큐베이션시켰다. 이후, 형질감염 배지를 완전히 제거하였고, cDMEM, cMEM 또는 cM199로

교체하였다. 24시간 후, 각각의 웰로부터 50  $\mu$ l의 배지를 제거하였고, 50  $\mu$ l의 ONE-Glo D-루시페린 기질(Promega, UK)을 첨가하였으며, 피펫팅에 의해 잘 혼합하였다. 이후, 분석을 위해 각각의 웰로부터의 총 부피를 백색 96-웰 플레이트(Costar)로 전달하였고, FLUOstar OMEGA 플레이트 판독기(BMG LABTECH, UK)에서 정량하였다. 대조군 웰로부터의 배경 형광을 각각의 웰로부터 차감하였다.

[0352] *마우스에서 생체내 f루시페라아제 발현*

[0353] 모든 동물은 영국 [Home Office Animals Scientific Procedures Act 1986] 및 지역 윤리 위원회에 따라 취급하였고, 영국 정부는 프로젝트 면허(P63FE629C) 및 개인 면허(IC37CBB8F)를 승인하였다. 음식 및 물은 자유롭게 공급하였다. 6-8주령의 암컷 BABL/c 마우스(Charles River, UK) 또는 C57BL/6 마우스(Charles River, UK)를 군(케이지 당 n=5)으로 수용하였고, 완전히 적응된 방에 수용하였다. 마우스를 50  $\mu$ l의 총 부피로 pABOL에 복합체화된 5  $\mu$ g의 fLuc saRNA로 양쪽 뒷다리에 또는 5  $\mu$ g의 GLuc로 한쪽 뒷다리에 근육내(IM) 주사하였다. 3, 4, 7, 10 또는 14일 후, 마우스를 이전에 기술된 것과 같이 fLuc에 대해 영상화하거나(54, 55), 제조사의 프로토콜에 따라 가우시아 루시페라아제 글로우 분석 키트(Pierce, Thermo Scientific, UK)를 이용해 GLuc를 분석하기 위하여 혈액을 수집하였다. 혈청에서의 단백질 발현을 FLUOstar OMEGA 플레이트 판독기(BMG LABTECH, UK)에서 정량하였다. 대조군 웰로부터의 배경 형광을 각각의 웰로부터 차감하였다. fLuc 분석을 위하여, 마우스를 150  $\mu$ l의 Xenolight RediJect D-루시페린 기질(PerkinElmer, UK)로 복강내(IP) 주사하였고, 10분 동안 휴식하게 하였다. 이후, 마우스를 이소플루란을 이용해 마취시켰고, 분자 영상화 소프트웨어 버전 5.0(Carestream Health, USA)이 장착된 생체내 영상화 시스템(IVIS) FX Pro(Kodak Co., Rochester, NY, USA)에서 2분동안 영상화하였다. 각각의 주사 부위로부터의 신호를 분자 영상화 소프트웨어를 이용해 정량하였고, 총 플럭스(p/s)로 표현하였다.

[0354] *마우스에서 생체내 g루시페라아제 발현*

[0355] 6-8 주령의 암컷 BABL/c 마우스(Charles River, UK) 또는 C57BL/6 마우스(Charles River, UK)를 군(n=5)으로 나누고, 완전히 적응된 방에 수용하였다. 마우스를 50  $\mu$ l의 총 부피로 5  $\mu$ g의 gLuc로 한쪽 뒷다리에 근육내(IM) 주사하였다. 3, 7 및 14일 후, 마우스를 꼬리 정맥을 통해 채혈하였다. 상기 혈액을 응고되게 하였고, 10,000 RPM으로 5분 동안 원심분리하였고, 혈청을 제거하였다. 이후, 모든 시점의 혈청을 20  $\mu$ l의 혈청 및 제조사의 프로토콜에 따라 제조된 100  $\mu$ l의 작업 용액을 이용하여 Pierce™ 가우시아 루시페라아제 글로우 분석 키트를 이용해 단일 96-웰 백색 플레이트(Costar)에서 분석하였다. FLUOstar Omega 플레이트 판독기(BMG LABTECH, UK)에서 발광을 분석하였고, 미처리(naive) 동물로부터의 배경을 각각의 샘플로부터 차감하였다.

[0356] *마우스, 래트 및 토끼의 백신화*

[0357] BALB/c 마우스, 스프라그 돌리 래트(Sprague Dawley) 및 뉴질랜드 백색 토끼를 50  $\mu$ l(마우스) 또는 100  $\mu$ l(래트, 토끼)의 총 부피로 pABOL과 함께 제형화된 1  $\mu$ g(마우스) 또는 20  $\mu$ g(래트, 토끼)의 RABV-암호화 saRNA로 한쪽 뒷다리에 IM으로 면역화하였다. 처음 프라임 주사 4주 후에 부스트 주사를 하였다. 연구 개시 0, 4 및 6주 후에 혈액을 수집하였고, 10,000 RPM으로 5분 동안 원심분리하였다. 이후, 혈청을 부었고, 추가적으로 분석할 때까지 -80℃에서 보관하였다.

[0358] *RABV-특이적 ELISA*

[0359] 반정량적 면역글로불린 ELISA 프로토콜을 이전에 기술된 것과 같이 수행하였다(56). 간략하게, 0.5  $\mu$ g/ml의 RABV-코팅된 ELISA 플레이트를 PBS 중의 1%(w/v) 소 혈청 알부민(BSA) 및 0.05%(v/v) Tween-20으로 차단하였다. 세척 후, 희석된 혈청 샘플을 상기 플레이트에 첨가하였고, 2시간 동안 인큐베이션하였다. 이후, 상기 플레이트를 세척하였으며, 마우스 ELISA를 위해 1:4000 희석된 항-마우스 IgG-HRP(Southern Biotech, UK)를 첨가하였고, 래트 ELISA를 위해 1:4000 희석된 염소 항-래트 IgG-HRP(Southern Biotech, UK)를 첨가하였고, 토끼 ELISA를 위해 1:10000 희석된 마우스 항-토끼 IgG-HRP(Sigma, UK)를 첨가하였다. ELISA 플레이트 웰을 항-마우스 카파(1:1,000) 및 람다(1:1,000) 경쇄(Serotec, UK)로 코팅하고, PBS 중의 1%(w/v) BSA/0.05%(v/v) Tween-20으로 차단하고, 세척하고, 정제된 IgG(Southern Biotech, UK)를 1000 ng/ml에서 출발하여 5배 연속 희석으로 적정해 첨가함으로써 마우스 표준물을 제조하였다. ELISA 플레이트 웰을 정제된 래트 IgG(R & D Systems, UK)를 1000 ng/ml에서 출발하여 5배 연속 희석으로 적정해 코팅함으로써 래트 표준물을 제조하였다. ELISA 플레이트 웰을 1:1250 항-토끼 IgG Fc(Milipore)로 코팅하고, PBS 중의 1%(w/v) BSA/0.05%(v/v) Tween-20으로 차단하고, 세척하고, 정제된 토끼 IgG(AbD Serotech, UK)를 1000 ng/ml에서 출발하여 5배 연속 희석으로 적정해 첨가함으로써 토끼 표준물을 제조하였다. 샘플 및 표준물을 3,3',5,5'-테트라메틸벤지딘(TMB)을 이용해 현상하였다. 5분 후에 중지 용액(Insight Biotechnologies, UK)으로 상기 반응을 중지하였다.

SoftMax Pro GxP v5 소프트웨어를 이용해 분광광도계(VersaMax, Molecular Devices, UK)로 흡광도를 관독하였다.

[0360] *RABV 미세중화 분석법*

[0361] 0, 4 및 6주 샘플에 대해 위형화된 광견병 미세중화를 수행하였다. BHK-21 세포를 cDMEM에서 10,000 세포/웰로 96-웰 플레이트에 분주하였다. 혈청을 56℃에서 열-불활성화시킨 후, cDMEM에서 1:5 연속 희석으로 희석하였다. 이후, 샘플을 50  $\mu$ l에서 100 TCID<sub>50</sub>의 농도로 동일한 부피의 유사-바이러스로 희석하였고, 37℃에서 1시간 동안 인큐베이션한 후, BHK-21 세포에 첨가하였고, 37℃에서 48시간 동안 배양하였다. 이후, 세포를 용해시켰고, Bright-Glo 루시퍼라아제 분석법(Promega, UK)을 이용해 루시퍼라아제 활성을 정량하였다. 이후, 분석을 위해 각각의 웰로부터의 총 부피를 백색 96-웰 플레이트(Costar)로 전달하였고, FLUOstar OMEGA 플레이트 판독기(BMG LABTECH, UK)에서 정량하였고, 각각의 샘플에 대해 IC<sub>50</sub>을 계산하였다.

[0362] *인간 피부 외식편 배양 및 주사*

[0363] 생체의 연구를 위하여, 인간 피부 조직의 외과적으로 절단된 검체를 찰링 크로스 병원(Charing Cross Hospital, Imperial NHS Trust, London, UK)에서 수집하였다. 모든 조직은 임페리얼 칼리지 런던(Imperial College London)의 지역 연구 윤리 위원회에 의해 승인된 프로토콜(MED\_RS\_11\_014) 하에 선택적 복부성형 또는 유방절제 수술을 받는 환자로부터 서면 동의를 받은 후 수집하였다. 사용시까지 피부 조직을 냉장하였고, 1 cm<sup>2</sup> 절편으로 잘라 내었고, 37℃, 5% CO<sub>2</sub>에서 2 ml의 cDMEM으로 12-웰 플레이트에서 배양하였다. Micro-Fine Demi 0.3 ml 주사기(Becton Dickinson, UK)를 이용하여 50  $\mu$ l의 총 부피에서 2  $\mu$ g의 saRNA의 용량으로 외식편에 진피내(ID) 주사하였다. 배양 기간 동안 매일 배지를 교체하였다.

[0364] *유세포 분석*

[0365] 주사 72시간 후, 피부 외식편을 손질하여 피하 지방 층을 제거하였고, 표피 및 진피 층을 가위로 잘 다지고, 1 mg/ml 콜라게나아제 P(Sigma Aldrich, UK) 및 5 mg/ml 디스파아제 II(Sigma Aldrich, UK)가 보충된 2 ml DMEM에서 회전형 교반기 상에서 37℃에서 4시간 동안 인큐베이션하였다. 이후, 소화물을 70  $\mu$ m 세포 여과기를 통해 여과하였고, 1,750 RPM으로 5분 동안 원심분리하였다. 이후, 세포를 100  $\mu$ l의 FACS 버퍼(PBS + 2.5% FBS)에 재현탁시켰고, FACS 버퍼에 1:400으로 희석된 고정형 수성 살아있는/죽은 세포 염색(ThermoFisher, UK)으로 얼음 상에서 20분 동안 염색하였다. 이후, 샘플을 1 ml의 FACS 버퍼로 세척하였고, 1,750 RPM으로 5분 동안 원심분리하였고, 다음의 항체 혼합물로 염색하였다: CD3-V450(BioLegend, UK), CD14-Qdot605(BioLegend, UK), CD19-BV650(BioLegend, UK), CD56-BV711(BioLegend, UK), CD1a-PerCP-eFluor710(BioLegend, UK), CD11c-PE(BioLegend, UK), CD90-PE-Cy7(BioLegend, UK) 및 CD45-AF700(BioLegend, UK). 이후, 샘플을 1 ml의 FACS 버퍼로 세척하였고, 1,750 RPM으로 5분 동안 원심분리하였고, 250  $\mu$ l의 PBS에 재현탁시킨 후, 1.5% 파라포름알데히드의 최종 농도로 250  $\mu$ l의 3% 파라포름알데히드로 고정하였고, 유세포 분석할 때까지 냉장시켰다. 샘플을 100,000 획득된 세포 이벤트로 FACSDiva 소프트웨어(BD Biosciences, UK)를 이용하여 LSRFortessa(BD Biosciences, UK) 유세포 분석기로 분석하였다. 이전에 기술된 것과 같이 게이팅 전략을 수행하였고(58), FlowJo version 10(FlowJo LLC, Oregon, USA)을 이용하여 GFP<sup>+</sup> 세포의 표현형적 동일성을 정량하였다. 살아있는 세포의 비지도 클러스터의 t-분포 확률적 임베딩(tSNE) 분석을 1000 반복, 30의 최적화(perplexity), 15196의 학습 속도(learning rate), Exact(vantage point tree) KNN 알고리즘 및 바르네스-허트(Barnes-Hut) 구배 알고리즘을 이용하여 FlowJo에서 수행하였다.

[0366] *SARS-CoV-2 당단백질의 시험관내 작업*

[0367] 본 발명자들은 플라스미드 벡터를 사용하여 트리니다드 당나귀 베네주엘라 말 뇌염 바이러스주(VEEV) 알파바이러스 계통에 기반하여 자가-중복 RNA(saRNA) 레플리콘을 합성하였다. 서브-계통 프로모터 유래의 바이러스 구조 단백질을 신규한 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2): 진뱅크 접근 번호: QHD43416.1의 표면 '스파이크' 당단백질로 교체하였다. 본 발명자들은 올리고뉴클레오타이드 스트링(GeneArt, Germany)을 이용하여 상기 SARS-CoV-2 유전자를 암호화하는 올리고뉴클레오타이드 단편을 합성하였고, 김슨(Gibson) 조립 방법(NEB Ltd, UK)을 이용해 플라스미드 벡터 내로 이들을 조립하였다. 상기 SARS-CoV-2 레플리콘 벡터의 추가적인 변형으로, MERS-CoV ORF4a(AHC74090.1)를 암호화하는 올리고뉴클레오타이드 스트링을 합성하였고(GeneArt, Germany), 연속적 개방 해독 틀에서 푸린/T2A 서열(서열번호 26)의 변이체를 이용하여 연결된 SARS-CoV-2 코딩 영역의 3'에 삽입하여, 새로운 SARS-CoV-2 ORF4a 벡터를 생성하였다. 상기 SARS-CoV-2 및 SARS-CoV-2 ORF4a saRNA 모두

를 이용하여 별도로 세포를 형질감염시킨 후, 폴리클론 항체로 염색하여 발현을 조사하였다. 간략하게, 형질감염 24시간 후, 세포를 수확하였고,  $1 \times 10^7$  세포/ml의 농도로 1 ml의 FACS 버퍼(PBS + 2.5% FBS)에 재현탁시켰다. 100 마이크로리터의 재현탁된 세포를 FACS 튜브에 넣었고, 50  $\mu$ l의 1:400으로 희석된 살아있는/죽은 고정형 수성 죽은 세포 염색(Thermo Fisher Scientific, UK)으로 얼음 상에서 20분 동안 염색하였다. 이후, 세포를 2.5 ml의 FACS 버퍼로 세척하였고, 1750 RPM으로 7분 동안 원심분리하였다. 원심분리 후, 세포를 2.5  $\mu$ g의 SARS-CoV 스파이크 단백질 폴리클론 항체(PA1-41165, Thermo Fisher Scientific, UK)로 얼음 상에서 30분 동안 염색하였고, 이후 2.5 ml의 FACS 버퍼로 세척하고, 1750 RPM으로 7분 동안 원심분리하였다. 이후, 세포를 0.4  $\mu$ g의 FITC 염소 항-토끼 IgG(BD Pharmingen, UK)로 얼음 상에서 30분 동안 염색하였다. 인큐베이션 후, 세포를 2.5 ml의 FACS 버퍼로 세척하였고, 1750 RPM으로 7분 동안 원심분리하였고, 250  $\mu$ l의 PBS로 재현탁시켰다. 세포를 1.5%의 최종 농도로 250  $\mu$ l의 3% 파라포름알데히드로 고정하였다. 샘플을 FACSDiva 소프트웨어(BD Biosciences, UK)를 이용하여 LSRFortessa(BD Biosciences, UK)로 분석하였다. 데이터를 FlowJo Version 10(FlowJo LLC, USA)을 이용하여 분석하였고, SARS-CoV-2 또는 SARS-CoV-2 ORF4a 레플리콘 saRNA 없이 상기와 동일한 방법론을 이용하여 모의(mock) 형질감염된 살아있고 단일한 것으로 게이트된 세포에 대해 세팅된 음성/양성 컷-오프(cut-off)를 이용하여 세포의 양성 개체군의 중앙 형광 강도(MFI)를 측정하였다.

[0368] 통계적 분석

[0369] 그래프 및 통계는 GraphPad Prism, 버전 8로 준비하였다. 통계적 차이는 2-원(two-way) ANOVA 또는 다중 비교를 위해 조정된 크루스칼-윌리스 테스트를 이용해 분석하였고,  $p < 0.05$ 를 사용하여 유의성을 시사하였다.

[0370] 결과 및 논의

[0371] RNA 레플리콘은 백신 및 치료제를 위한 관심있는 유전자의 운반 및 발현을 위한 잠재적인 도구가 될 것으로 상정되었다. 그러나, 이중 가닥 RNA(dsRNA)는 단백질 번역을 억제하는 신호전달 캐스케이드를 촉발하는 선천적 감지 메커니즘에 의해 세포내에서 검출된다. 그 결과, 레플리콘에서 암호화되는 관심있는 유전자의 발현이 현저하게 손상되고, 따라서 RNA 레플리콘의 치료적 잠재력이 제한된다.

[0372] 본 발명자들은 saRNA의 선천적 인식을 약화시키기 위하여 선천적 억제 단백질을 암호화하는 RNA 레플리콘을 개발함으로써 상기 문제를 극복하기 시작하였다. saRNA를 이용한 인터페론 반응을 약화시키기 위한 기존에 공개된 유일한 접근법은 백시니아 바이러스, E3, K3 및 B18 유래의 인터페론 억제 단백질을 사용하였다. 그러나, 상기 연구에서 인터페론 억제 단백질은 saRNA와 조합된 별도의 mRNA 분자로 운반 및 제형화되었다. 이것은 saRNA 및 mRNA 모두의 제조를 필요로 하고, 동일한 세포 내로의 공동-운반을 보장하고 단백질 발현에 있어서 임의의 관찰 가능한 향상을 제공하기 위하여 레플리콘 RNA보다 3-6배 더 많은 백시니아 mRNA의 사용을 필요로 한다. 아울러, mRNA 및 saRNA에 대한 발현의 동역학이 상이하여서, mRNA로부터 발현된 IIP의 임의의 유익한 효과는 수반하는 레플리콘과 비교하여 짧은 기간일 것이다.

[0373] 치료법에서 saRNA를 사용하는 현재의 어려움의 측면에서, 본 발명자들은 이들에 대한 면역 반응을 종래 기술의 방법과 비교하여 보다 효과적으로 제한하여 백신화 및 치료제에 있어서의 그 유용성을 증가시키는 신규한 saRNA를 디자인하였다.

[0374] 신규한 IIP로서 PIV-5 및 ORF4a의 용도

[0375] PIV-5 및 ORF4a는 각각 도입된 합성 saRNA의 번역을 감소시키는 제한 인자의 생산을 유발하는 인터페론 조절 인자 3 및 7(IRF3 및 IRF7)의 유도를 야기하는 MAV라 불리는 어댑터 분자를 통한 신호를 전달하는 세포질성 RNA 헬리카아제인 MDA-5를 차단하는 것으로 알려져 있다(도 1). 상기 2개의 IIP는 다양한 상이한 바이러스 유래의 10개 IIP의 초기 시험관내 스크리닝에서 확인되었다.

[0376] 1. HSV-2 Us1 - IRF-3와 IFN- $\beta$  프로모터의 결합을 저해함으로써 IFN-B 생산을 억제함[1]

[0377] 2. HSV-1 Us1(HSV-1 유래의 조정 인자)

[0378] 3. HSV-1 Us11 - RIG-I 신호전달을 방지함[2,3]

[0379] 4. OV20.0L - dsRNA에 결합하고 PKR & PACT를 모두 억제하여 RIG-I 신호전달을 차단함[4,5]

[0380] 5. BVDV Npro: IRF3 인산화 및 S100A9 신호전달을 차단함[6, 7]

[0381] 6. PIV5 V: MDA-5에 결합함으로써 MDA-5 및 IRF3를 차단함[8,9]

- [0382] 7. MERS-CoV M: TRAF3과 상호작용하고 TRAF3-TBK1 결합을 방해하여 IRF3 활성화의 감소를 유발함[10-12]
- [0383] 8. MERS-CoV ORF4a: dsRNA에 결합하고(긴 RNA에 대한 선호도를 가짐), PACT를 저해하여 MDA5 및 RIG-I, PKR 및 스트레스 반응을 촉발함[13,14].
- [0384] 9. 란가트 바이러스 NS5: IFNA1R을 하향 조절하고, JAK-STAT 신호전달을 손상시킴[15-16]
- [0385] 10. 인플루엔자 NS1: dsRNA에 결합하고, RIG-I 신호전달을 차단함[17]
- [0386] 상기 IIP는 GOI로서, 그리고 발현의 마커로서 사용되는 f루시페라아제와 함께 본 발명자들의 표준 VEEV saRNA 레플리콘 내에 통합되었다(도 1). 상기 구조체는 3가지 인간 세포주에서 평가되었다: 손상된 선천적 감지 경로를 갖는 HEK293T 세포, HeLa 세포 및 원발성 MRC5 배아 상피 세포(도 3a 및 도 3b). 모든 IIP 후보 saRNA는 HEK293T 세포에서 선천적 인식 없이 야생형 saRNA와 유사하게 복제하였다. (MERS-CoV M 및 인플루엔자 NS1를 제외한) 다양한 IIP는 HeLa 세포에서 발현을 향상시킬 수 있었지만, 가장 확연한 향상(3 log)은 PIV-V 및 ORF4a에서 보였다. 가장 중요하게는, 원발성 MRC5 세포에서의 평가는 PIV-V 및 ORF4a만이 루시페라아제 발현을 2 log만큼 향상시킬 수 있음을 시사하였다. 상기 데이터는 PIV-V 및 ORF4a가 원발성 인간 세포에서 GOI의 발현을 향상시키는 능력에 있어서 특별함을 제시한다. RNA 벡터에 포함시키기 위해 PIV-5 및 ORF4a를 확인한 것은 상기 실험 데이터에 기반하며, 평가된 다른 IIP가 유사한 메커니즘을 통해 작동하는 것으로 생각되는 것을 고려하면 그 활성은 예측불가하였다. 유전자 운반에서 그 용도를 추가로 뒷받침하기 위하여, 본 발명자들은 마우스에서 생체내 실험을 수행하였다. 여기서 본 발명자들은 BalBc 마우스보다 가장 강력한 선천적 감지 메커니즘 및 하류 인터페론 반응을 갖는 것으로 알려져 있고, 이에 따라 2가지 모델을 비교하는 것이 유용한 Black 6(BL6) 마우스를 이용하였다. 본 발명자들은 리포터 단백질로서 반딧불이 루시페라아제(fLuc)를 이용하여 BL/6 마우스에서 WT, PIV-5 및 ORF4a 레플리콘의 상대적 발현을 평가하였다(도 4).
- [0387] fLuc는 세포내에서 발현되기 때문에, 상기 관심있는 유전자의 발현은 기질인 D-루시페린과 함께 복강내(IP) 주사한 이후 루시페라아제 발현에 기반하여 시각화되고, 총 플럭스(p/s)로서 표현된다(방법 참조). 상기 생체내 실험은 PIV5 및 ORF4a는 BL/6 세포에서 14일까지 루시페라아제 발현의 기간을 증가시켰지만, WT 구조체는 이 시점에서 음성이었음을 실증하였다. 추가 연구에서, 본 발명자들은 (GOI로서) gLuc의 g루시페라아제의 발현에 대한 상기 2가지 RNA 레플리콘 구조체의 영향을 평가하였다(도 5). gLuc는 가용성 단백질로서 분비되며, 그 활성은 혈액에서 측정된다(방법 참조). 상기 생체내 실험은 PIV5 및 ORF4a가 BL/6 세포에서 14일까지 루시페라아제 발현의 기간을 증가시켰음을 실증하였다. gLuc의 발현은 14일에 PIV5 및 ORF4a RNA 레플리콘 모두의 경우에 현저하게 더 높았다(각각  $p=0.0244$  및  $0.00422$ ). 14일에 걸친 "곡선 하 면적"의 계산은 다음의 값을 시사하고, AUC는 다음과 같다: fLuc BALBc= 1950; PIV5 BALBc= 2742; ORF4a BALBc= 1596; fLuc BL6= 2012; PIV5 BL6= 4513; ORF4a BL6= 6972(총 플럭스(p/s)). 상기 생체내 실험이 뒷받침하고 있지만, 선천적 제한 인자 및 인터페론 자극 유전자의 특이도와 관련하여 인간 및 마우스 사이에 현저한 차이가 있고(>100임), 따라서 상기 데이터는 인간 세포주를 이용한 시험관내 관찰에 기반한 인간에서 예상되는 영향을 과소평가할 수 있음을 유의하는 것이 중요하다.
- [0388] 인터페론 억제 단백질은 시험관내에서 saRNA의 단백질 발현을 향상시킨다.
- [0389] 본 발명자들은 인터페론 억제 단백질의 라이브러리가 시험관내에서 반딧불이 루시페라아제(fLuc) 단백질 발현을 향상시키는지 여부를 결정하고자 하였다. 본 발명자들은 IRF-3, MDA5, RIG-I, 및 JAK/STAT(도 9b)을 포함하는 다양한 세포질성 인터페론 표적(표 1)과 함께 T2A 절단 부위로 fLuc로부터 분리된 IIP를 갖는 saRNA VEEV 레플리콘의 라이브러리를 제조하였다(도 9a).

## 표 1

[0390] 인터페론을 억제하는 VEEV 레플리콘 및 관련 IFN 표적

| 구조체         | 경로 표적                                                |
|-------------|------------------------------------------------------|
| HSV-2 Us1   | IRF-3과 IFN-B 프로모터의 결합을 저해함으로써 IFN-B 생산을 억제함          |
| HSV-1 Us1   | 대조군                                                  |
| HSV-1Us11   | PACT에 결합하고, MDA5 및 RIG-I 신호전달을 차단함                   |
| Orf OV20.0L | dsRNA에 결합하고, PKR 및 PACT 모두를 억제하여, RIG-I 신호전달을 차단함    |
| BVDV Npro   | IRF3 인산화를 차단함                                        |
| PIV-5 V     | MDA-5에 결합함으로써 MDA-5 및 IRF3을 차단함                      |
| MERS-CoV M  | TRAF3과 상호작용하고, TRAF3-TBK1 결합을 방해하여 IRF3 활성화의 감소를 유발함 |

|                |                                                        |
|----------------|--------------------------------------------------------|
| MERS-CoV ORF4a | 긴 RNA에 대한 선호도로 dsRNA에 결합하고, MDA5 및 RIG-I의 PACT 촉발을 억제함 |
| 란가트 NS5        | IFN $\alpha$ 1R을 하향 조절하고, JAK/STAT 신호전달을 손상시킴          |
| 인플루엔자 NS1      | dsRNA에 결합하고, RIG-I 신호전달을 차단함                           |

[0391]

이후, 본 발명자들은 상대적으로 높은 단백질 발현을 생성하지만 그 생체감소(bioreducible) 본성으로 인해 상대적으로 면역 침묵인 것으로 기존에 특징분석 되어 있는 폴리머성 운반 시스템인 pABOL(도 9c, 보충적 도 1)을 이용하여 saRNA를 HEK293T.17, HeLa 및 MRC5 세포 내로 형질감염시켰다(2). 본 발명자들은 IFN 경로의 완전함에 있어서의 변동성을 위해 상기 3종의 세포주를 선택하였다; HEK293T.17 세포는 내인성 RIG-I 및 MDA5 발현이 없기 때문에 완전한 경로를 갖지 않고(37), 따라서 상기 경로에 영향을 미치는 단백질에 대해 덜 민감할 것인 반면, HeLa 및 MRC5는 보다 차별적이다(38, 39). 본 발명자들은 IIP 레플리콘 모두 HEK293T.17 세포에서 단백질 발현을 향상시키지 않았지만(도 1c), 흥미롭게도 란가트 및 인플루엔자 IIP는 모두 각각  $p=0.0097$  및  $0.0061$ 으로 단백질 발현을 0.06배 현저하게 감소시켰음을 관찰하였다. HeLa 세포에서, 많은 IIP는 단백질 발현을 향상시켰다; HSV-2, HSV-1\_1, HSV-1\_2, Orf 및 BVDV는 fLuc 발현을 20-150배 범위로 증가시켰다. 그러나, PIV-5 V 및 MERS-CoV ORF4a 단백질은 각각 796 및 893배로 가장 많이 단백질 발현을 향상시켰지만, PIV-5 군만이 통계적으로 유의미하였고( $p=0.0272$ ), ORF4a 군은 그렇지 않았다( $p=0.0689$ ). MRC5 세포에서, 본 발명자들은 PIV-5 V 및 MERS-CoV ORF4a 단백질로부터 각각  $p=0.0485$  및  $0.025$ 로 72 및 109배 더 큰 fLuc 발현의 가장 큰 향상을 유사하게 관찰하였다. 모든 세포 타입 및 각각의 구조체에 대해 2가지 별도로 만들어진 RNA의 배치로부터의 발현 레벨 사이에 양호한 일치가 있었다(도 14).

[0392]

본 발명자들은 PIV-5 V 및 MERS-CoV ORF4a 단백질에 대한 2개의 돌연변이가 어떻게 마우스(MEF), 토끼(RK13), 비-인간 영장류(LLC) 및 인간 세포(MRC5)에서 단백질 발현에 영향을 미치는지를 추가로 조사하였다(도 15a 내지 도 15d). PIV-5 V에서의 R172A 돌연변이는 MDA5를 차단하는 능력을 폐지하지만 STAT는 그렇지 않고(40), MERS-CoV ORF4a에서 K63A/K67A 돌연변이는 dsRNA에 대한 결합을 차단한다(41). 본 발명자들은 PIV-5 V 및 MERS-CoV ORF4a 단백질이 MEF 또는 RK13 세포에서 단백질 발현을 향상시키지 않음을 관찰하였다. 상기 MERS-CoV ORF4a 단백질은 LLC 및 MRC5 세포에서 단백질 발현을 향상시켰고(도 15c, 도 15d), 상기 K63A/K67A 돌연변이는 단백질 발현을 매우 감소시켰다. 상기 PIV-5 V 단백질은 MRC5 세포에서는 단백질 발현을 향상시켰지만 LLC 세포에서는 그렇지 않았고, 상기 R172A 돌연변이는 MRC5 세포에서 단백질 발현을 감소시켰다. 종합적으로, 상기 데이터는 PIV-5 V 및 MERS-CoV ORF4a 단백질이 인터페론-적격(competent) 인간 세포에서 단백질 발현을 향상시키고, K63A/K67A 및 R172A 치환을 갖는 단백질을 돌연변이시켜 saRNA 발현을 약화시킴을 시사한다.

[0393]

MERS-CoV ORF4a 단백질은 생체내에서 부분적으로 용량 비선형성의 증가를 약화시킨다.

[0394]

PIV-5 V 및 MERS-CoV ORF4a 단백질로부터 시험관내 단백질 발현이 향상되었으므로, 본 발명자들은 상기 구조체가 생체내에서 단백질 발현을 향상시키고 saRNA의 용량 증가의 비선형성을 약화시킬지 여부를 결정하고자 하였다. 본 발명자들은 세포내 단백질인 반딧불이 루시페라아제, 및 분비된 단백질인 가우시아 루시페라아제 모두를 암호화하는 saRNA를 생체내에서 테스트하였다(표 2).

## 표 2

[0395]

14일의 과정에 걸쳐  $n=5$ 로 BALB/c 및 C57BL/6J에서 총 루시페라아제 발현의 곡선 하 면적(AUC)

|      | BALB/c              |                     |                     | C57BL/6             |                    |                     |
|------|---------------------|---------------------|---------------------|---------------------|--------------------|---------------------|
|      | WT                  | +PIV-5              | +MERS-CoV_2         | WT                  | +PIV-5             | +MERS-CoV_2         |
| fLuc | 434210 $\pm$ 173923 | 235231 $\pm$ 103702 | 301294 $\pm$ 170309 | 263783 $\pm$ 191231 | 246182 $\pm$ 87859 | 453411 $\pm$ 331433 |
| gLuc | 1950 $\pm$ 1270     | 2742 $\pm$ 493      | 1596 $\pm$ 915      | 2012 $\pm$ 1373     | 4513 $\pm$ 1651    | 6972 $\pm$ 2789     |

[0396]

본 발명자들은 인터페론 생성 능력에서의 차이에 대해 BALB/c 및 C57BL/6 마우스 모두에서 상기 구조체를 테스트하기 위해 선택하였다: 시험관내에서 HEK293T.17 및 HeLa/MRC5 세포의 차이와 유사하게, BALB/c는 IFN을 불량하게 생산하지만, C57BL/6 마우스는 IFN- $\alpha/\beta$  및 IFN- $\gamma$ 를 높게 생산하는 것으로 기존에 알려져 있다(42). 본 발명자들은 상기 PIV-5 V 및 MERS-CoV ORF4a 단백질을 통합해도 BALB/c 마우스에서 fLuc 또는 gLuc의 단백질 발현을 향상시키지 않았음을 관찰하였다(표 2, 도 16). 본 발명자들은, 비록 그 차이가 통계적으로 유의미하지는 않았지만, MERS-CoV ORF4a 단백질을 갖는 C57BL/6 마우스에서 fLuc, 및 PIV-5 V 및 MERS-CoV ORF4a 단백질을 갖는 마우스에서 gLuc의 단백질 발현의 총 곡선 하 면적(AUC)이 약간 향상되었음을 관찰하였다.

- [0397] 본 발명자들은 이전에 saRNA의 용량을 증가시키면 결국 단백질 발현 레벨의 감소로 귀결됨을 관찰하였고, 따라서 MERS-CoV ORF4a 단백질이 생체내에서 saRNA의 비선형 용량 의존성을 약화시킬 수 있는지 여부를 특징분석하고자 하였다. 본 발명자들은 0.2, 2 및 20  $\mu\text{g}$  용량의 야생형 fLuc 및 fLuc + MERS-CoV ORF4a 레플리콘을 테스트하였고, 근육내(IM) 주사 7 및 10일 후에 단백질 발현을 정량하였다(도 10). 본 발명자들은 7일 후 0.2  $\mu\text{g}$ 의 용량에서 양쪽 구조체가 유사한 단백질 발현( $\sim 5000$  p/s)을 갖고, 용량이 2  $\mu\text{g}$ 으로 증가되었을 때, 양쪽 모두의 경우에 단백질 발현이 증가하였으며(WT의 경우  $\sim 50,000$  p/s 및 MERS-CoV ORF4a 구조체의 경우  $\sim 200,000$  p/s 까지), MERS-CoV ORF4a 단백질을 통합한 경우  $p=0.0029$ 로 단백질 발현이 4배 향상됨을 관찰하였다. 흥미롭게도, 양쪽 구조체 모두 7일 후 20  $\mu\text{g}$ 의 용량에서 낮은 단백질 발현을 나타내었지만, WT는  $p<0.0001$ 으로 MERS-CoV 구조체보다 18배 더 낮았다. 10일 후, 단백질 발현 레벨은 2  $\mu\text{g}$  용량의 경우와 동일하였고, 0.2 및 20  $\mu\text{g}$  용량에서 발현이 관찰되지 않았다. 임의의 특정 이론에 구애되지 않음으로써, 상기 데이터는 MERS-CoV ORF4a 단백질이 생체내에서 saRNA의 비선형 용량 의존성을 부분적으로 구제할 수 있음을 시사한다.
- [0398] *록솔리티닙은 생체내에서 saRNA의 단백질 발현을 향상시킨다.*
- [0399] 하류의 인터페론 반응에서 JAK/STAT 경로의 역할을 고려하여, 본 발명자들은 saRNA, MERS-CoV ORF4a 인터페론 억제 단백질 및 JAK1 및 JAK2 단백질 키나아제의 강력하고 선택적인 억제제인 록솔리티닙(36)의 조합이 생체내에서 어떻게 단백질 발현에 영향을 미치는지를 특징분석하고자 하였다(도 3). 본 발명자들은 5  $\mu\text{g}$ 의 fLuc  $\pm$  MERS-CoV ORF4a를 암호화하는 saRNA를 100  $\mu\text{g}$ 의 록솔리티닙과의 공동-제형과 함께 또는 없이 마우스에 IM 주사하였고, 주사 4, 7, 10 및 14일 후에 단백질 발현을 정량하였다. 4일 후(도 11a), 록솔리티닙을 함유하는 제형은 모두 WT 또는 MERS-CoV ORF4a 구조체( $\sim 5 \times 10^5$  p/s)와 비교하여 약간 더 높은 단백질 발현( $\sim 10^6$  p/s)을 가졌지만, 통계적으로 유의미하지는 않았다. 그러나, 7일 후, 록솔리티닙을 함유하는 제형은 모두 saRNA-단독 병렬 군과 비교하여 각각  $p=0.0347$  및  $0.0447$ 으로 더 높은 단백질 발현을 가졌다. 10일까지, 상기 군들은 여전히 약간 상승되어 있었지만, 그 차이는 더 이상 통계적으로 유의미하지 않았다. 14일 후, 록솔리티닙이 없는 saRNA 군에서는 단백질 발현이 관찰되지 않았고, 록솔리티닙 군의 경우에 단지 일부 양성 샘플이 있었다. 임의의 특정 이론에 구애되지 않음으로써, 상기 데이터는 록솔리티닙이 saRNA 단백질 발현에서 고도한 증가를 가능하게 하지만, 조합시에 MERS-CoV ORF4a 단백질 및 록솔리티닙 사이에 부가 효과는 없음을 시사한다.
- [0400] *PIV-5 V 및 MERS-CoV ORF4a 단백질은 인간 피부 외식편에서 생체외에서 용량 비선형성의 증가를 약화시킨다.*
- [0401] 본 발명자들은 IIP가 시험관내에서 세포 타입의 종에 따라 단백질 발현에 있어서 차이를 나타냄을 관찰하였으므로, 본 발명자들은 보다 임상적으로 관련있는 인간 피부 외식편 모델에서 saRNA IIP 구조체를 테스트하고자 하였다. 본 발명자들은 PIV-5 V 및 MERS-CoV ORF4a 단백질(도면에서 MERS-CoV\_2로 나타냄)뿐만 아니라, 록솔리티닙과의 공동-제형이 통합된 고유(resident) 인간 피부 세포에서 단백질 발현의 정량(eGFP<sup>+</sup> 세포의 %) 및 정성(세포 당 중앙 eGFP 형광 강도) 모두를 특징분석하였다(도 12). 본 발명자들은 0.2, 2 및 20  $\mu\text{g}$  용량의 eGFP saRNA를 PIV-5 V 및 MERS-CoV ORF4a 단백질과 함께 테스트하였고(도 12a, 도 12b), WT 구조체의 용량을 0.2에서 2  $\mu\text{g}$ 로 증가시키면 eGFP<sup>+</sup> 세포의 백분율이 10%에서 18%로 증가하는 것으로 귀결되지만, 용량이 20  $\mu\text{g}$ 으로 증가되었을 때, eGFP<sup>+</sup> 세포의 백분율은  $\sim 5\%$ 로 급락하였음을 관찰하였다. 그러나, PIV-5 및 MERS-CoV ORF4a 구조체의 경우, saRNA의 용량 증가와 함께 선형으로 용량이 증가되었다. 0.2  $\mu\text{g}$  용량은 유사하게 상기 구조체 모두에 대해  $\sim 12\%$ 의 eGFP<sup>+</sup>로 귀결되었고, 2  $\mu\text{g}$ 에서 15% 및 20  $\mu\text{g}$ 에서 25%로 추가로 증가되었으며, 이 시점에서 상기 PIV-5 및 MERS-CoV ORF4a 구조체는 모두 양쪽 모두에 대해  $p<0.0001$ 로 통계적으로 유의미하게 eGFP<sup>+</sup> 세포의 백분율이 더 높았다. 흥미롭게도, PIV-5 또는 MERS-CoV ORF4a 단백질의 용량 또는 통합 모두 eGFP MFI에 영향을 미치지 않았으며, 모든 샘플에 대해  $\sim 350$ 이었다(도 12b).
- [0402] 본 발명자들은 t-분포 확률적 임베딩을 이용하여 어느 세포가 saRNA를 발현하는지와, 오버레이 정의된 단백질 및 표현형 게이팅으로 세포를 비지도 클러스터링함으로써 시각화하도록 하는 유세포 분석 데이터를 위한 주된 성분 분석의 타입을 추가로 특징분석하였다(도 17-25)(43). 본 발명자들은 가장 높은 용량의 saRNA(20  $\mu\text{g}$ )에서 상기 PIV-5 V 및 MERS-CoV ORF4a 단백질은 고유 상피 세포 및 섬유아세포와 대조적으로 T 세포, 수지상 세포, 단핵구, B 세포, 랑게르한스 세포, 백혈구 및 NK 세포를 포함하는 면역 세포에서 단백질을 발현할 수 있음을 관찰하였다.
- [0403] 다음으로, 본 발명자들은 0-100  $\mu\text{g}$  범위 용량의 록솔리티닙을 통합하면 어떻게 인간 피부 고유 세포에서 saRNA 발현에 영향을 미치는지를 테스트하였다. 본 발명자들은 록솔리티닙과 saRNA의 공동-제형은 eGFP<sup>+</sup> 세포의 백분

율에 어떠한 영향도 갖지 않았지만(도 4c), 록솔리티닙의 용량을 증가시키면 eGFP<sup>+</sup> 세포의 백분율이 ~8%에서 ~5%로 실제로 감소되는 약간의 경향이 있음을 관찰하였다. 그러나, 본 발명자들은 eGFP 발현의 세포 당 품질에 대한 고도한 효과는 관찰하지 않았다(도 12d); 록솔리티닙의 용량을 증가시키면 ~100으로부터 10  $\mu$ g 용량의 록솔리티닙에서 ~2000으로 eGFP MFI를 증가시켰지만, MFI는 100  $\mu$ g 용량의 록솔리티닙에서는 ~1000으로 감소하였다. saRNA PIV-5 및 MERS-CoV ORF4a 단백질을 발현하는 세포와 유사하게, 본 발명자들은 록솔리티닙이 상피 세포 및 섬유아세포와 대조적으로 면역 세포에서 단백질 발현을 향상시켰고, T 세포, 랑게르한스 세포, 백혈구 및 NK 세포에서의 섭취를 특이적으로 증가시켰음을 발견하였다(도 26a, 도 26b).

[0404] 함께 고려하면, 임의의 특정 이론에 구애되길 원하지 않으면서, 상기 데이터는 IIP 레플리콘은 saRNA를 발현하는 세포의 백분율을 증가시킴으로써 면역 세포에서 발현을 향상시키지만, 록솔리티닙은 세포 당 기준으로 단백질 발현을 향상시킴을 보여준다.

[0405] *MERS-CoV ORF4a* 단백질은 토끼에서 생체내 RABV 당단백질의 면역원성을 향상시킨다.

[0406] 핵산 제형의 단백질 발현이 항상 면역원성의 직접적인 예측자가 되는 것은 아니기 때문에(2), 본 발명자들은 본 발명의 saRNA 구조체에서 (선천적 억제제 단백질 또는 IIP로 작용하는) MERS-CoV ORF4a 단백질과 조합될 때, 치료 생체분자로서 작용하는 모델 단백질, 즉 진뱅크 ID 번호: NP\_056796.1로 표시되는 광견병 당단백질(RABV)의 면역원성을 특징분석하고자 하였다. 상기 RABV 단백질은 부가적으로 세포의 p75NTR 표면 수용체에 대한 결합을 방지하는 아미노산 치환인 F318V 변형과 함께 사용되었다. 본 발명자들은 토끼를 20  $\mu$ g의 saRNA의 1차 용량으로 주사하고, 4주 후에 부스트한 후, 0, 4, 및 6주 후에 그 혈액에서 RABV-특이적 IgG 항체를 샘플링하였다(도 13a). 본 발명자들은 야생형 및 MERS-CoV ORF4a 구조체 모두에 대한 토끼는 모두 단일 주사 후에 혈청전환되었음을 관찰하였다. 4주 후, MERS-CoV 군에 대한 IgG 역가는 야생형( $\sim 5 \times 10^3$  ng/ml)과 비교하여 약간 더 높았지만( $\sim 10^4$  ng/ml), 이것은 통계적으로 유의미하지는 않았다. 그러나, 6주 후, MERS-CoV ORF4a 군에서 동물에 대한 항체 역가는  $p=0.0061$ 로 WT( $\sim 10^4$  ng/ml)와 비교하여 현저하게 더 높았다( $\sim 10^5$  ng/ml). 상기 RABV 위형화된 중화는 항체의 경향을 반영하였다(도 13b). 4주 후, WT 군은  $\sim 10^3$ 의 평균 IC<sub>50</sub>을 가졌지만, MERS-CoV 군은  $\sim 10^4$ 의 IC<sub>50</sub>을 가졌다. 6주 후, MERS-CoV 군은  $\sim 10^5$ 의 훨씬 더 높은 IC<sub>50</sub>을 가졌지만, WT 군은  $\sim 10^3$ 에서 안정화되었다. 본 발명자들은 또한 마우스 및 래트에서 WT RABV 및 RABV-MERS-CoV ORF4a saRNA의 면역원성을 비교하였지만(도 27), 상기 중 모두에서 항체 역가 또는 중화 IC<sub>50</sub> 사이에 임의의 차이를 관찰할 수 없었다. 임의의 특정 이론에 구애되길 원하지 않으면서, 상기 데이터는 MERS-CoV ORF4a 단백질이 토끼에서 saRNA에 의해 암호화되는 RABV 당단백질의 면역원성을 향상시킴을 시사한다.

[0407] *MERS-CoV ORF4a* 단백질은 시험관내에서 SARS-CoV-2 당단백질의 발현을 향상시킨다.

[0408] 본 발명자들은 또한 본 발명의 saRNA 구조체에서 MERS-CoV ORF4a 단백질(즉, 선천적 억제제 단백질 또는 IIP)과 조합될 때 다른 모델 단백질(즉, 치료 생체분자)인 진뱅크 ID 번호: QHD43416.1로 표시되는 SARS-CoV-2 당단백질의 면역원성을 특징분석하였다.

[0409] HeLa 세포주 내로 형질감염된 후 24시간에, SARS-CoV-2(IIP 없음) 및 co MERS-CoV ORF4a(IIP)와 조합된 SARS-CoV-2 사이의 표면 발현의 레벨을 비교하였다. 도 29를 참조하면, 모의 형질감염된 살아있고 단일한 것으로 게이트된 세포에 대해 세팅된 음성/양성 컷-오프를 이용하여 세포의 양성 개체군의 중앙 형광 강도(MFI)를 측정하였다. 놀랍게도, SARS-CoV-2 당단백질만을 암호화하는 saRNA(즉, IIP 없음)와 비교할 때, MERS-CoV ORF4a 단백질은 SARS-CoV-2 당단백질의 발현 레벨이 세포 당 거의 2배였다.

[0410] 논의

[0411] 본 발명자들은 마우스, 토끼, 비-인간 영장류 및 인간 세포에서 시험관내, 인간 피부 외식편에서 생체외, 및 마우스에서 생체내 단백질 발현에 대한 시스-암호화된 인터페론 억제 단백질들을 갖는 자가-중폭 RNA의 라이브러리 뿐만 아니라 마우스, 래트 및 토끼에서 면역원성을 스크리닝하였다. 본 발명자들은 PIV-5 V 및 MERS-CoV ORF4a 단백질이 IFN-적격 HeLa 및 MRC5 세포에서 시험관내 단백질 발현을 100-500배 향상시켰음을 관찰하였다. 본 발명자들은 MERS-CoV ORF4a 단백질은 생체내에서 용량 비선형성을 부분적으로 약화시키지만, 록솔리티닙은 생체내에서 saRNA의 단백질 발현을 향상시키고, IIP는 그렇지 않음을 발견하였다. 상기 PIV-5 V 및 MERS-CoV ORF4a 단백질은 모두 saRNA를 발현하는 인간 피부 외식편에서 고유 세포의 백분율을 향상시키고 saRNA의 용량 비선형성

을 완전히 구제하지만, 록솔리티닙은 세포 기준 당 단백질 발현을 증가시킴을 발견하였다. 마지막으로, 본 발명자들은 상기 MERS-CoV ORF4a가 토끼에서 RABV-특이적 IgG 역가 및 중화 IC<sub>50</sub>을 ~10배까지 증가시켰지만, 마우스 및 래트에서는 그렇지 않음을 관찰하였다.

[0412] 상기 실험에서 특징분석된 단백질 디자인, 세포 및 돌연변이는 PIV-5 V 및 MERS-CoV ORF4a 단백질이 단백질 발현을 증가시키는 메커니즘으로의 통찰력을 제공한다. 상기 PIV-5 V 단백질은 MDA-5에 결합함으로써 MDA-5 및 IRF3을 차단하지만(26, 27), 상기 MERS-CoV ORF4a 단백질은 dsRNA에 결합하고 MDA-5 및 RIG-I를 촉발하는 PACT를 저해한다(29-31). 본 발명자들의 시험관내 스크리닝 후, 본 발명자들은 MERS-CoV ORF4a 레플리콘으로 앞으로 나아가기로 선택하였는데, 그 이유는 생체내에서 모든 10개 후보자를 스크리닝하는 것이 실험 가능하지 않았고, PIV-5 V 단백질은 종들 사이에 보존되지 않았지만(예컨대, 마우스에 적응하기 위해서는 N100D 돌연변이가 필요함(44)), ORF4a 단백질은 종들 사이에 매우 보존되어 있기 때문이다(29-31). 본 발명자들은 PIV-5 V 단백질에 대한 R172A 돌연변이는, MDA-5에 대한 결합을 차단하는 능력은 폐지하지만 STAT에 대해서는 그렇지 않으며(45), MRC5 세포에서 단백질 발현을 약간 억제함을 관찰하였고(도 15d), 따라서 MDA5에 대한 결합은 부분적으로 단백질 발현을 향상시키는데 책임이 있음을 시사한다. 유사하게, MERS-CoV ORF4a 단백질에 대한 K63A/K67A 돌연변이는 dsRNA에 결합하는 능력을 제한하고(46, 47), 비-인간 영장류 및 인간 세포 모두에서 단백질 발현을 감소시키는 것으로 관찰되었다(도 15c, 도 15d). 다양한 IIP가 PIV-5 V 및 MERS-CoV ORF4a와 유사한 메커니즘에 의해 인터페론을 억제하지만, 작용 메커니즘은 필수적으로 단백질 발현의 향상을 예측하는 것으로 관찰되지는 않았다.

[0413] 본 발명자들은 saRNA 제형의 단백질 발현이 필수적으로 면역원성을 예측하는 것은 아님을 이전에 관찰하였고(2), 이에 따라 MERS-CoV ORF4a 단백질이 어떻게 마우스, 래트 및 토끼에서 광견병 당단백질의 면역원성에 영향을 미치는지도 특징분석하였다. 본 발명자들은 토끼에서 RABV 및 MERS-CoV ORF4a를 암호화하는 saRNA가 항체 역가 및 중화 IC<sub>50</sub> 모두를 증가시킴을 관찰하였지만(도 13a, 도 13b), 마우스 또는 래트에서 면역원성의 증가는 관찰하지 않았다. 인간 반응을 완벽하게 예측하는 전임상 동물 모델은 없지만, 토끼는 면역학적으로 마우스 또는 래트보다 인간과 보다 유사한 것으로 간주된다(48-50). 아울러, 본 발명자들은 무린(murine) 세포에서 PIV-5 V 또는 MERS-CoV ORF4a 단백질에 의한 단백질 발현의 임의의 향상을 관찰하지 않았고(도 15a), 따라서 면역원성의 향상 부재도 예측불가가 아니다. 본 발명자들은 전임상 동물 모델에서의 특징을 세포가 자연 조직구조 내에 있고 내재하는 인간 IFN 반응을 갖고 있는 인간 외식편 모델과 쌍을 이루었다. 그 지식 안에서, 본 발명자들은 IIP는 saRNA를 발현하는 세포의 백분율을 향상시키지만, 록솔리티닙은 세포 당 발현을 향상시키는 것을 처음으로 관찰한다. 상기 유망한 결과를 고려할 때, 본 발명자들은 상기 MERS-CoV ORF4a 단백질이 인간에서 saRNA 백신의 면역원성을 향상시킬 수 있고, 또한 saRNA를 단백질 교체 치료법에 적용하는데 유용할 수 있음을 상정한다(51, 52).

[0414] 상기 실험은 IIP가 saRNA 벡터 내로 직접 암호화되고, 비선형 용량 의존성을 효과적으로 약화시키고, 면역원성을 향상시킬 수 있다는 개념 증명을 제공한다. 기계론적 연구에 의해 시사된 것과 같이, 인터페론 경로의 상이한 측면이 표적화되고, saRNA 발현을 증가시킬 수 있으며, 따라서 IIP 및 다른 IFN 억제 전략, 예컨대 록솔리티닙의 조합을 탐색하는 동기부여를 제공한다.

#### [0415] 재정 지원에 대한 설명

[0416] 본 출원을 허용하는 프로젝트는 마리 스클로돕스카-큐리(Marie Skłodowska-Curie) 보조금 협약 번호 794059 하의 유럽 연합 호라이즌 2020 연구 및 혁신 프로그램으로부터 기금을 받았다.

#### [0417] 참고문헌

- [0418] 1. S. Perri *et al.*, An alphavirus replicon particle chimera derived from venezuelan equine encephalitis and sindbis viruses is a potent gene-based vaccine delivery vector. *J Virol* **77**, 10394-10403 (2003).
- [0419] 2. A. K. Blakney *et al.*, Big is Beautiful: Enhanced saRNA Delivery and Immunogenicity by a Higher Molecular Weight, Bioreducible, Cationic Polymer. *ACS Nano* 10.1021/acsnano.0c00326 (2020).
- [0420] 3. L. A. Brito *et al.*, A cationic nanoemulsion for the delivery of next-generation RNA vaccines. *Molecular therapy : the journal of the American Society of Gene Therapy* **22**, 2118-2129 (2014).
- [0421] 4. A. K. Blakney, P. F. McKay, B. I. Yus, Y. Aldon, R. J. Shattock, Inside out: optimization of lipid

- 60 –

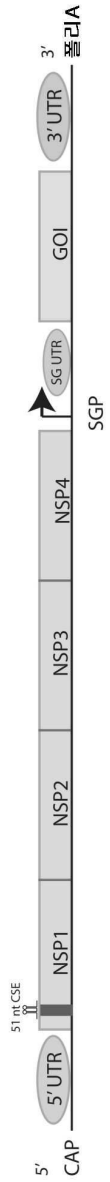
inhibits an interferon-inducible, double-stranded RNA-dependent kinase. *Immunology* **93**, 335-340 (1998).

- [0441] 24. N. Horscroft *et al.*, Establishment of a subgenomic replicon for bovine viral diarrhea virus in Huh-7 cells and modulation of interferon-regulated factor 3-mediated antiviral response. *J Virol* **79**, 2788-2796 (2005).
- [0442] 25. M. F. Darweesh, M. K. S. Rajput, L. J. Braun, J. S. Rohila, C. C. L. Chase, BVDV Npro protein mediates the BVDV induced immunosuppression through interaction with cellular S100A9 protein. *Microb Pathog* **121**, 341-349 (2018).
- [0443] 26. K. R. Rodriguez, C. M. Horvath, Paramyxovirus V protein interaction with the antiviral sensor LGP2 disrupts MDA5 signaling enhancement but is not relevant to LGP2-mediated RLR signaling inhibition. *J Virol* **88**, 8180-8188 (2014).
- [0444] 27. R. Mandhana, L. K. Qian, C. M. Horvath, Constitutively Active MDA5 Proteins Are Inhibited by Paramyxovirus V Proteins. *J Interferon Cytokine Res* **38**, 319-332 (2018).
- [0445] 28. P. Y. Lui *et al.*, Middle East respiratory syndrome coronavirus M protein suppresses type I interferon expression through the inhibition of TBK1-dependent phosphorylation of IRF3. *Emerg Microbes Infect* **5**, e39 (2016).
- [0446] 29. Y. Yang *et al.*, The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell* **4**, 951-961 (2013).
- [0447] 30. S. Shokri, S. Mahmoudvand, R. Taherkhani, F. Farshadpour, Modulation of the immune response by Middle East respiratory syndrome coronavirus. *J Cell Physiol* **234**, 2143-2151 (2019).
- [0448] 31. M. Batool, M. Shah, M. C. Patra, D. Yesudhas, S. Choi, Structural insights into the Middle East respiratory syndrome coronavirus 4a protein and its dsRNA binding mechanism. *Sci Rep* **7**, 11362 (2017).
- [0449] 32. C. E. Comar *et al.*, Antagonism of dsRNA-Induced Innate Immune Pathways by NS4a and NS4b Accessory Proteins during MERS Coronavirus Infection. *mbio* **10** (2019).
- [0450] 33. S. M. Best, The Many Faces of the Flavivirus NS5 Protein in Antagonism of Type I Interferon Signaling. *J Virol* **91** (2017).
- [0451] 34. K. Werme, M. Wigerius, M. Johansson, Tick-borne encephalitis virus NS5 associates with membrane protein scribble and impairs interferon-stimulated JAK-STAT signalling. *Cell Microbiol* **10**, 696-712 (2008).
- [0452] 35. R. M. Krug, Functions of the influenza A virus NS1 protein in antiviral defense. *Curr Opin Virol* **12**, 1-6 (2015).
- [0453] 36. E. M. Elli, C. Barate, F. Mendicino, F. Palandri, G. A. Palumbo, Mechanisms Underlying the Anti-inflammatory and Immunosuppressive Activity of Ruxolitinib. *Front Oncol* **9**, 1186-1186 (2019).
- [0454] 37. M. E. Fitzgerald *et al.*, Selective RNA targeting and regulated signaling by RIG-I is controlled by coordination of RNA and ATP binding. *Nucleic Acids Research* **45**, 1442-1454 (2016).
- [0455] 38. K. Cantell, Production and action of interferon in HeLa cells. *Archiv fur die gesamte Virusforschung* **10**, 510-521 (1961).
- [0456] 39. E. Meurs, A. G. Hovanessian, L. Montagnier, Interferon-mediated Antiviral State in Human MRC5 Cells in the Absence of Detectable Levels of 2-5A Synthetase and Protein Kinase. *Journal of Interferon Research* **1**, 219-234 (1981).
- [0457] 40. A. Ramachandran, C. M. Horvath, Dissociation of paramyxovirus interferon evasion activities: universal and virus-specific requirements for conserved V protein amino acids in MDA5 interference. *Journal of virology* **84**, 11152-11163 (2010).

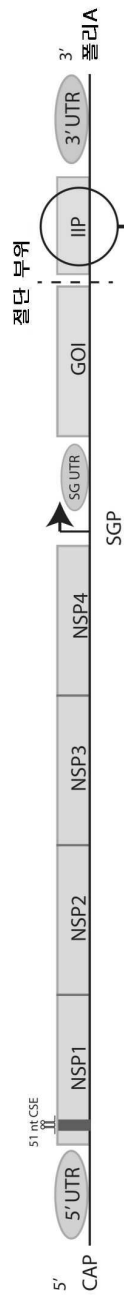
- [0458] 41. K.-L. Siu *et al.*, Middle East Respiratory Syndrome Coronavirus 4a Protein Is a Double-Stranded RNA-Binding Protein That Suppresses PACT-Induced Activation of RIG-I and MDA5 in the Innate Antiviral Response. *Journal of Virology* **88**, 4866 (2014).
- [0459] 42. T. Shirahata, A. Mori, H. Ishikawa, H. Goto, Strain Differences of Interferon-Generating Capacity and Resistance in Toxoplasma-Infected Mice. *Microbiology and Immunology* **30**, 1307-1316 (1986).
- [0460] 43. L. Van Der Maaton, G. Hinton, Visualizing data using t-SNE. . *J. Mach. Learn. Res.* **9**, 2579-2625 (2008).
- [0461] 44. T. A. Kraus, L. Garza, C. M. Horvath, Enabled interferon signaling evasion in an immune-competent transgenic mouse model of parainfluenza virus 5 infection. *Virology* **371**, 196-205 (2008).
- [0462] 45. A. Ramachandran, C. M. Horvath, Dissociation of Paramyxovirus Interferon Evasion Activities: Universal and Virus-Specific Requirements for Conserved V Protein Amino Acids in MDA5 Interference. *Journal of Virology* **84**, 11152 (2010).
- [0463] 46. H. H. Rabouw *et al.*, Middle East Respiratory Coronavirus Accessory Protein 4a Inhibits PKR-Mediated Antiviral Stress Responses. *PLoS pathogens* **12**, e1005982-e1005982 (2016).
- [0464] 47. K.-L. Siu *et al.*, Middle east respiratory syndrome coronavirus 4a protein is a double-stranded RNA-binding protein that suppresses PACT-induced activation of RIG-I and MDA5 in the innate antiviral response. *Journal of virology* **88**, 4866-4876 (2014).
- [0465] 48. S. C. Jameson, D. Masopust, What Is the Predictive Value of Animal Models for Vaccine Efficacy in Humans? Reevaluating the Potential of Mouse Models for the Human Immune System. *Cold Spring Harbor perspectives in biology* **10**, a029132 (2018).
- [0466] 49. V. Gerdts *et al.*, Large animal models for vaccine development and testing. *Ilar j* **56**, 53-62 (2015).
- [0467] 50. C. Vaure, Y. Liu, A comparative review of toll-like receptor 4 expression and functionality in different animal species. *Front Immunol* **5**, 316 (2014).
- [0468] 51. A. Magadum, K. Kaur, L. Zangi, mRNA-Based Protein Replacement Therapy for the Heart. *Mol. Ther.* **27**, 785-793 (2019).
- [0469] 52. P. S. Kowalski, A. Rudra, L. Miao, D. G. Anderson, Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Molecular therapy : the journal of the American Society of Gene Therapy* **27**, 710-728 (2019).
- [0470] 53. A. K. Blakney, P. F. McKay, R. J. Shattock, Structural Components for Amplification of Positive and Negative Strand VEEV Splitzicons. *Frontiers in Molecular Biosciences* **5**, 71 (2018).
- [0471] 54. A. K. Blakney *et al.*, Effects of cationic adjuvant formulation particle type, fluidity and immunomodulators on delivery and immunogenicity of saRNA. *Journal of Controlled Release* **304**, 65-74 (2019).
- [0472] 55. A. K. Blakney, G. Yilmaz, P. F. McKay, C. R. Becer, R. J. Shattock, One Size Does Not Fit All: The Effect of Chain Length and Charge Density of Poly(ethylene imine) Based Copolymers on Delivery of pDNA, mRNA, and RepRNA Polyplexes. *Biomacromolecules* **19**, 2870-2879 (2018).
- [0473] 56. A. Badamchi-Zadeh *et al.*, Intramuscular Immunisation with Chlamydial Proteins Induces Chlamydia trachomatis Specific Ocular Antibodies. *PLoS ONE* **10**, e0141209 (2015).
- [0474] 57. A. K. Blakney *et al.*, The Skin You Are In: Design-of-Experiments Optimization of Lipid Nanoparticle Self-Amplifying RNA Formulations in Human Skin Explants. *ACS Nano* **13**, 5920-5930 (2019).

도면  
도면1

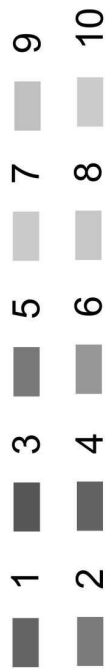
### WT VEEV 레플리콘



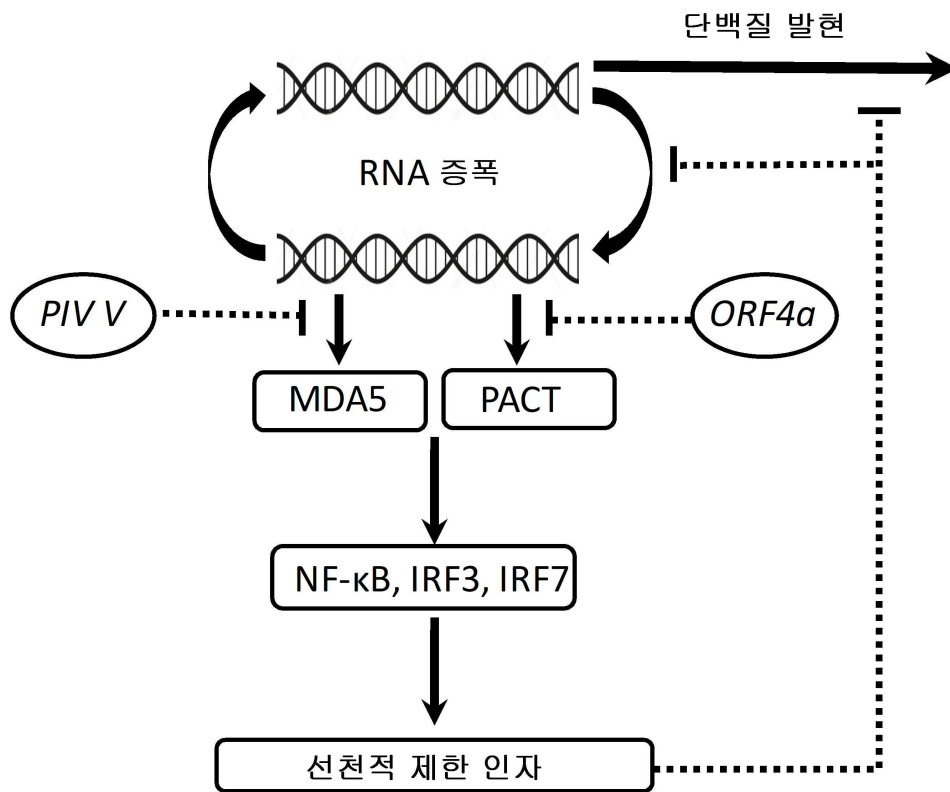
### IIP VEEV 레플리콘



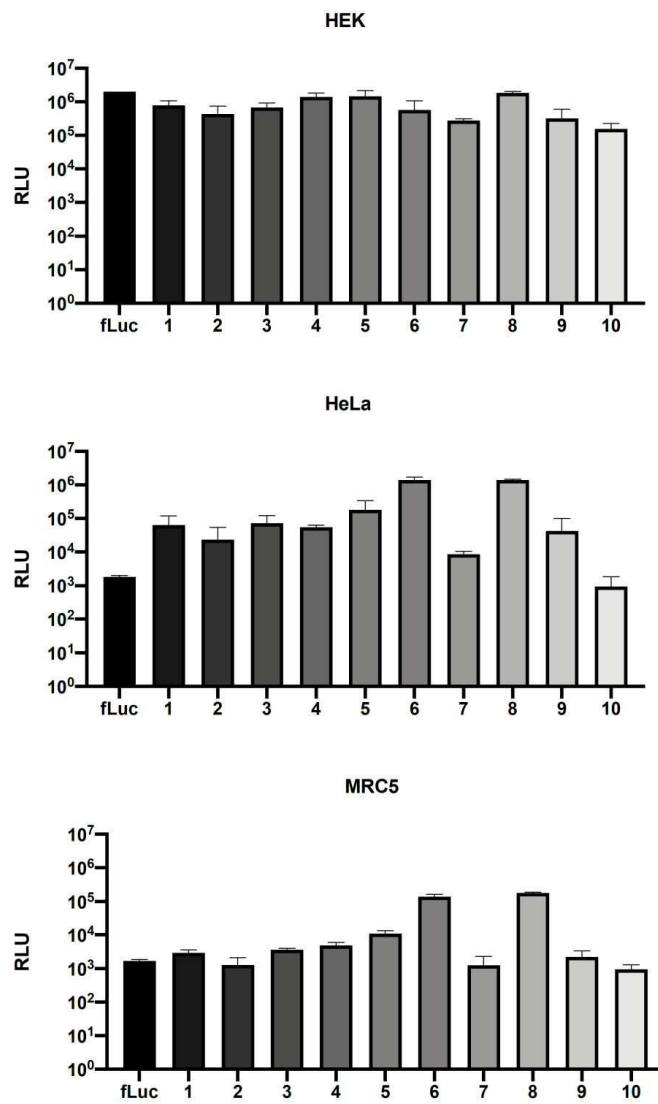
### IIP의 라이브러리



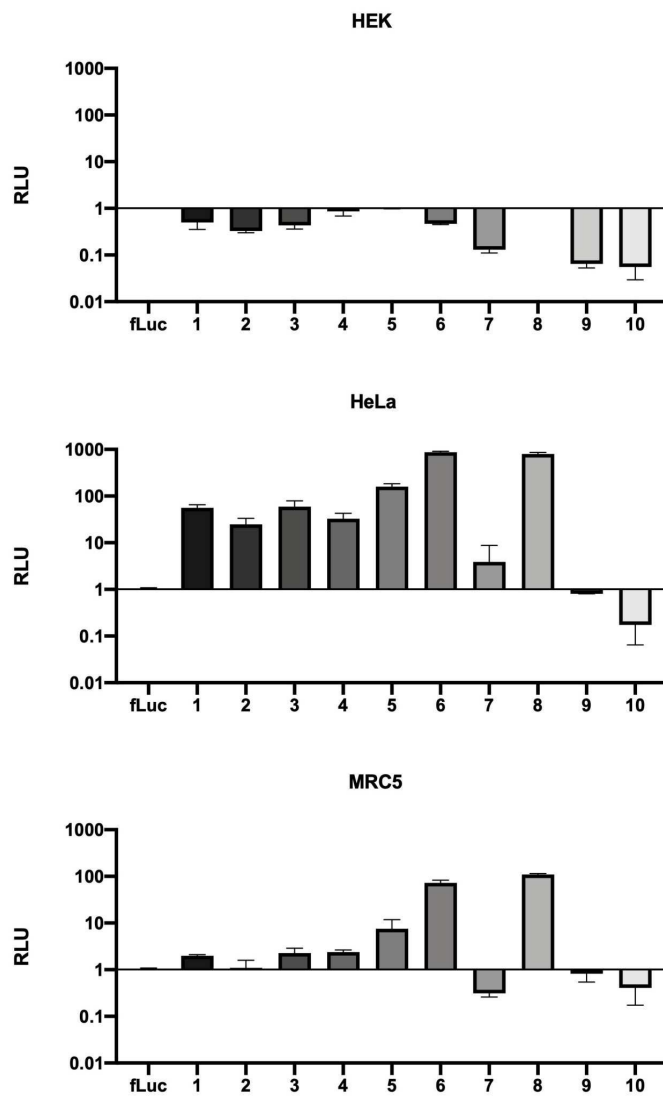
도면2



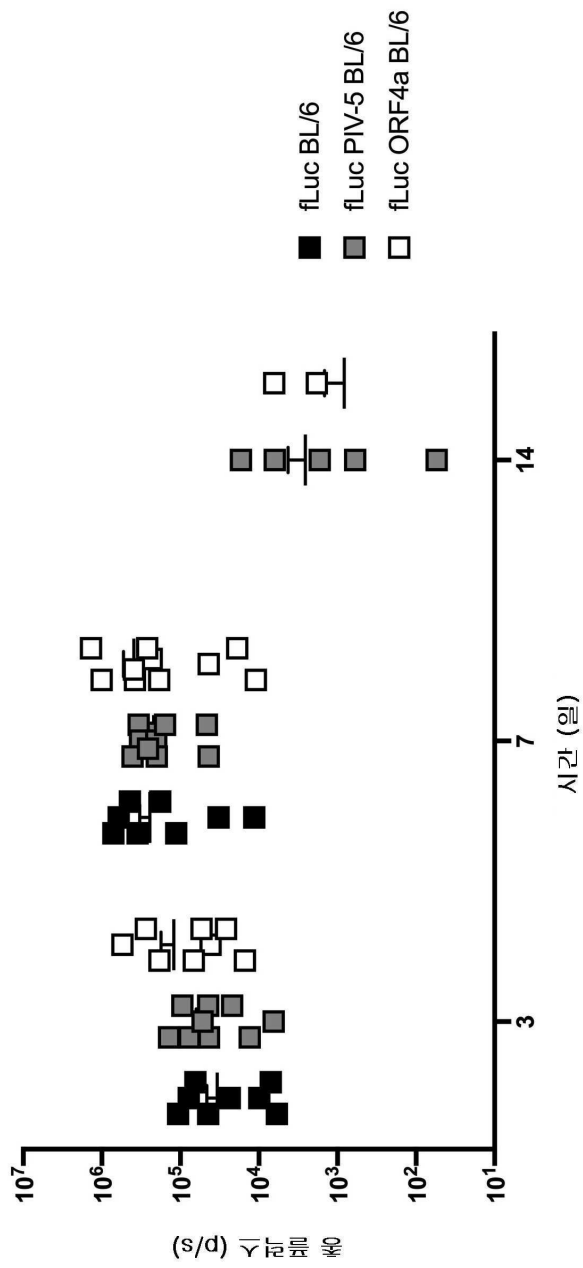
도면3a



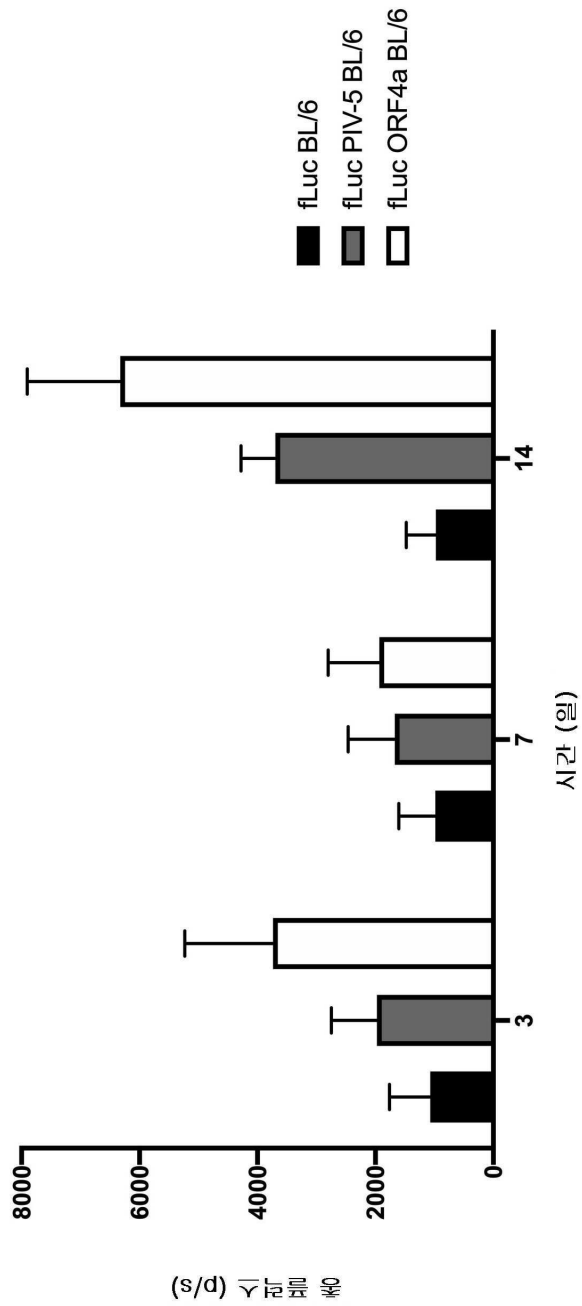
도면3b



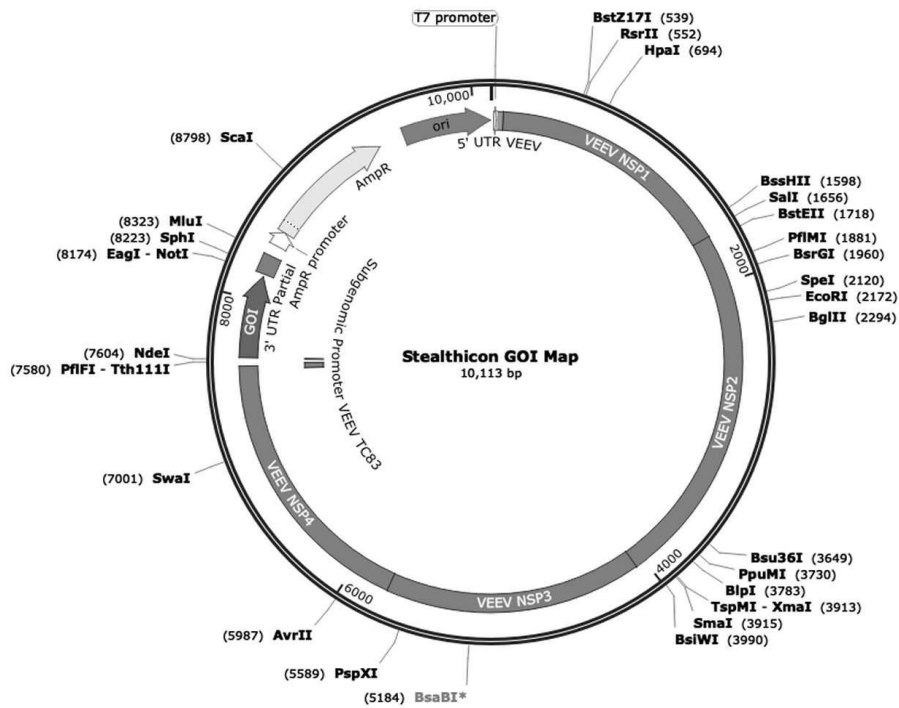
도면4



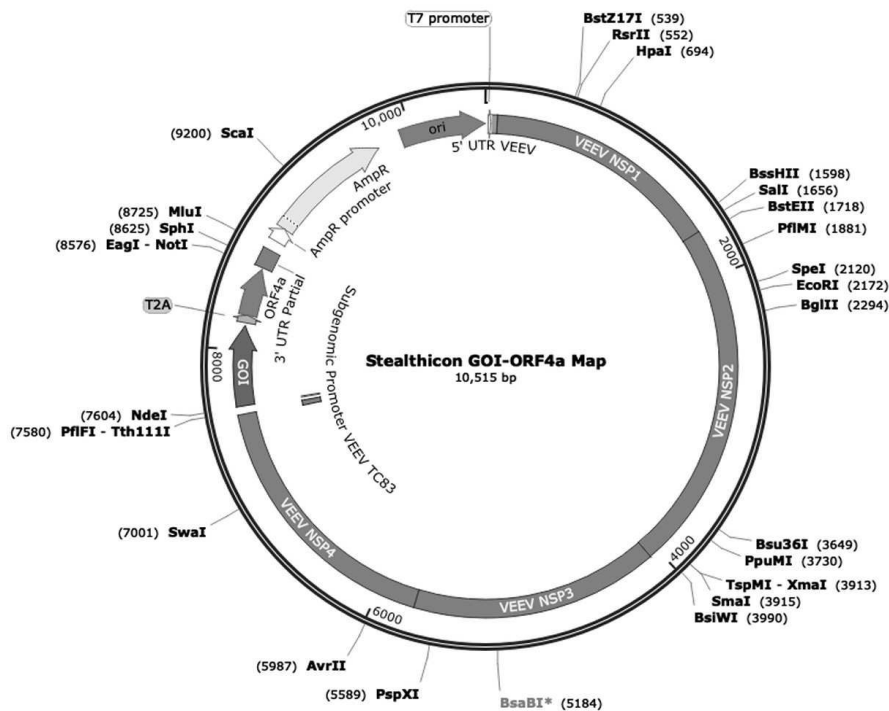
도면5



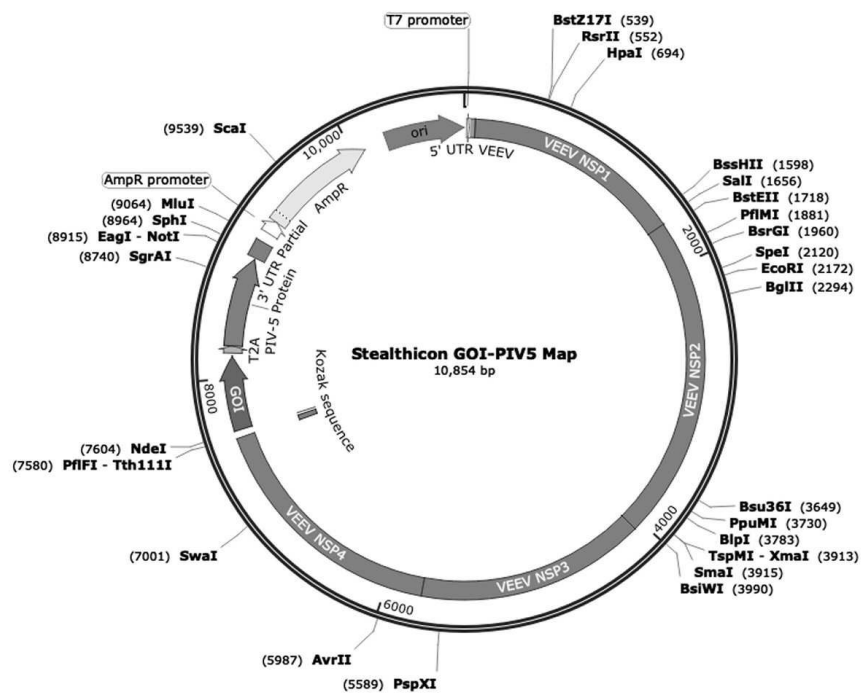
도면6



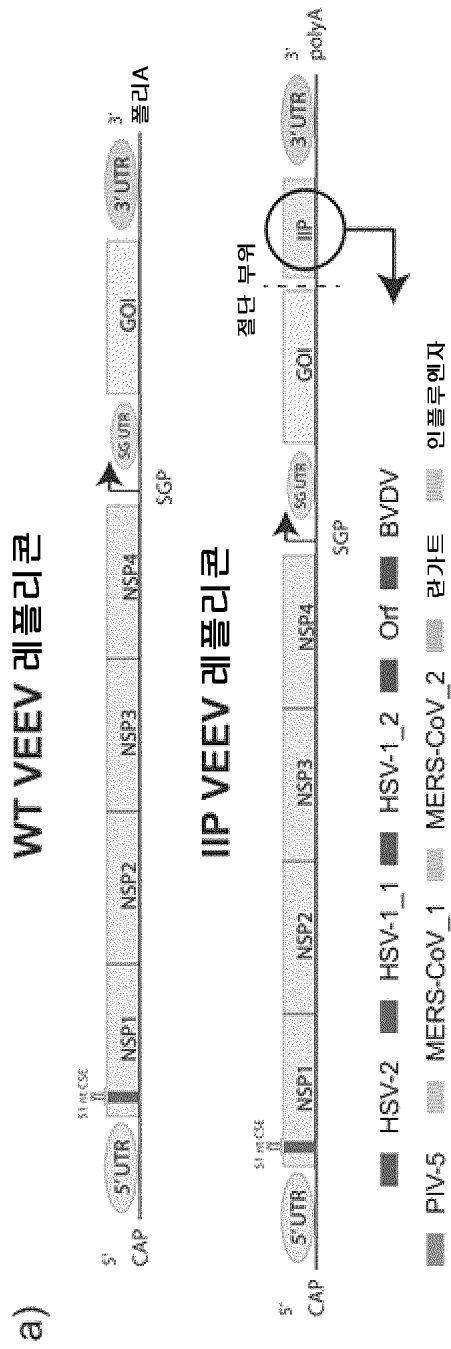
도면7



도면8

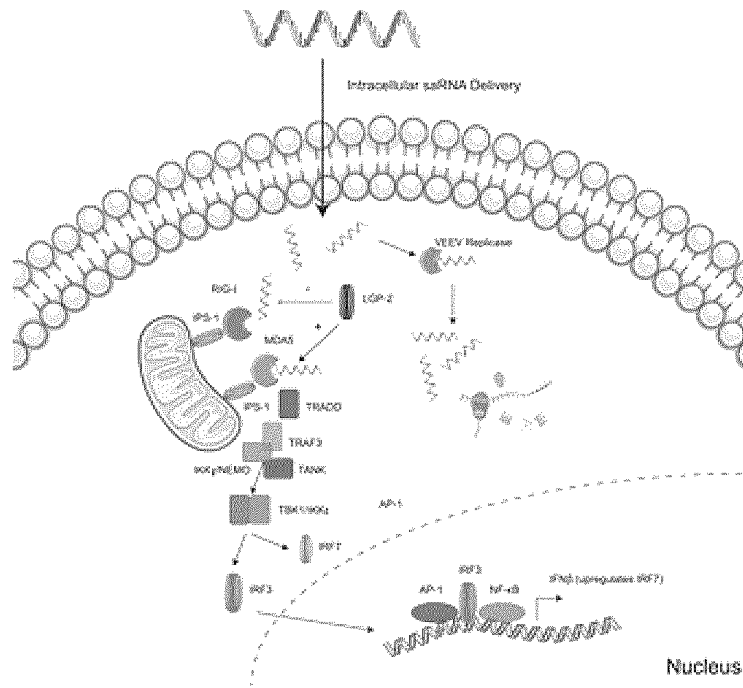


도면9a



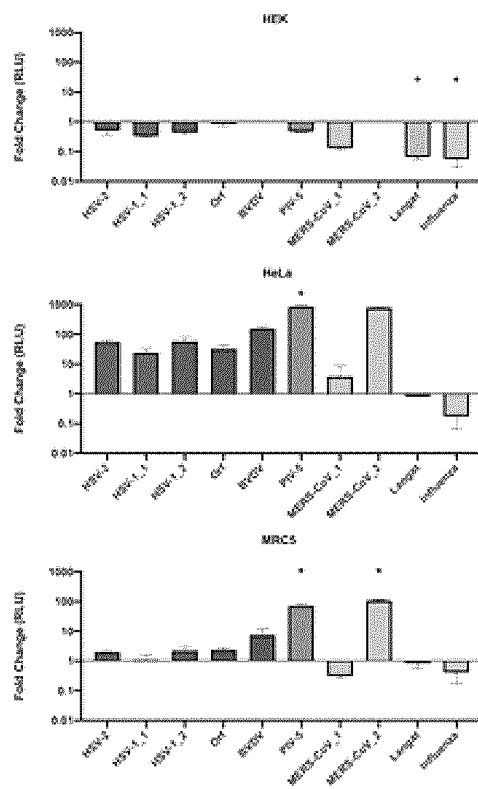
도면9b

b)

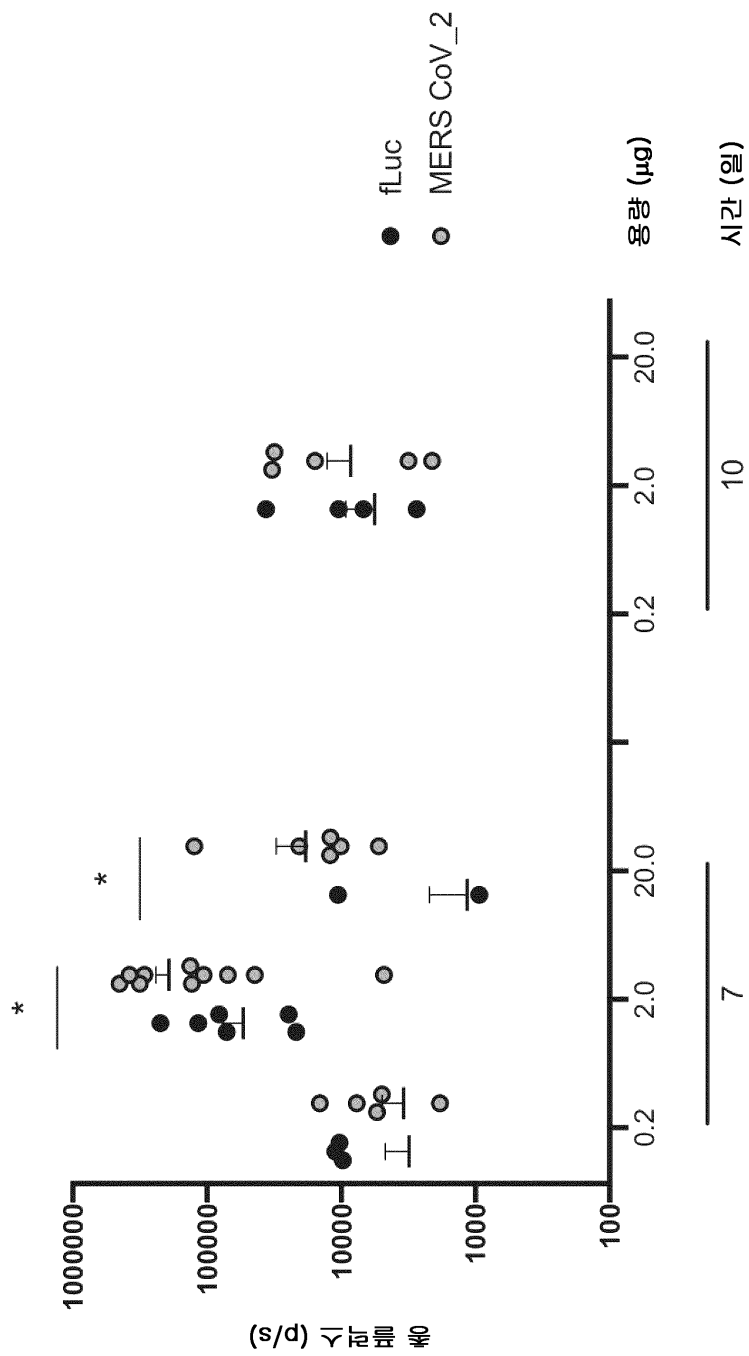


도면9c

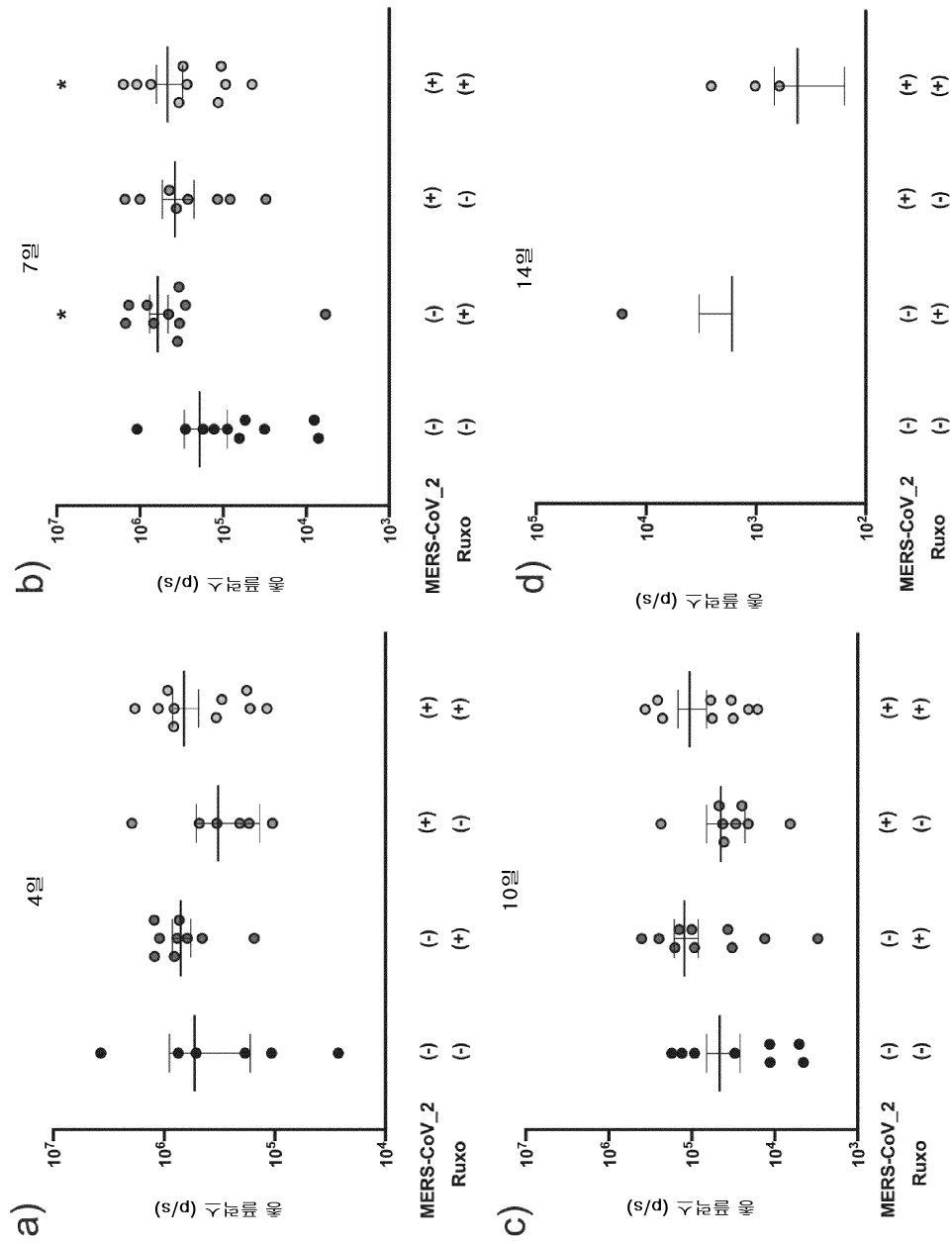
c)



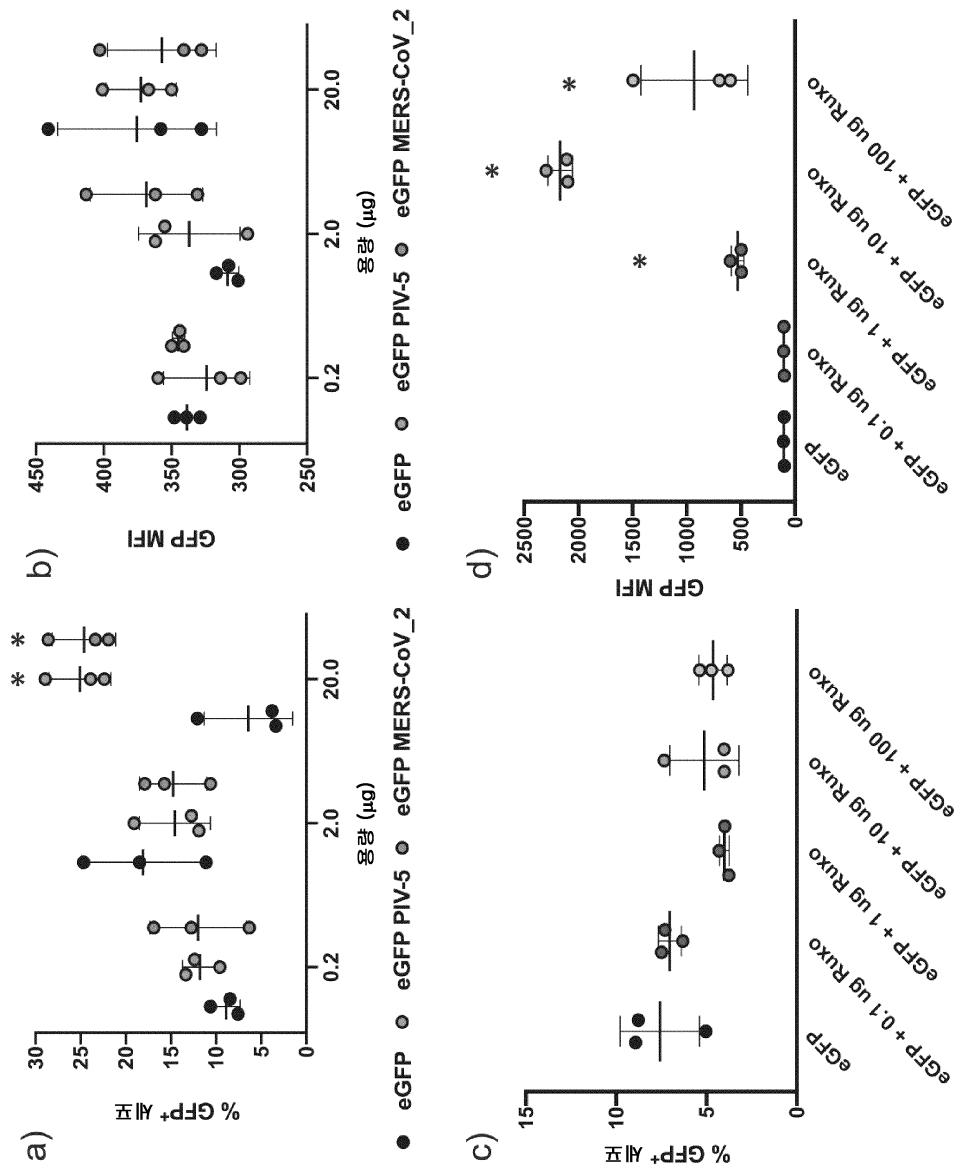
도면10



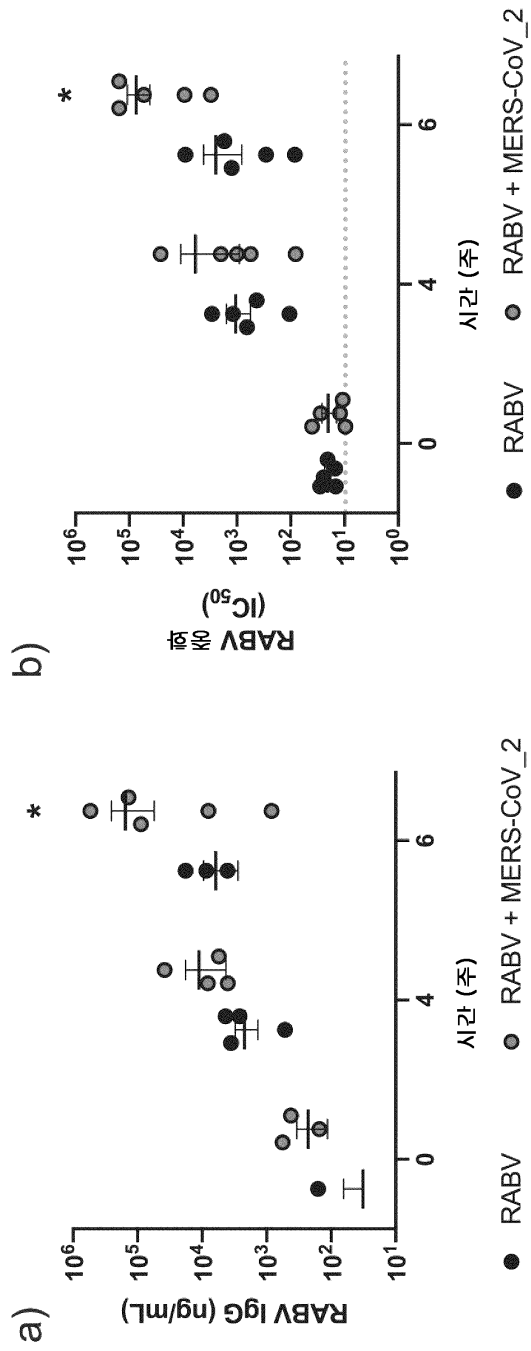
도면11



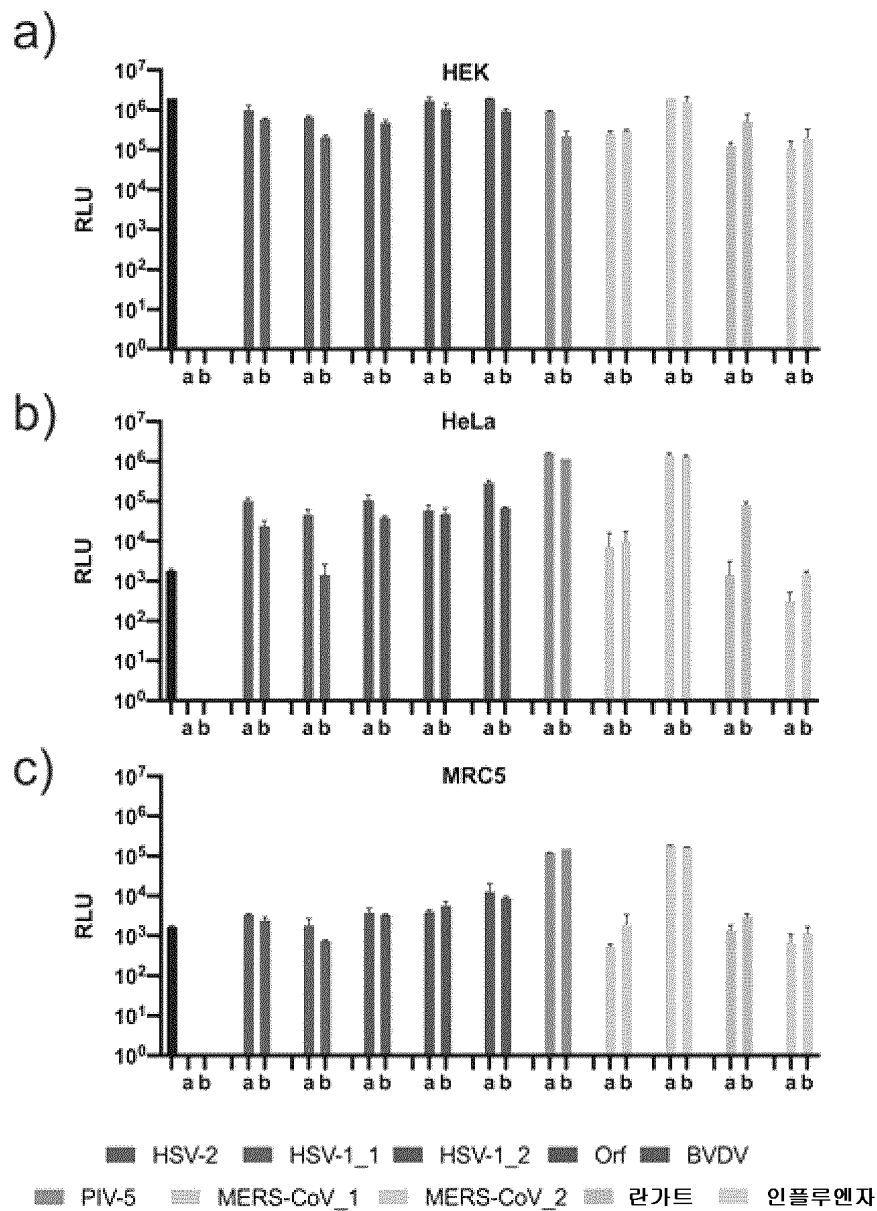
도면12



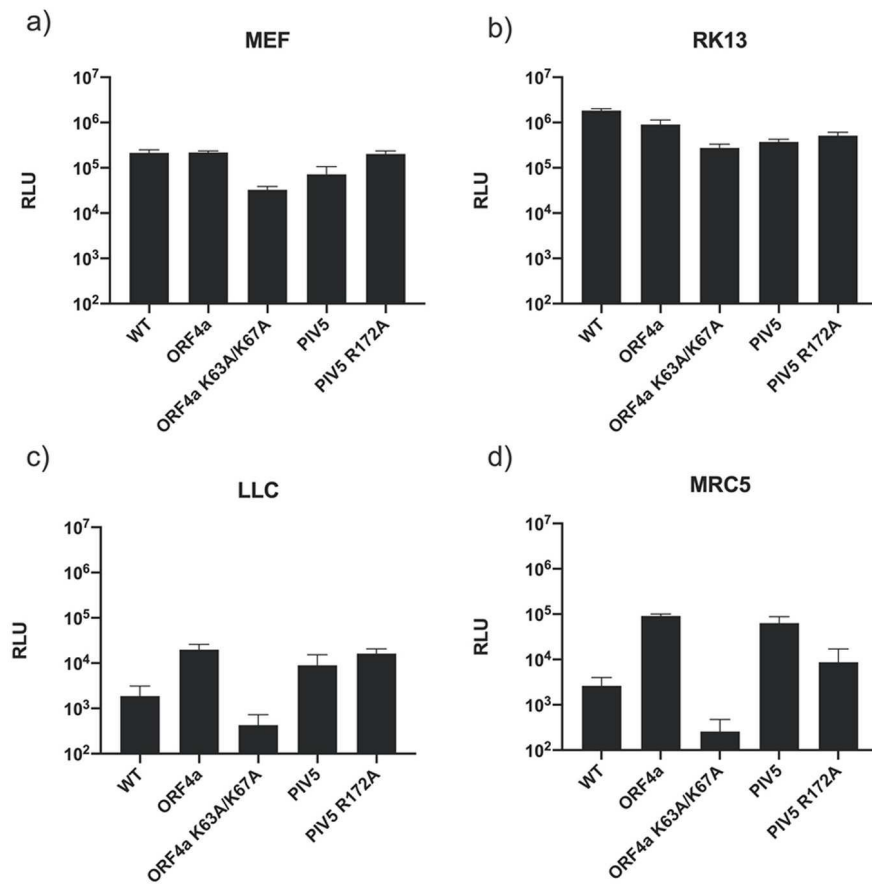
도면13



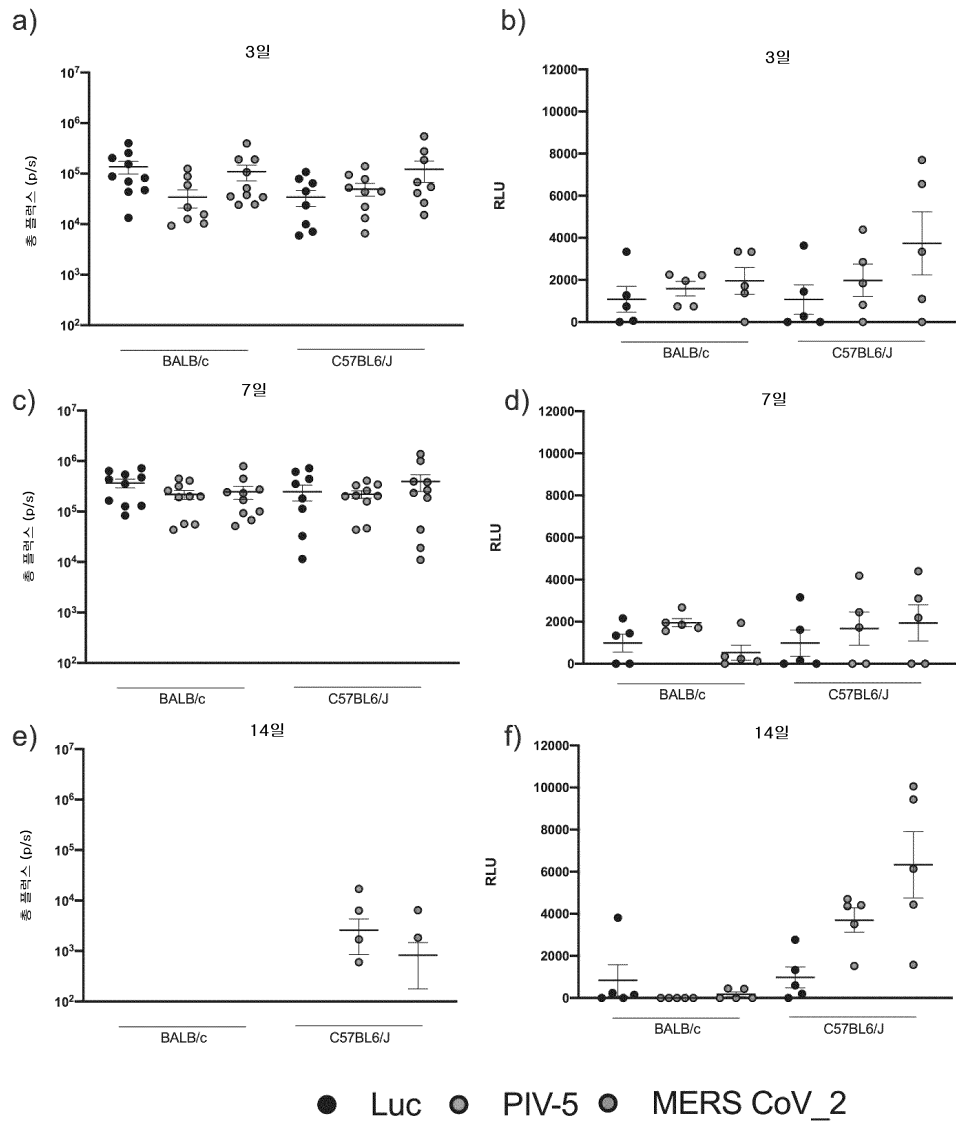
도면14



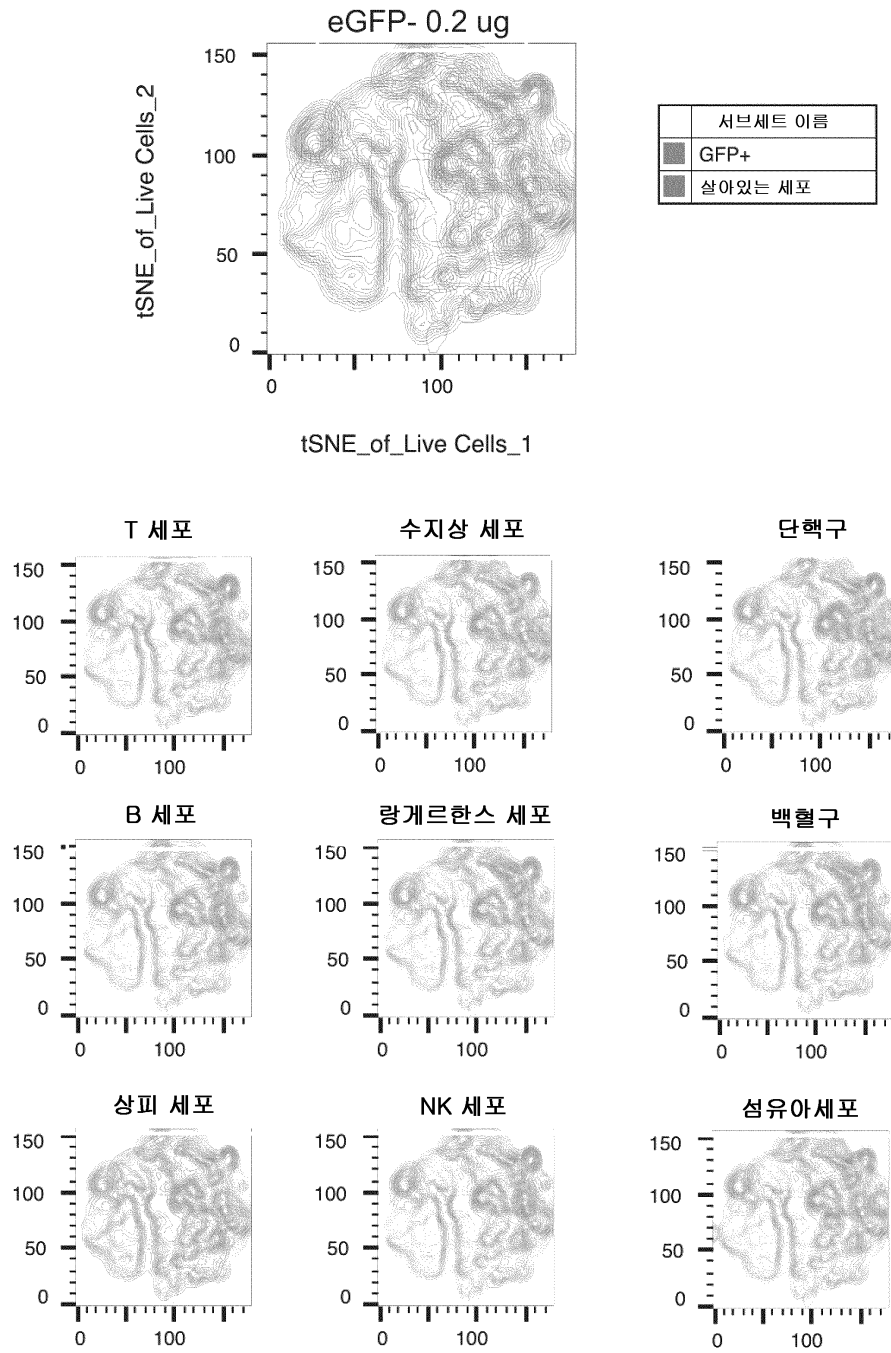
도면15



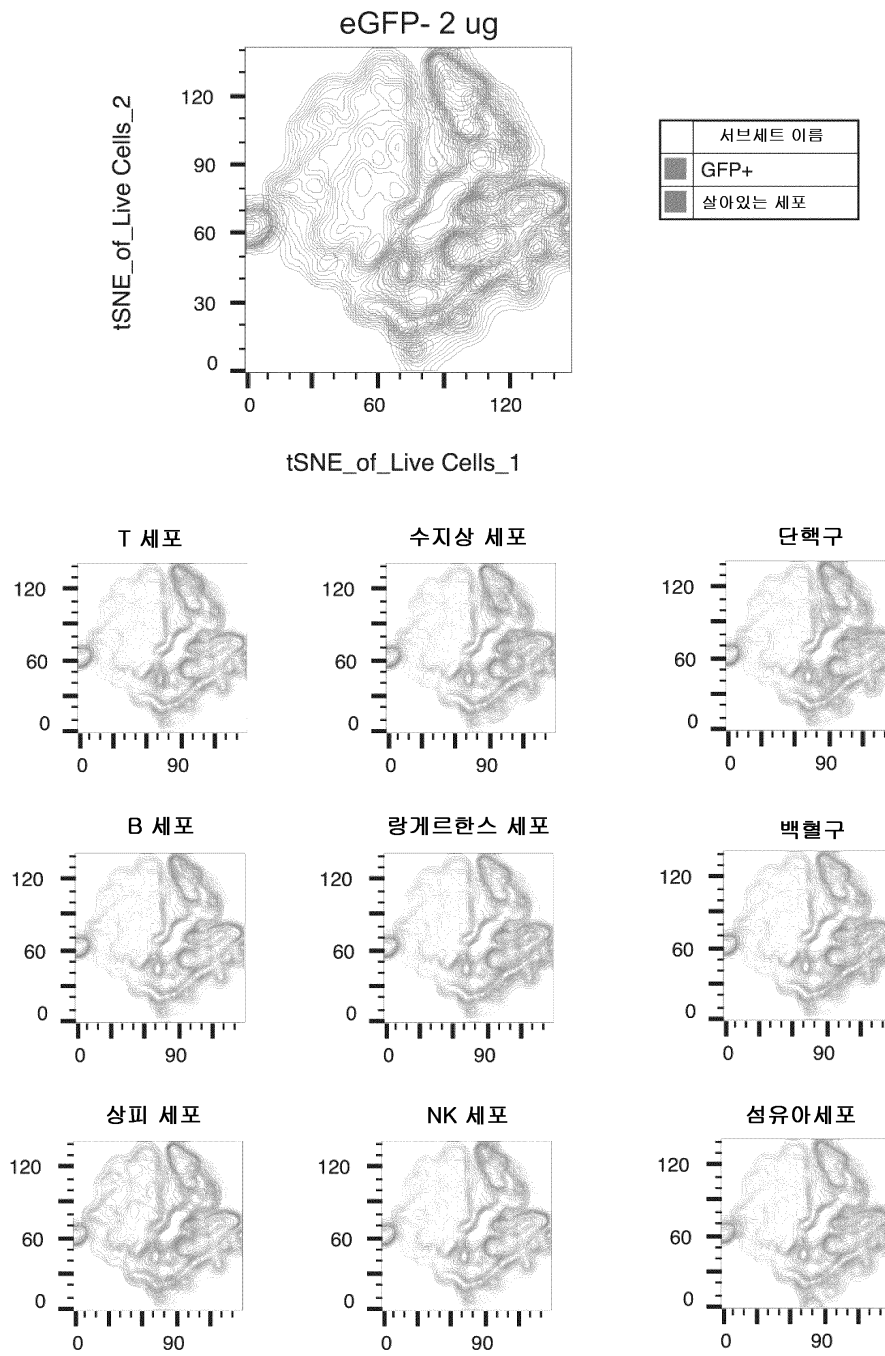
도면16



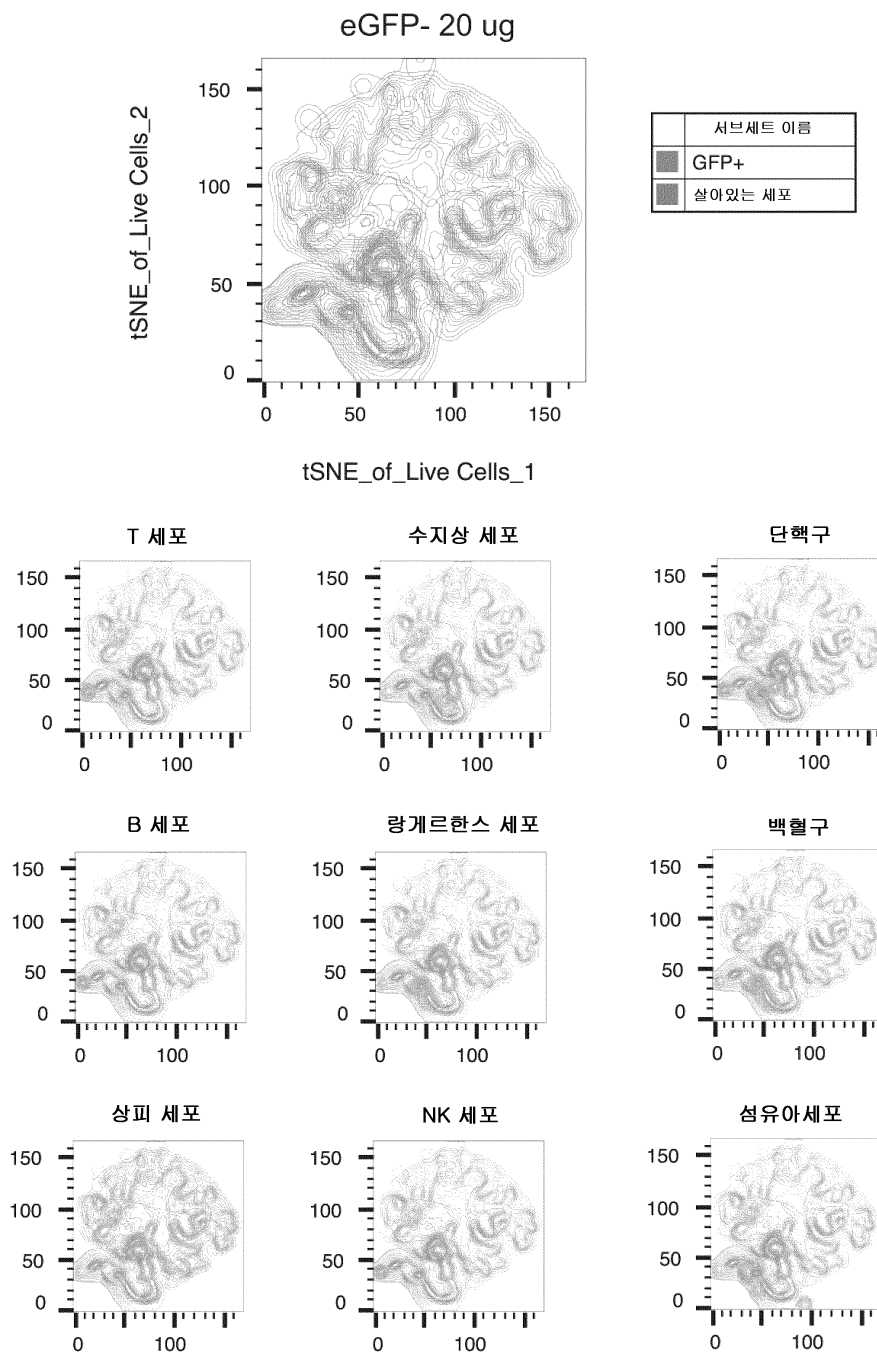
도면17



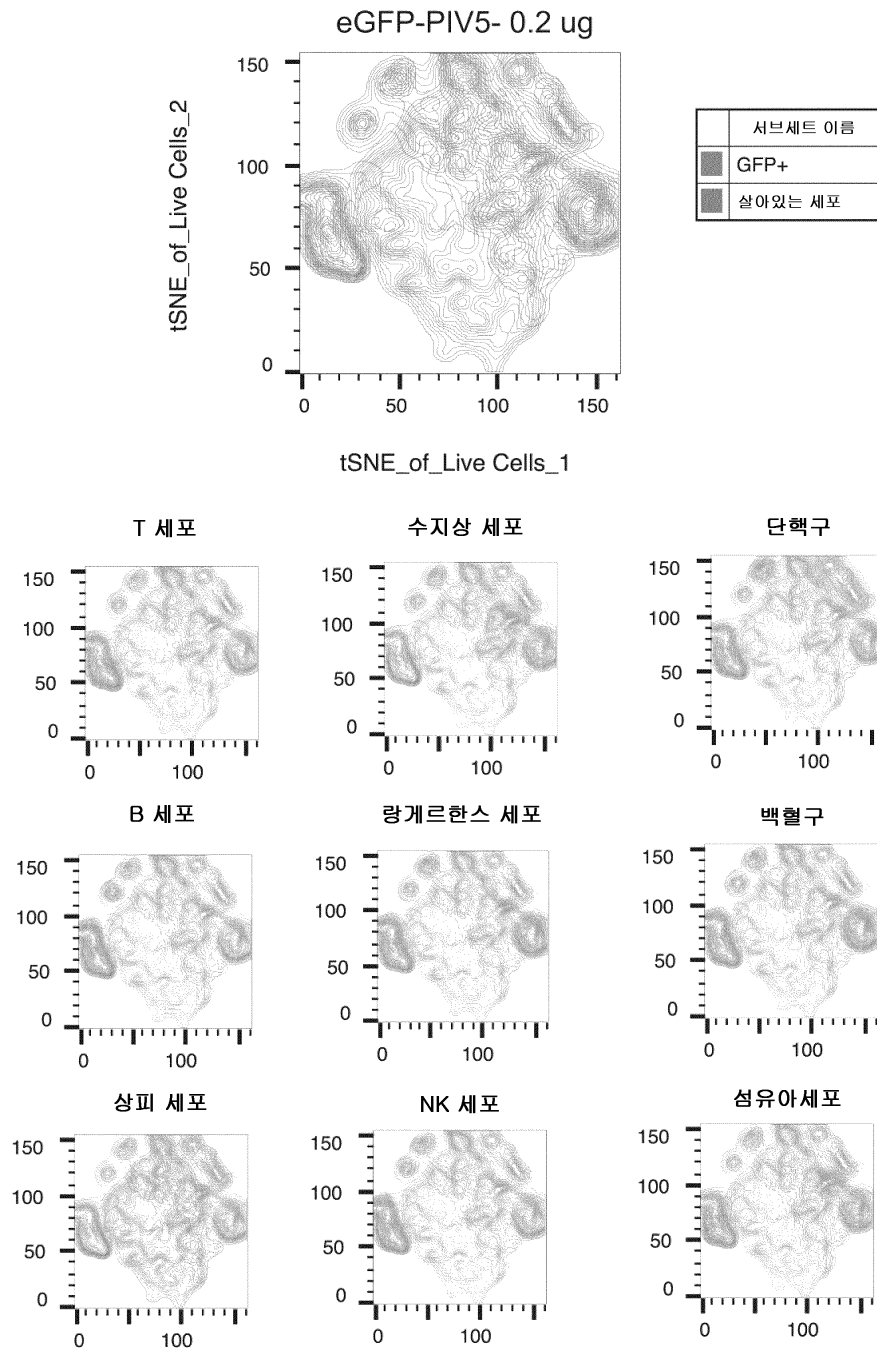
도면18



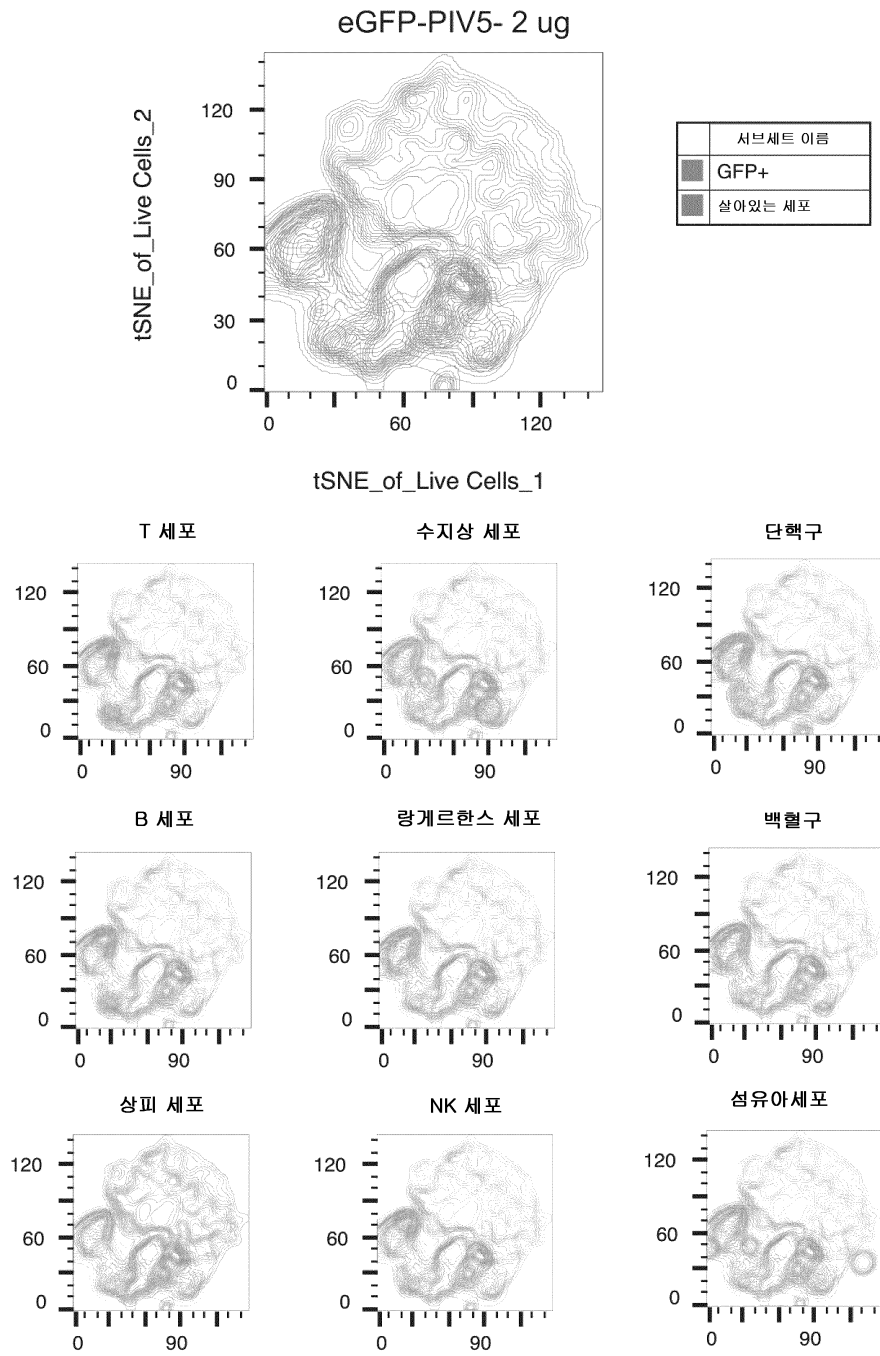
도면19



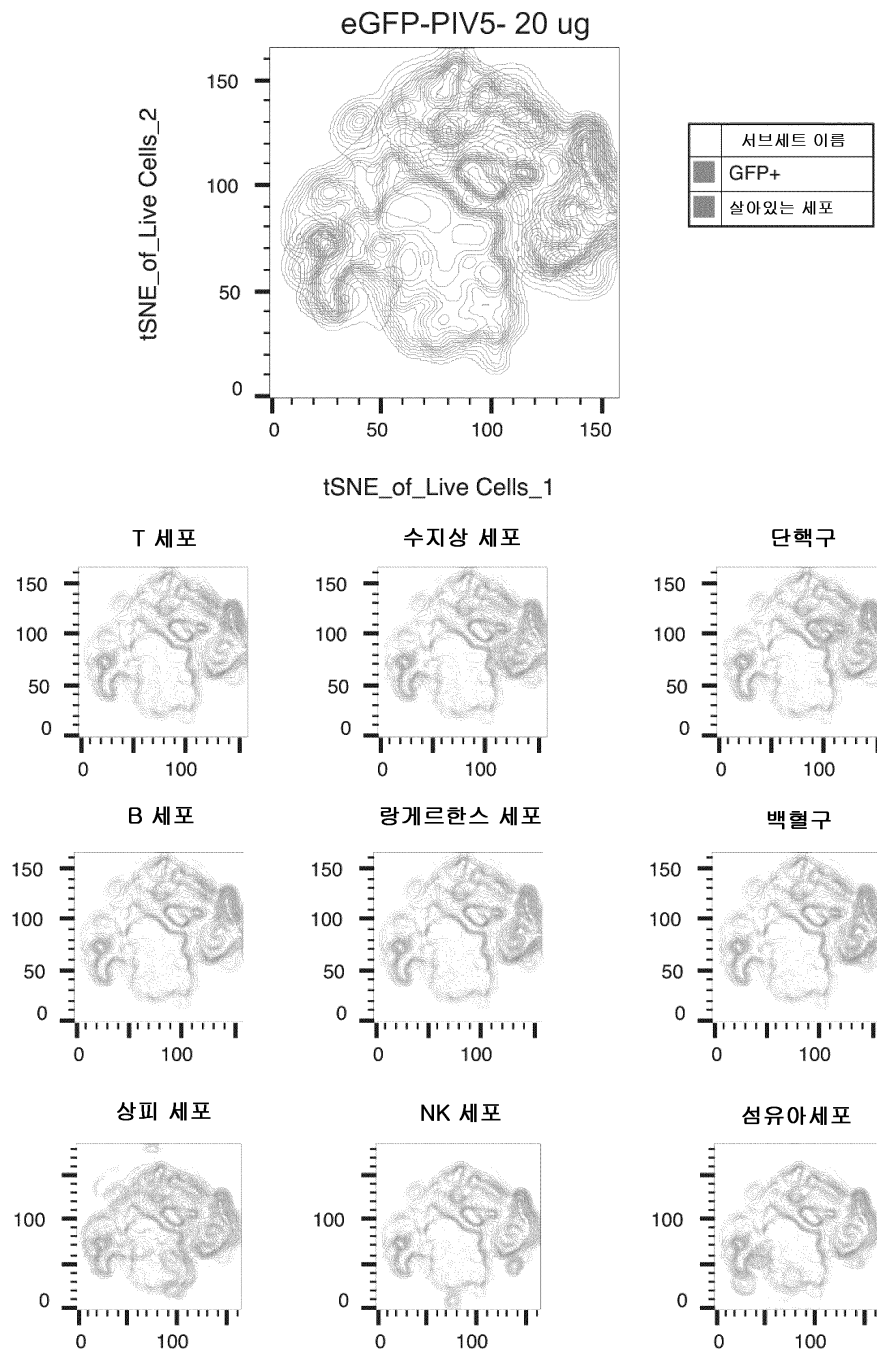
도면20



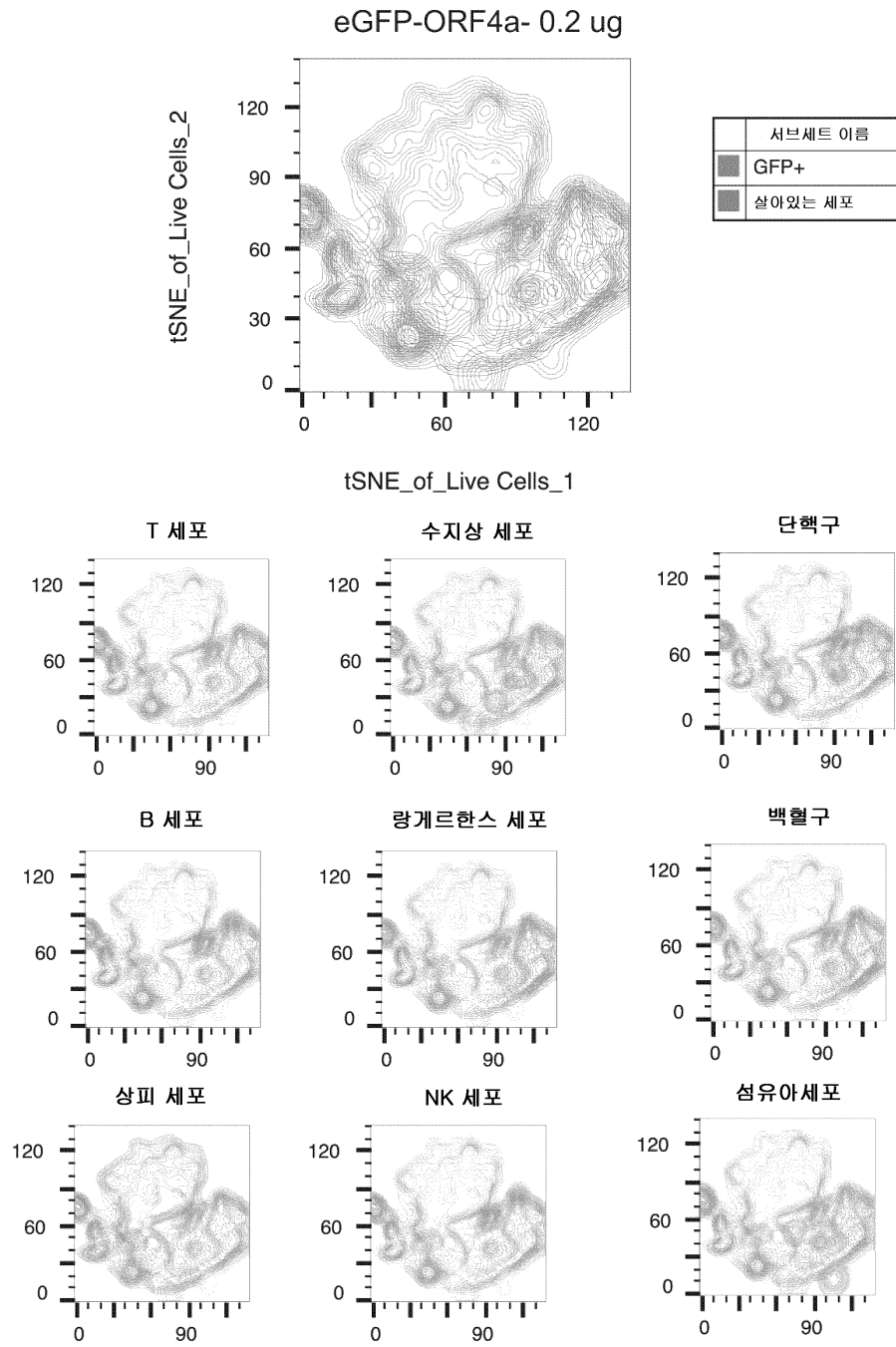
도면21



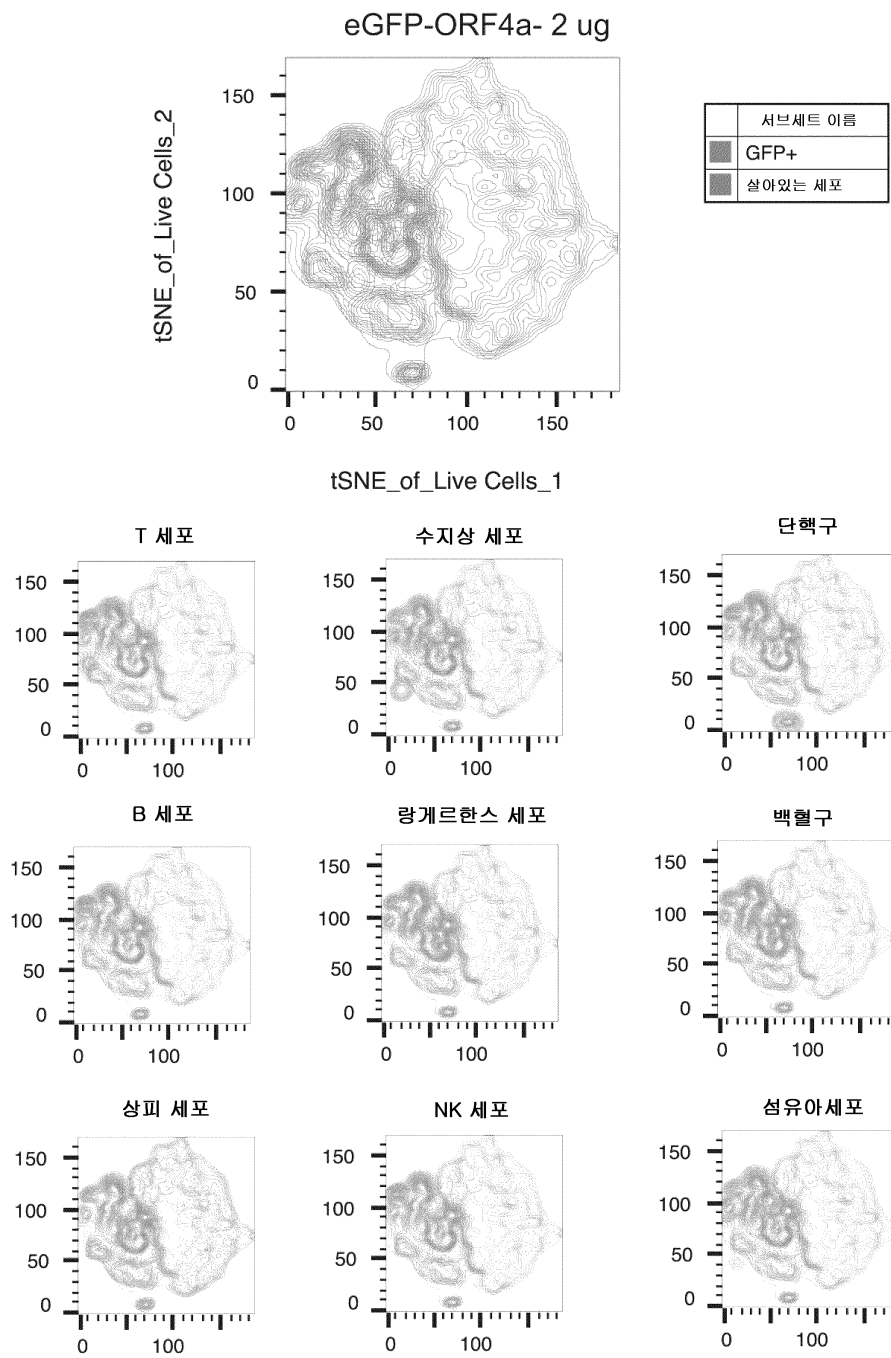
도면22



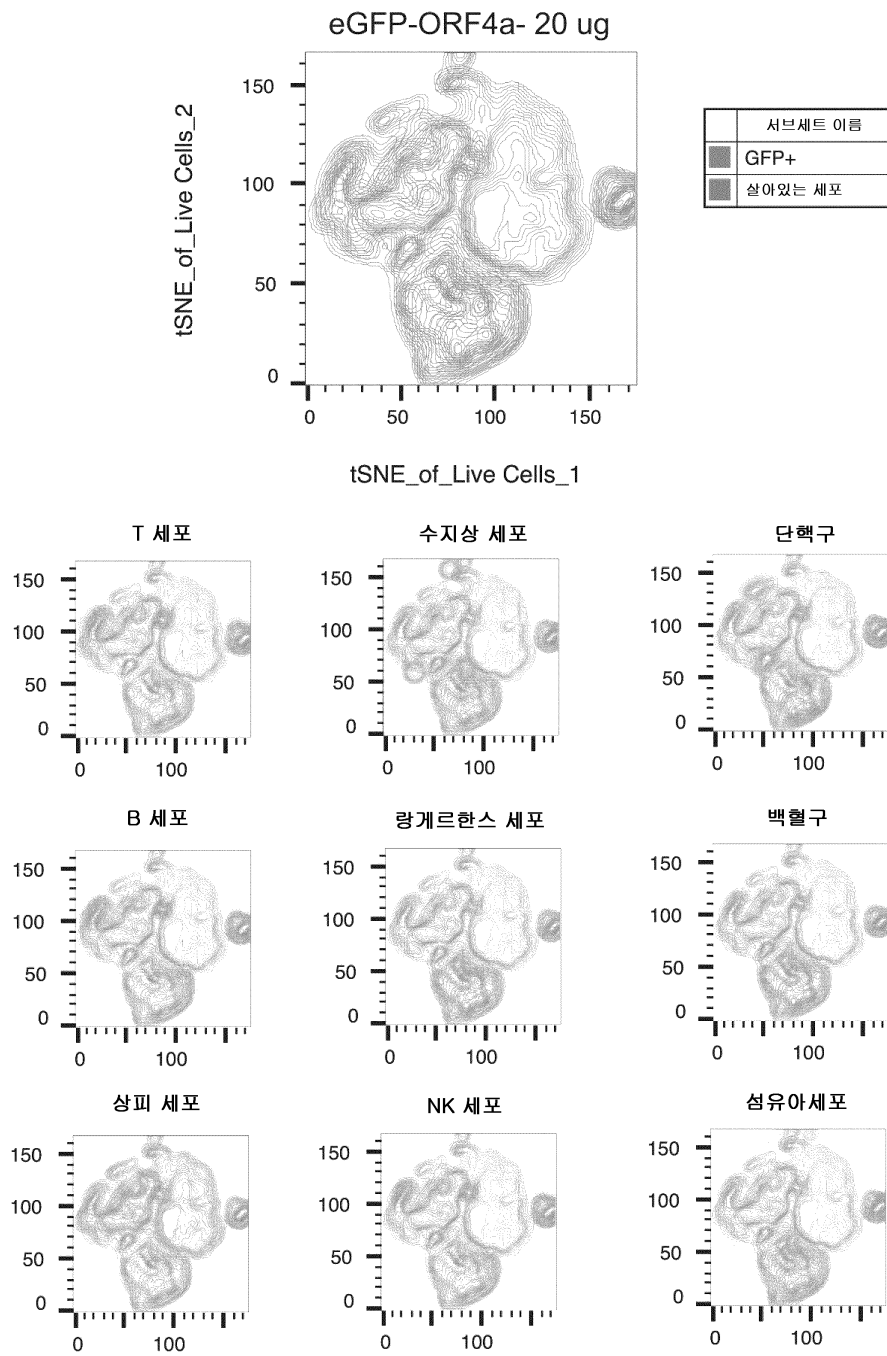
도면23



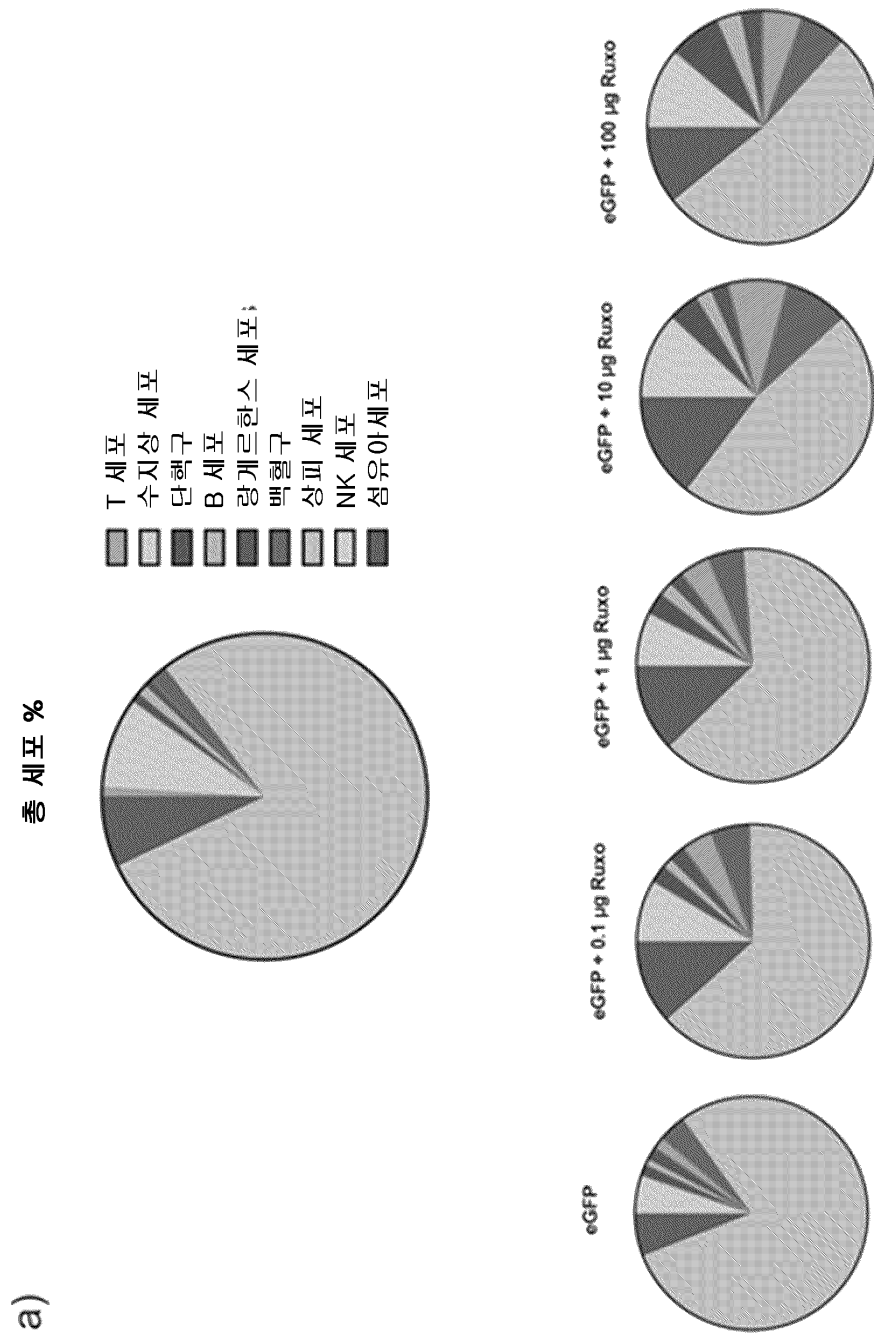
도면24



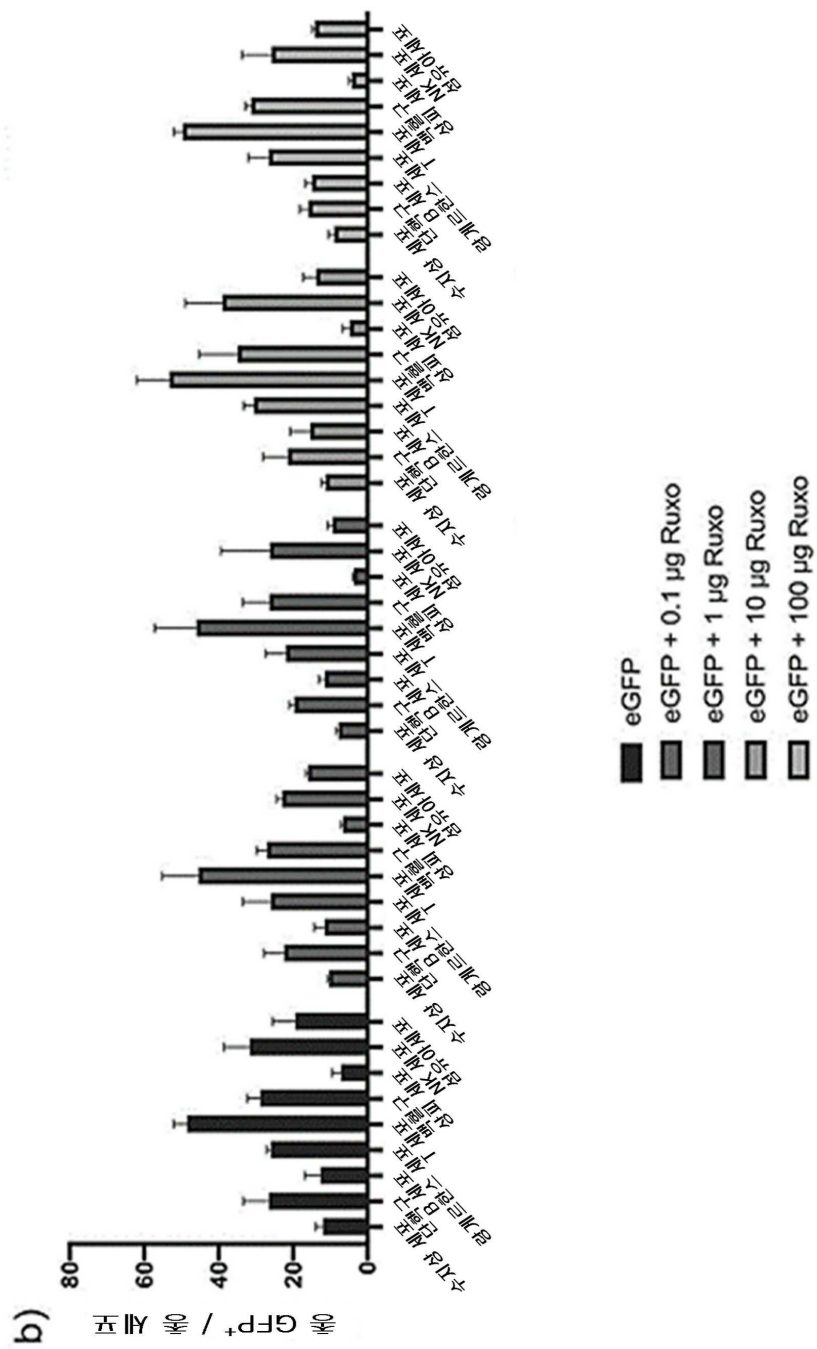
도면25



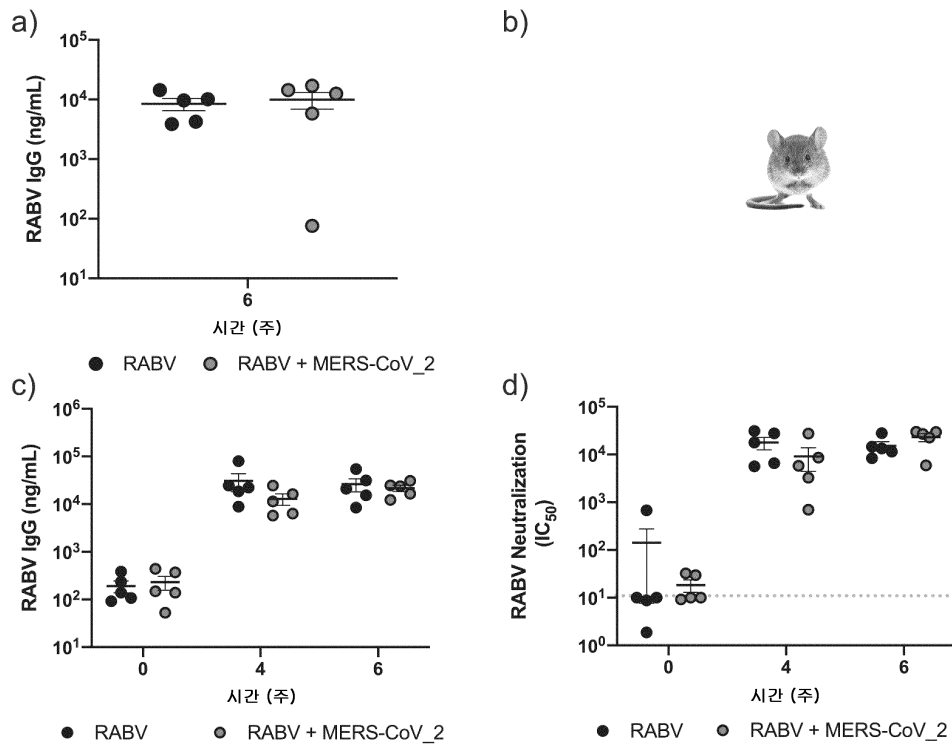
도면26a



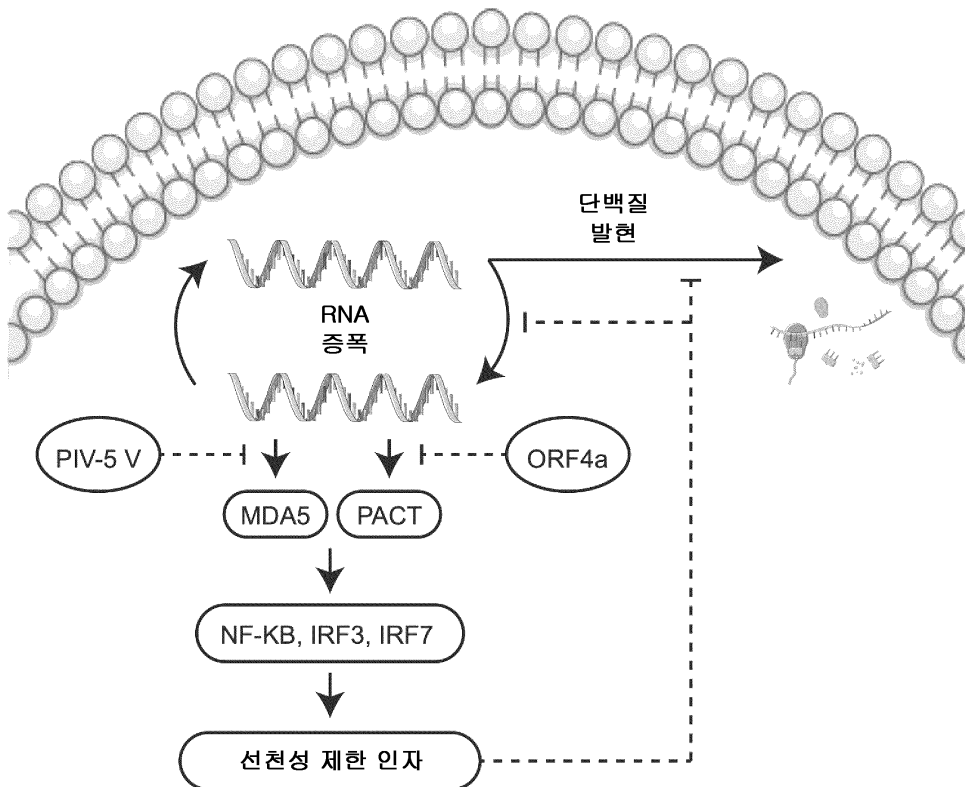
도면26b



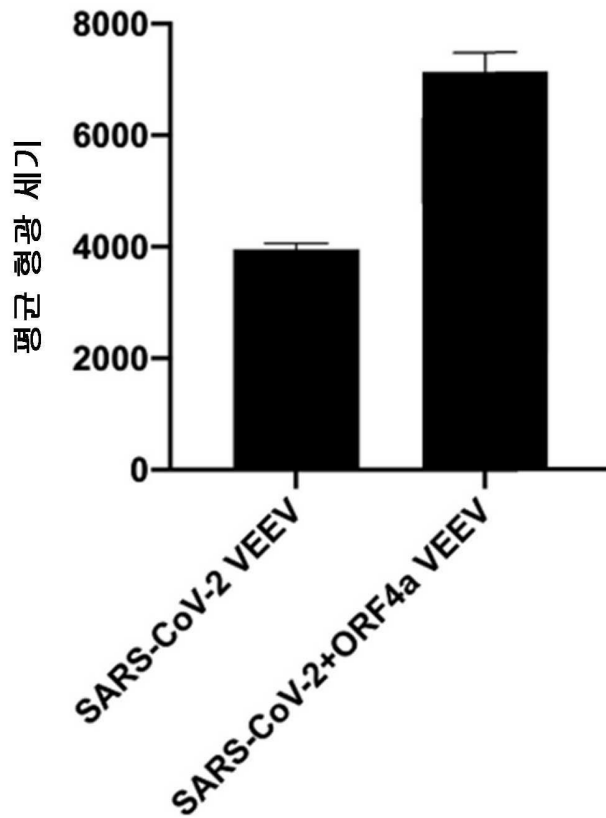
도면27



도면28



도면29



## 서열 목록

<110> Imperial College Innovations Limited

<120> RNA construct

<130> 2021-FPA-1131

<150> 1908729.5

<151> 2019-06-18

<160> 57

<170> PatentIn version 3.5

<210> 1

<211> 196

<212> PRT

<213> human herpesvirus 2

<400> 1

Val Arg Asp Cys Tyr Leu Met Gly Tyr Cys Arg Thr Arg Leu Gly Pro

1 5 10 15

Arg Thr Trp Gly Arg Leu Leu Gln Ile Ser Gly Gly Thr Trp Asp Val

20 25 30

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

100 105 110  
 Glu Ser Ala Ser Gly Gly Ala Ile Ala Ala Arg Leu Glu Cys Glu Phe  
 115 120 125  
 Gly Thr Phe Asp Trp Thr Ser Glu Glu Gly Ser Gln Pro Trp Leu Ser  
 130 135 140  
 Ala Val Val Ala Asp Thr Ser Ser Ala Glu Arg Ser Gly Leu Pro Ala  
 145 150 155 160  
 Pro Gly Ala Cys Arg Ala Thr Glu Ala Pro Glu Arg Glu Asp Gly Cys  
 165 170 175

Arg Lys Met Arg Phe Pro Ala Ala Cys Pro Tyr Pro Cys Gly His Thr  
 180 185 190  
 Phe Leu Arg Pro

195

<210> 2

<211> 588

<212> DNA

<213> human herpesvirus 2

<400> 2

gtgagagact gctacctgat gggctactgc cggacaagac tgggacctag aacatggggc 60  
 agactgctgc agatcagcgg cggaacatgg gatgtgcggc tgagaaacgc catcagagag 120  
 gtggaagccc acttcgagcc tcccgctgaa cctgtgtgtg aactgccctg tctgaacgct 180  
 agaagatacg gccctgagtg cgacgtgggc aacctggaaa caaatggcgg cagcaccagc 240  
 gacgacgaga tttccgatgc caccgacagc gacgatacac tggccagcca cagcgataca 300

gaaggcggac catctcctgc cggaagagag aatcctgagt ctgcctctgg cggagccatt 360  
gccgctagac tggaatgcga gttcggcacc ttcgactgga caagcgagga aggctctcag 420  
ccttggtgtgt ctgctgtggt ggccgataca agctctgccg agagaagtgg acttcctgct 480  
cctggcgcct gtagagctac agaggctcct gaaagagagg acggctgcag aaagatgcgg 540  
ttccctgccg cctgtcctta tccttgccgc cacacatttc tgcggccc 588

<210> 3

<211> 395

<212> PRT

<213> human herpesvirus 1

<400> 3

Val Arg Asp Cys Tyr Leu Met Gly Tyr Cys Arg Thr Arg Leu Gly Pro

1 5 10 15

Arg Thr Trp Gly Arg Leu Leu Gln Ile Ser Gly Gly Thr Trp Asp Val

20 25 30

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

35 40 45

Ala Glu Pro Val Cys Glu Leu Pro Cys Leu Asn Ala Arg Arg Tyr Gly

50 55 60

Pro Glu Cys Asp Val Gly Asn Leu Glu Thr Asn Gly Gly Ser Thr Ser

65 70 75 80

Asp Asp Glu Ile Ser Asp Ala Thr Asp Ser Asp Asp Thr Leu Ala Ser

85 90 95

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

100 105 110

Glu Ser Ala Ser Gly Gly Ala Ile Ala Ala Arg Leu Glu Cys Glu Phe

115 120 125

Gly Thr Phe Asp Trp Thr Ser Glu Glu Gly Ser Gln Pro Trp Leu Ser

130 135 140

Ala Val Val Ala Asp Ile Arg Asp Cys Tyr Leu Met Gly Tyr Cys Arg

145 150 155 160

Ala Arg Leu Ala Pro Arg Thr Trp Cys Arg Leu Leu Gln Val Ser Gly

165 170 175

Gly Thr Trp Gly Met His Leu Arg Asn Thr Ile Arg Glu Val Glu Ala  
180 185 190

Arg Phe Asp Ala Thr Ala Glu Pro Val Cys Lys Leu Pro Cys Leu Glu  
195 200 205

Thr Arg Arg Tyr Gly Pro Glu Cys Asp Leu Ser Asn Leu Glu Ile His  
210 215 220

Leu Ser Ala Thr Ser Asp Asp Glu Ile Ser Asp Ala Thr Asp Leu Glu  
225 230 235 240

Ala Ala Gly Ser Asp His Thr Leu Ala Ser Gln Ser Asp Thr Glu Asp  
245 250 255

Ala Pro Ser Pro Val Thr Leu Glu Thr Pro Glu Pro Arg Gly Ser Leu  
260 265 270

Ala Val Arg Leu Glu Asp Glu Phe Gly Glu Phe Asp Trp Thr Pro Gln  
275 280 285

Glu Gly Ser Gln Pro Trp Leu Ser Ala Val Val Ala Asp Thr Ser Ser  
290 295 300

Val Glu Arg Pro Gly Pro Ser Asp Ser Gly Ala Gly Arg Ala Ala Glu  
305 310 315 320

Asp Arg Lys Cys Leu Asp Gly Cys Arg Lys Met Arg Phe Ser Thr Ala  
325 330 335

Cys Pro Tyr Pro Cys Ser Asp Thr Phe Leu Arg Pro Thr Ser Ser Ala  
340 345 350

Glu Arg Ser Gly Leu Pro Ala Pro Gly Ala Cys Arg Ala Thr Glu Ala  
355 360 365

Pro Glu Arg Glu Asp Gly Cys Arg Lys Met Arg Phe Pro Ala Ala Cys  
370 375 380

Pro Tyr Pro Cys Gly His Thr Phe Leu Arg Pro  
385 390 395

<210> 4  
<211> 597  
<212> DNA  
<213> human herpesvirus 1

<400> 4

atcagagact gctacctgat gggctactgc cgggctagac tggcccctag aacatggtgc 60  
agactgctgc aagtgtctgg cggcacatgg ggcatgcacc tgagaaacac catcagagag 120

gtggaagcca gattcgacgc cacagccgag cctgtgtgca agctgccttg tctggaaact 180  
cggagatacg gccccgagtg cgacctgagc aatctggaaa ttcacctgag cgccaccagc 240  
gacgacgaga tttctgatgc caccgacctg gaagccgccg gatctgatca tacactggcc 300  
agccagagcg acaccgagga tgctccatct ccagtgactc tggaaacccc tgagcctaga 360  
ggatctctgg ccgtgcgact ggaagatgag ttcggcgagt tcgactggac ccctcaagag 420  
ggatctcagc cttagctgtc tgccgtggtg gccgatacaa gcagcgtgga aagaccgga 480  
cctagcgatt ctggtgctgg cagagccgcc gaggatagaa agtgcctgga tggctgccgg 540

aagatgcggt tctctaccgc ctgtccatat ccttgcagcg acaccttctt gcggcct 597

<210> 5

<211> 161

<212> PRT

<213> human herpesvirus 1

<400> 5

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

1 5 10 15

Val Tyr Leu Lys Gly Val Pro Ser Ala Gly Met His Pro Arg Gly Val

20 25 30

His Ala Pro Arg Gly His Pro Arg Met Ile Ser Gly Pro Pro Gln Arg

35 40 45

Gly Asp Asn Asp Gln Ala Ala Gly Gln Cys Gly Asp Ser Gly Leu Leu

50 55 60

Arg Val Gly Ala Asp Thr Thr Ile Ser Lys Pro Ser Glu Ala Val Arg

65 70 75 80

Pro Pro Thr Ile Pro Arg Thr Pro Arg Val Pro Arg Glu Pro Arg Val

85 90 95

Pro Arg Pro Pro Arg Glu Pro Arg Glu Pro Arg Val Pro Arg Ala Pro

100 105 110

Arg Asp Pro Arg Val Pro Arg Asp Pro Arg Asp Pro Arg Gln Pro Arg

115 120 125  
 Ser Pro Arg Glu Pro Arg Ser Pro Arg Glu Pro Arg Ser Pro Arg Glu  
 130 135 140  
 Pro Arg Thr Pro Arg Thr Pro Arg Glu Pro Arg Thr Ala Arg Gly Ser  
 145 150 155 160  
 Val

<210> 6  
 <211> 483  
 <212> DNA  
 <213> human herpesvirus 1  
 <400> 6  
 atgagccaga cacagcctcc agctccagtt ggacctggcg accctgatgt gtatctgaag 60  
 ggcgtgccaa gcgccggcat gcctcctaga ggtgttcatg cccctagagg acaccccaga 120  
 atgatctctg gccctctca gagaggcgac aacgatcagg ctgctggaca gtgtggcgat 180  
 agcggactgc tgagagtggg cgccgatacc acaatcagca agccatctga ggctgtgcgg 240  
 cctctctaca tccccagaac acctagagtg ccccgcgagc caagagtgcc tagacctct 300  
 agagagccca gagaaccag agtgccaagg gctcccagag atcctagagt ccctcgggac 360  
 cctagggacc caagacaacc tagatcacc agagagcctc ggagcccaag agagccaaga 420  
 agccctaggg aaccccgac accaagaaca cccagggaac ctagaaccgc cagaggcagc 480  
 gtg 483

<210> 7  
 <211> 183  
 <212> PRT  
 <213> Orf virus  
 <400> 7

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

Val Cys Cys Glu Asp Gly Asn Pro Pro Arg Trp Phe Val Glu Cys Ala

50 55 60  
Pro Ser Ala Pro Thr Glu Glu Asp Glu Asn Ser Asp Thr Glu Pro Met  
65 70 75 80  
Glu Thr Glu Ala Gly Cys Asp Thr Leu Phe Gly Gly Asp Ile Asp Ile  
85 90 95  
Met Thr Gln Ser Ala Val Ile Arg Leu Lys Ser Leu Asn Pro Val Ser  
100 105 110  
Ala Val Asn Glu Phe Cys Met Met Thr His Arg Pro Leu Glu Phe Cys  
115 120 125

Glu Thr Arg Ala Gly Gly Glu Asp His Cys Pro Arg Phe Thr Cys Thr  
130 135 140  
Ile Thr Ile Ser Gly Lys Val Val Ala Val Ala Asp Gly Ala Ser Lys  
145 150 155 160  
Lys Leu Ala Arg His Thr Ala Cys Ser Ser Ala Leu Thr Ile Leu Ile  
165 170 175  
Asn Asn Cys Gly Ile Ser Phe  
180

<210> 8  
<211> 549  
<212> DNA  
<213> Orf virus  
<400> 8

atggcctgtg aatgcgccag cctgatacctg gaactgctga gaaagagcga cgacaagctg 60  
cccgccaagc agatcgccaa agagctgggc atctctaagc acgaggccaa ccggcagctg 120  
taccggctgc tggattctga cgaagtgtgc tgcgaggacg gcaatcctcc tcgttggttc 180  
gtggaatgtg cccctagcgc tcccaccgaa gaggacgaga atagcgacac cgagcctatg 240  
gaaaccgagg ccggctgcga tacactgttt ggcgagaca tcgacatcat gacccagagc 300  
gccgtgatcc ggctgaagtc cctgaatcct gtgtccgccg tgaacgagtt ctgcatgatg 360  
accacccggc ctctggaatt ttgcgagaca agagccggcg gagaggatca ctgccccaga 420  
ttcacctgta ccataccat cagcggcaag gtggtggctg ttgccgatgg cgcctctaag 480

aaactggcca gacacaccgc ctgttagcagc gccctgacaa tcctgatcaa caactgcggc 540  
atcagcttc 549

<210> 9  
<211> 168  
<212> PRT  
<213> pestivirus type 1  
<400> 9

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

<210> 10  
<211> 504  
<212> DNA

<213> pestivirus type 1

<400> 10

atggaactga tcaccaacga gctgctgtac aagacctaca agcagaaacc cgtgggcgtc 60  
gaggaacccg tgiatgatca agctggcgac cctctgtttg gcgagagagg cgctgttcac 120  
cctcagagca cactgaagct gccccacaag cggggcgaaa gagatgtgcc taccaacctg 180  
gccagcctgc ctaagagagg cgattgcaga accggcaata gcagaggccc tgtgtccggc 240  
atctacctga aacctggacc actgttctac caggactaca agggacccgt gtaccacaga 300

gcccctctgg aactgtttga agagggcagc atgtgcgaaa ccaccaagcg gatcggaaga 360  
gtgaccggct ctgacggcaa gctgtaccac atctacgtgt gcatcgacgg ctgcatcatc 420  
atcaagagcg ccaccagatc ctaccagcgg gtgttcagat gggcgcacaa cagactggac 480  
tgccctctgt gggtcaccag ctgc 504

<210> 11

<211> 222

<212> PRT

<213> Simian parainfluenza virus 5

<400> 11

Met Asp Pro Thr Asp Leu Ser Phe Ser Pro Asp Glu Ile Asn Lys Leu  
1 5 10 15

Ile Glu Thr Gly Leu Asn Thr Val Glu Tyr Phe Thr Ser Gln Gln Val  
20 25 30

Thr Gly Thr Ser Ser Leu Gly Lys Asn Thr Ile Pro Pro Gly Val Thr  
35 40 45

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

His Gln Lys Gly Ser Val Gly Gly Gly Ala Lys Pro Lys Lys Pro Arg  
65 70 75 80

Pro Lys Ile Ala Ile Val Pro Ala Asp Asp Lys Thr Val Pro Gly Lys

85 90 95  
Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Pro Ser Thr Gln

100 105 110  
Thr Val Leu Asp Leu Ser Gly Lys Thr Leu Pro Ser Gly Ser Tyr Lys

115 120 125

Gly Val Lys Leu Ala Lys Phe Gly Lys Glu Asn Leu Met Thr Arg Phe  
130 135 140

Ile Glu Glu Pro Arg Glu Asn Pro Ile Ala Thr Ser Ser Pro Ile Asp  
145 150 155 160

Phe Lys Arg Gly Arg Asp Thr Gly Gly Phe His Arg Arg Glu Tyr Ser  
165 170 175

Ile Gly Trp Val Gly Asp Glu Val Lys Val Thr Glu Trp Cys Asn Pro  
180 185 190

Ser Cys Ser Pro Ile Thr Ala Ala Ala Arg Arg Phe Glu Cys Thr Cys  
195 200 205

His Gln Cys Pro Val Thr Cys Ser Glu Cys Glu Arg Asp Thr  
210 215 220

<210> 12

<211> 666

<212>

> DNA

<213> Simian parainfluenza virus 5

<400> 12

atggacccta ccgacctgag cttcagcccc gacgagatca acaagctgat cgagacaggc 60  
ctgaacaccg tggaatactt caccagccag caagtgaccg gcacaagcag cctgggcaag 120  
aacacaattc ctccaggcgt gaccggcctg ctgacaaatg ctgccgaggc caagatccaa 180  
gagagcacca accaccagaa gggctctgtt ggaggcggag ccaagcctaa gaagcccaga 240  
cctaagatcg ccatcgtgcc cgccgacgat aagacagtgc ctggcaagcc cattccta 300  
cctctgctgg gcctcgacag caccctagc acacagacag tgctggatct gagcggcaag 360

acactgccta gcggcagcta taagggcgtg aagctggcca agttcggcaa agaaaacctg 420  
atgacccggt tcacgagga acccagagag aaccctatcg ccaccagctc tcccatcgac 480  
ttcaagagag gcagagacac cggcggcttc cacagaagag agtacagcat tggctgggtc 540  
ggagatgaag tgaaagtgc cgagtgggtgc aaccccagct gcagccctat tacagccgcc 600  
gctagaagat tcgagtgcac ctgtcaccag tgtcctgtga cctgtagcga gtgcgagcgg 660  
gacaca 666

<210> 13

<211> 230

<212> PRT

<213> Influenza virus

<400> 13

```

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

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

          130            135            140
Thr Leu Leu Arg Ala Phe Thr Glu Glu Gly Ala Ile Val Gly Glu Ile
  145            150            155            160
Ser Pro Leu Pro Ser Leu Pro Gly His Thr Asn Glu Asp Val Lys Asn
          165            170            175
Ala Ile Gly Val Leu Ile Gly Gly Leu Glu Trp Asn Asp Asn Thr Val
          180            185            190
Arg Val Ser Glu Thr Leu Gln Arg Phe Ala Trp Arg Ser Ser Asn Glu
          195            200            205

Asn Gly Gly Pro Pro Leu Thr Pro Thr Gln Lys Arg Lys Met Ala Gly
  210            215            220

```

Lys Ile Arg Ser Glu Val

225 230

<210> 14

<211> 693

<212> DNA

<213> Influenza virus

<400> 14

atggattcca acactgtgtc aagctttcag gtagattgct tcctttggca tgtccgcaaa 60

caagttgcag accaagagct aggtgatgcc ccattccttg atcggttcg ccgagatcag 120

aagtccttaa agggaagagg cagcactctc ggtctgaaca tcgaaacagc cacctgtgtt 180

ggaaagcaaa tagtagagag gattctgaag gaagaatccg atgaggcatt tagaatgacc 240

atggcctccg cacttgcttc gcgataccta actgacatga ctattgaaga gatgtcaagg 300

gactggttca tgctcatgcc caagcagaaa gtggcaggcc ctctttgtgt cagaatggac 360

caggcgataa tggataagaa catcatactg aaagcgaatt tcagtgtgat ttttgaccgg 420

ttggagactc tgacattact aagggtcttc accgaagagg gagcaattgt tggcgaaatt 480

tcaccattgc ctctcttcc aggacatact aatgaggatg tcaaaaatgc aattggggtc 540

ctcatcgagg gacttgaatg gaatgataac acagttcgag tctctgaaac tctacagaga 600

ttcgcttga gaagcagtaa tgagaatggg ggacctccac tcaactcaac acagaaacgg 660

aaaatggcgg gaaaaattag gtcagaagtt tga 693

<210> 15

<211> 109

<212> PRT

<213> Middle East respiratory syndrome related coronavirus

<400> 15

Met Asp Tyr Val Ser Leu Leu Asn Gln Ile Trp Gln Lys Tyr Leu Asn

1 5 10 15

Ser Pro Tyr Thr Thr Cys Leu Tyr Ile Pro Lys Pro Thr Ala Lys Tyr

20 25 30

Thr Pro Leu Val Gly Thr Ser Leu His Pro Val Leu Trp Asn Cys Gln

35 40 45

Leu Ser Phe Ala Gly Tyr Thr Glu Ser Ala Val Asn Ser Thr Lys Ala

50 55 60

Leu Ala Lys Gln Asp Ala Ala Gln Arg Ile Ala Trp Leu Leu His Lys  
65 70 75 80  
Asp Gly Gly Ile Pro Asp Gly Cys Ser Leu Tyr Leu Arg His Ser Ser  
85 90 95  
Leu Phe Ala Gln Ser Glu Glu Glu Glu Ser Phe Ser Asn  
100 105

<210> 16  
<211> 327  
<212> DNA  
<213> Middle East respiratory syndrome related coronavirus  
<400> 16

atggactacg tgtccctgct gaaccagatt tggcagaagt acctgaacag ccctacacc 60  
acctgtctgt acatcccaa gcctaccgcc aagtacacac ctctcgtggg cacatctctg 120  
caccctgtgc tgtggaattg ccagctgagc ttgccggct acaccgagtc tgccgtgaac 180  
agcacaaagg ccctggccaa acaggacgcc gctcagagaa ttgcctggct gctgcacaag 240  
gatggcggca tcctgatgg ctgtagcctg tacctgagac acagcagcct gttcgcccag 300

agcgaggaag aggaatcctt cagcaac 327

<210> 17  
<211> 381  
<212> PRT  
<213> Langat virus  
<400> 17

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

Lys Val Arg Ser Asn Ala Ala Leu Gly Ala Trp Ser Asp Glu Gln Asn  
50 55 60  
Lys Trp Lys Ser Ala Arg Glu Ala Val Glu Asp Pro Glu Phe Trp Ser  
65 70 75 80

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

|                                                                   |              |     |
|-------------------------------------------------------------------|--------------|-----|
| 325                                                               | 330          | 335 |
| Asp Thr Gly Glu Trp Glu Pro Ser Thr Gly Phe Ala Ser Trp Glu Glu   |              |     |
| 340                                                               | 345          | 350 |
| Val Pro Phe Cys Ser His His Phe His Glu Leu Val Met Lys Asp Gly   |              |     |
| 355                                                               | 360          | 365 |
| Arg Ala Leu Val Val Pro Cys Arg Asp Gln Asp Glu Leu               |              |     |
| 370                                                               | 375          | 380 |
| <210>                                                             | 18           |     |
| <211>                                                             | 1143         |     |
| <212>                                                             | DNA          |     |
| <213>                                                             | Langat virus |     |
| <400>                                                             | 18           |     |
| gtgttcaagg acaaggtgga caccaaggct caagagcctc agcctggcac caagatcatc | 60           |     |
| atgagagccg tgaacgactg gctgctggaa cggctggtea agaaaagcag accccggatg | 120          |     |
|                                                                   |              |     |
| tgcagccggg aagagtttat cgccaaagtg cggagcaatg ccgctctcgg agcttggagt | 180          |     |
| gacgagcaga acaagtggaa gtccgccaga gaagccgtgg aagatcccga gttttggagc | 240          |     |
| ctggtggaag ccgagagaga gaggcacatg cagggaagat gtgcccactg cgtgtacaac | 300          |     |
| atgatgggca agagagagaa gaagctgggc gatttcggag tggccaaagg cagcagagcc | 360          |     |
| atctggtata tgtggctggg cagccgcttc ctggaatttg aggccctggg ctctctgaac | 420          |     |
| gaggatcact gggctagcag agcctcttct ggtgctggcg tggaaggcat cagcctgaat | 480          |     |
| tatctcggtt ggcacctgaa gaaactggcc tctctgtctg gcggcctgtt ctacgccgat | 540          |     |
|                                                                   |              |     |
| gatacagccg gatgggacac aaagatcacc aacgccgacc tggacgacga ggaacagatc | 600          |     |
| ctgagatata tggacggcga ccacaaaag ctggccgcca ccgtgctgag aaaggcctat  | 660          |     |
| cacgccaagg tcgtcagagt ggccagacct agtagagaag gcggctgcgt gatggacatc | 720          |     |
| atcaccagaa gggaccagcg cggctctggc caggttgtga catagcctt gaacaccatc  | 780          |     |
| accaacatca aggtgcagct cgtgcggatg atggaaggcg agggcgtgat cgaagtggcc | 840          |     |
| gacagccata atctcggtt gctgagagtg gaaaagtggc tggaagaaca cggcgaagaa  | 900          |     |
| cggctgagca gaatgctggt gtccggcgac gattgtgttg tgcggcccgt ggacgacaga | 960          |     |
|                                                                   |              |     |
| ttcagcaagg cctgtactt tctgaatgac atggccaaga ccagaaagga caccggcgag  | 1020         |     |
| tgggagcctt ctacagctt tgccagctgg gaagaagtgc ctttctgcag ccaccacttc  | 1080         |     |
| cacgagctgg tcatgaagga tggcagagcc ctggtgtgtc cctgcagaga tcaggacgaa | 1140         |     |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ctg                                                                | 1143 |
| <210> 19                                                           |      |
| <211> 693                                                          |      |
| <212> RNA                                                          |      |
| <213> Influenza virus                                              |      |
| <400> 19                                                           |      |
| auggauucca acacuguguc aagcuuucag guagauugcu uccuuuggca uguccgcaaa  | 60   |
| caaguugcag accaagagcu aggugaugcc ccauuccuug aucggcuucg ccgagaucag  | 120  |
| aagucccuuaa agggaagagg cagcacucuc ggucugaaca ucgaaacagc caccuguguu | 180  |
| ggaaagcaaa uaguagagag gauucugaag gaagaaucg augaggcauu uagaugacc    | 240  |
| auggccuccg cacuugcuuc gcgauaccua acugacauga cuauugaaga gaugucaagg  | 300  |
| gacugguuca ugcucaugcc caagcagaaa guggcaggcc cucuuugugu cagaauaggac | 360  |
| caggcgauaa uggauaagaa caucauacug aaagcgaaau ucagugugau uuuugaccgg  | 420  |
| uuggagacuc ugacauuacu aagggcuuuc accgaagagg gagcaauugu uggcgaaaau  | 480  |
| ucaccauugc cuucucuucc aggacauacu aaugaggaug ucaaaaaugc aaugggguc   | 540  |
| cucaucgggg gacuugaug gaugauaac acaguucgag ucucugaaac ucuacagaga    | 600  |
| uucgcuugga gaagcaguaa ugagaauagg ggaccuccac ucacuccaac acagaaacgg  | 660  |
| aaaauggcgg gaaaaauag gucagaaguu uga                                | 693  |
| <210> 20                                                           |      |
| <211> 78                                                           |      |
| <212> PRT                                                          |      |
| <213> Severe acute respiratory syndrome coronavirus 2              |      |
| <400> 20                                                           |      |
| Met Met Pro Thr Ile Phe Phe Ala Gly Ile Leu Ile Val Thr Thr Ile    |      |
| 1 5 10 15                                                          |      |
| Val Tyr Leu Thr Ile Val Gln Leu Leu Gln Leu Ser Leu Leu Gln Val    |      |
| 20 25 30                                                           |      |
| Met Ala Gln Gln Val Leu Phe Leu Asn Met Thr Thr Arg Leu Val Val    |      |
| 35 40 45                                                           |      |
| Ile Leu Lys Asn Gly Asn Leu Glu Gln Lys Thr Val Leu Tyr Tyr Thr    |      |
| 50 55 60                                                           |      |
| Val Thr Ser Leu Gln Thr Ile Thr Ser Cys Thr Gln Leu Asn            |      |

|                                                                 |                     |    |
|-----------------------------------------------------------------|---------------------|----|
| 65                                                              | 70                  | 75 |
| <210>                                                           | 21                  |    |
| <211>                                                           | 19                  |    |
| <212>                                                           | PRT                 |    |
| <213>                                                           | porcine teschovirus |    |
| <400>                                                           | 21                  |    |
| Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn |                     |    |

Pro Gly Pro

|                                                                 |                     |
|-----------------------------------------------------------------|---------------------|
| <210>                                                           | 22                  |
| <211>                                                           | 20                  |
| <212>                                                           | PRT                 |
| <213>                                                           | Thosea asigna virus |
| <400>                                                           | 22                  |
| Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser |                     |

1                      5                      10                      15

Asn Pro Gly Pro                      Gly                      Gly                      Gly

|       |                          |
|-------|--------------------------|
|       | 20                       |
| <210> | 23                       |
| <211> | 22                       |
| <212> | PRT                      |
| <213> | Equine Rhinovirus type 1 |
| <400> | 23                       |

Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val

1 5 10 15

|            |                     |     |     |     |     |
|------------|---------------------|-----|-----|-----|-----|
| Glu        | Ser                 | Asn | Pro | Gly | Pro |
| 20         |                     |     |     |     |     |
| <210>      | 24                  |     |     |     |     |
| <211>      | 4                   |     |     |     |     |
| <212>      | PRT                 |     |     |     |     |
| <213>      | Artificial Sequence |     |     |     |     |
| <220><223> | furin cleavage site |     |     |     |     |

<400> 24  
 Arg Asn Asn Arg  
 1  
 <210> 25  
 <211> 6  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Optimised furin cleavage site  
 <400> 25  
 Arg Arg Arg Arg Arg Arg  
 1 5  
 <210> 26  
 <211> 27  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> furin/T2A  
 <400> 26  
 Arg Arg Arg Arg Arg Arg Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu  
 1 5 10 15  
 Thr Cys Gly Asp Val Glu Glu Asn Pro Gly Pro  
 20 25  
 <210> 27  
 <211> 535  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> nsP1  
 <400> 27  
 Met Glu Lys Val His Val Asp Ile Glu Glu Asp Ser Pro Phe Leu Arg  
 1 5 10 15  
 Ala Leu Gln Arg Ser Phe Pro Gln Phe Glu Val Glu Ala Lys Gln Val  
 20 25 30  
 Thr Asp Asn Asp His Ala Asn Ala Arg Ala Phe Ser His Leu Ala Ser  
 35 40 45

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

Ala Thr Met His Arg Glu Gly Phe Leu Cys Cys Lys Val Thr Asp Thr  
305 310 315 320

Leu Asn Gly Glu Arg Val Ser Phe Pro Val Cys Thr Tyr Val Pro Ala  
325 330 335

Thr Leu Cys Asp Gln Met Thr Gly Ile Leu Ala Thr Asp Val Ser Ala  
340 345 350

Asp Asp Ala Gln Lys Leu Leu Val Gly Leu Asn Gln Arg Ile Val Val  
355 360 365

Asn Gly Arg Thr Gln Arg Asn Thr Asn Thr Met Lys Asn Tyr Leu Leu  
370 375 380

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

385 390 395 400

Asp Gln Glu Asp Glu Arg Pro Leu Gly Leu Arg Asp Arg Gln Leu Val  
405 410 415

Met Gly Cys Cys Trp Ala Phe Arg Arg His Lys Ile Thr Ser Ile Tyr  
420 425 430

Lys Arg Pro Asp Thr Gln Thr Ile Ile Lys Val Asn Ser Asp Phe His  
435 440 445

Ser Phe Val Leu Pro Arg Ile Gly Ser Asn Thr Leu Glu Ile Gly Leu  
450 455 460

Arg Thr Arg Ile Arg Lys Met Leu Glu Glu His Lys Glu Pro Ser Pro  
465 470 475 480

Leu Ile Thr Ala Glu Asp Val Gln Glu Ala Lys Cys Ala Ala Asp Glu  
485 490 495

Ala Lys Glu Val Arg Glu Ala Glu Glu Leu Arg Ala Ala Leu Pro Pro  
500 505 510

Leu Ala Ala Asp Val Glu Glu Pro Thr Leu Glu Ala Asp Val Asp Leu  
515 520 525

Met Leu Gln Glu Ala Gly Ala

530 535

<210> 28

<211> 1605

<212> DNA

<213> Artificial Sequence

<220><223> nsP1 coding sequence

<400> 28

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggagaaag ttcacgttga catcgaggaa gacagcccat tctcagagc ttgcagcgg    | 60   |
| agcttcccgc agtttgaggt agaagccaag caggctactg ataatgacca tgctaagcc   | 120  |
| agagcgtttt cgcacttgcc ttcaaaactg atcgaaacgg aggtggaccc atccgacacg  | 180  |
| atccttgaca ttggaagtgc gcccggccgc agaatgtatt ctaagcacia gtatcattgt  | 240  |
| atctgtccga tgagatgtgc ggaagatccg gacagattgt ataagtatgc aactaagctg  | 300  |
|                                                                    |      |
| aagaaaaact gtaaggaaat aactgataag gaattggaca agaaaatgaa ggagctggcc  | 360  |
| gccgtcatga gcgaccctga cctggaaaact gagactatgt gcctccacga cgacgagtcg | 420  |
| tgctgctacg aagggaagt cgctgtttac caggatgtat acgcggttga cggaccgaca   | 480  |
| agtctctatc accaagccaa taaggaggtt agagtcgcct actggatagg ctttgacacc  | 540  |
| acccctttta tgtttaagaa cttggctgga gcataatccat catactctac caactgggcc | 600  |
| gacgaaacgg tgtaaacggc tcgtaacata ggctatgca gctctgacgt tatggagcgg   | 660  |
| tcacgtagag ggatgtccat tcttagaaag aagtatttga aaccatccaa caatgttcta  | 720  |
|                                                                    |      |
| ttctctgttg gctcgacat ctaccacgag aagagggact tactgaggag ctggcacctg   | 780  |
| ccgtctgtat ttcacttacg tggcaagcaa aattacacat gtcggtgtga gactatagtt  | 840  |
| agttgcgacg ggtacgtcgt taaaagaata gctatcagtc caggcctgta tgggaagcct  | 900  |
| tcaggctatg ctgctacgat gcaccgcgag ggattcttgt gctgcaaagt gacagacaca  | 960  |
| ttgaacgggg agagggtctc ttttcccgtg tgcacgtatg tgccagctac attgtgtgac  | 1020 |
| caaatgactg gcatactggc aacagatgtc agtgcggacg acgcgcaaaa actgctgggt  | 1080 |
| gggctcaacc agcgtatagt cgtcaacggt cgcaccaga gaaacaccaa taccatgaaa   | 1140 |
|                                                                    |      |
| aattaccttt tgcccgtagt ggcccaggca tttgctaggt gggcaaagga atataaggaa  | 1200 |
| gatcaagaag atgaaaggcc actaggacta cgagatagac agttagtcac ggggtgttgt  | 1260 |
| tgggctttta gaaggcacia gataacatct atttataagc gcccggatag ccaaaccatc  | 1320 |
| atcaaagtga acagcgattt ccactcattc gtgctgcca ggataggcag taacacattg   | 1380 |
| gagatcgggc tgagaacaag aatcaggaaa atgttagagg agcacaagga gccgtcacct  | 1440 |
| ctcattaccg ccgaggacgt acaagaagct aagtgcgcag ccgatgaggc taaggaggtg  | 1500 |
| cgtaagccg aggagttgcg cgcagctcta ccacctttgg cagctgatgt tgaggagccc   | 1560 |

actctggaag ccgatgtcga cttgatgtta caagaggctg gggcc

1605

<210> 29

<211> 794

<212> PRT

<213> Artificial Sequence

<220><223> nsP2

<400> 29

Gly Ser Val Glu Thr Pro Arg Gly Leu Ile Lys Val Thr Ser Tyr Asp

1 5 10 15

Gly Glu Asp Lys Ile Gly Ser Tyr Ala Val Leu Ser Pro Gln Ala Val

20 25 30

Leu Lys Ser Glu Lys Leu Ser Cys Ile His Pro Leu Ala Glu Gln Val

35 40 45

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

50 55 60

Tyr His Gly Lys Val Val Val Pro Glu Gly His Ala Ile Pro Val Gln

65 70 75 80

Asp Phe Gln Ala Leu Ser Glu Ser Ala Thr Ile Val Tyr Asn Glu Arg

85 90 95

Glu Phe Val Asn Arg Tyr Leu His His Ile Ala Thr His Gly Gly Ala

100 105 110

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

115 120 125

Asp Gly Glu Tyr Leu Tyr Asp Ile Asp Arg Lys Gln Cys Val Lys Lys

130 135 140

Glu Leu Val Thr Gly Leu Gly Leu Thr Gly Glu Leu Val Asp Pro Pro

145 150 155 160

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

165 170 175

Tyr Gln Val Pro Thr Ile Gly Val Tyr Gly Val Pro Gly Ser Gly Lys

180 185 190

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

195                      200                      205  
 Ala Lys Lys Glu Asn Cys Ala Glu Ile Ile Arg Asp Val Lys Lys Met  
 210                      215                      220  
 Lys Gly Leu Asp Val Asn Ala Arg Thr Val Asp Ser Val Leu Leu Asn  
 225                      230                      235                      240  
 Gly Cys Lys His Pro Val Glu Thr Leu Tyr Ile Asp Glu Ala Phe Ala  
 245                      250                      255  
 Cys His Ala Gly Thr Leu Arg Ala Leu Ile Ala Ile Ile Arg Pro Lys  
  
 260                      265                      270  
 Lys Ala Val Leu Cys Gly Asp Pro Lys Gln Cys Gly Phe Phe Asn Met  
 275                      280                      285  
 Met Cys Leu Lys Val His Phe Asn His Glu Ile Cys Thr Gln Val Phe  
 290                      295                      300  
 His Lys Ser Ile Ser Arg Arg Cys Thr Lys Ser Val Thr Ser Val Val  
 305                      310                      315                      320  
 Ser Thr Leu Phe Tyr Asp Lys Lys Met Arg Thr Thr Asn Pro Lys Glu  
 325                      330                      335  
  
 Thr Lys Ile Val Ile Asp Thr Thr Gly Ser Thr Lys Pro Lys Gln Asp  
 340                      345                      350  
 Asp Leu Ile Leu Thr Cys Phe Arg Gly Trp Val Lys Gln Leu Gln Ile  
 355                      360                      365  
 Asp Tyr Lys Gly Asn Glu Ile Met Thr Ala Ala Ala Ser Gln Gly Leu  
 370                      375                      380  
 Thr Arg Lys Gly Val Tyr Ala Val Arg Tyr Lys Val Asn Glu Asn Pro  
 385                      390                      395                      400  
 Leu Tyr Ala Pro Thr Ser Glu His Val Asn Val Leu Leu Thr Arg Thr  
  
 405                      410                      415  
 Glu Asp Arg Ile Val Trp Lys Thr Leu Ala Gly Asp Pro Trp Ile Lys  
 420                      425                      430  
 Thr Leu Thr Ala Lys Tyr Pro Gly Asn Phe Thr Ala Thr Ile Glu Glu  
 435                      440                      445  
 Trp Gln Ala Glu His Asp Ala Ile Met Arg His Ile Leu Glu Arg Pro

450                      455                      460  
 Asp Pro Thr Asp Val Phe Gln Asn Lys Ala Asn Val Cys Trp Ala Lys  
 465                      470                      475                      480  
  
 Ala Leu Val Pro Val Leu Lys Thr Ala Gly Ile Asp Met Thr Thr Glu  
                                  485                      490                      495  
 Gln Trp Asn Thr Val Asp Tyr Phe Glu Thr Asp Lys Ala His Ser Ala  
                                  500                      505                      510  
 Glu Ile Val Leu Asn Gln Leu Cys Val Arg Phe Phe Gly Leu Asp Leu  
                                  515                      520                      525  
 Asp Ser Gly Leu Phe Ser Ala Pro Thr Val Pro Leu Ser Ile Arg Asn  
                                  530                      535                      540  
 Asn His Trp Asp Asn Ser Pro Ser Pro Asn Met Tyr Gly Leu Asn Lys  
  
 545                      550                      555                      560  
 Glu Val Val Arg Gln Leu Ser Arg Arg Tyr Pro Gln Leu Pro Arg Ala  
                                  565                      570                      575  
 Val Ala Thr Gly Arg Val Tyr Asp Met Asn Thr Gly Thr Leu Arg Asn  
                                  580                      585                      590  
 Tyr Asp Pro Arg Ile Asn Leu Val Pro Val Asn Arg Arg Leu Pro His  
                                  595                      600                      605  
 Ala Leu Val Leu His His Asn Glu His Pro Gln Ser Asp Phe Ser Ser  
                                  610                      615                      620  
  
 Phe Val Ser Lys Leu Lys Gly Arg Thr Val Leu Val Val Gly Glu Lys  
 625                      630                      635                      640  
 Leu Ser Val Pro Gly Lys Met Val Asp Trp Leu Ser Asp Arg Pro Glu  
                                  645                      650                      655  
 Ala Thr Phe Arg Ala Arg Leu Asp Leu Gly Ile Pro Gly Asp Val Pro  
                                  660                      665                      670  
 Lys Tyr Asp Ile Ile Phe Val Asn Val Arg Thr Pro Tyr Lys Tyr His  
                                  675                      680                      685  
 His Tyr Gln Gln Cys Glu Asp His Ala Ile Lys Leu Ser Met Leu Thr  
  
 690                      695                      700

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

Arg Thr His Asn Ser Tyr Lys Leu Ser Ser Thr Leu Thr Asn Ile Tyr  
 770 775 780

Thr Gly Ser Arg Leu His Glu Ala Gly Cys

785 790

<210> 30

<211> 2382

<212> DNA

<213> Artificial Sequence

<220><223> nsP2 coding sequence

<400> 30

ggctcagtgg agacacctcg tggcttgata aaggttacca gctacgatgg cgaggacaag 60

atcggtctctt acgtgtgtct ttctccgcag gctgtactca agagtgaata attatcttgc 120

atccaccctc tcgtgaaca agtcatagt ataacacact ctggccgaaa agggcggttat 180

gccgtggaac cataccatgg taaagtagtg gtgccagagg gacatgcaat acccgtccag 240

gactttcaag ctctgagtga aagtgccacc attgtgtaca acgaacgtga gticgtaaac 300

aggtacctgc accatattgc cacacatgga ggagcgtga acactgatga agaataattac 360

aaaactgtca agcccagcga gcacgacggc gaatacctgt acgacatcga caggaaacag 420

tgcgtcaaga aagaactagt cactgggcta gggtcacag gcgagctggt ggatcctccc 480

ttccatgaat tcgcctacga gagtctgaga acacgaccag ccgtcctta ccaagtacca 540

accatagggg tgtatggcgt gccaggatca ggcaagtctg gcatcattaa aagcgagtc 600

accaaaaag atctagtggg gagcgccaag aaagaaaact gtgcagaaat tataagggac 660

gtcaagaaaa tgaaagggtt ggagctcaat gccagaactg tggactcagt gctcttgaat 720

ggatgcaaac accccgtaga gacctgtat attgacgaag cttttgcttg tcatgcaggt 780

actctcagag cgctcatagc cattataaga cctaaaaagg cagtgtcttg cggggatccc 840

aaacagtgcg gtttttttaa catgatgtgc ctgaaagtgc attttaacca cgagatttgc 900  
 acacaagtct tccacaaaag catctctcgc cgttgacta aatctgtgac ttcggctcgc 960  
 tcaaccttgt tttagacaaa aaaaatgaga acgacgaatc cgaaagagac taagattgtg 1020  
  
 attgacacta ccggcagtac caaacctaag caggacgac tcattctcac ttgtttcaga 1080  
 ggggtgggtga agcagtttga aatagattac aaaggcaacg aaataatgac ggcagctgcc 1140  
 tctcaagggc tgaccctgaa aggtgtgtat gccgttcggt acaaggtgaa tgaaaatcct 1200  
 ctgtacgcac ccacctcaga acatgtgaac gtctactga cccgcacgga ggaccgcatc 1260  
 gtgtggaaaa cactagccgg cgacccatgg ataaaaacac tgactgcaa gtacctggg 1320  
 aatttactg ccacgataga ggagtggcaa gcagagcatg atgccatcat gaggcacatc 1380  
 ttggagagac cggaccctac cgacgtcttc cagaataagg caaacgtgtg ttgggccaag 1440  
  
 gctttagtgc cgggtgtgaa gaccgtggc atagacatga ccaactgaaca atggaacact 1500  
 gtggattatt ttgaaacgga caaagctcac tcagcagaga tagtattgaa ccaactatgc 1560  
 gtgaggttct ttggactcga tctggactcc ggtctatttt ctgcaccac tgttcgtta 1620  
 tccattagga ataactctg ggataactcc ccgtgccta acatgtacgg gctgaataaa 1680  
 gaagtggctc gtcagctctc tcgcaggtac ccacaactgc ctccggcagt tgccactgga 1740  
 agagtctatg acatgaacac tggtagactg cgcaattatg atccgcgcac aaacctagta 1800  
 cctgtaaaca gaagactgcc tcatgcttta gtctccacc ataataaaca cccacagagt 1860  
  
 gacttttctt cattcgtcag caaattgaag ggcagaactg tcctggtggt cggggaaaag 1920  
 ttgtccgtcc caggcaaaat ggttgactgg ttgtcagacc ggctgaggc taccttcaga 1980  
 gctcggctgg atttaggcat ccaggtgat gtgccccaat atgacataat atttgtaat 2040  
 gtgaggacc catataaata ccatcactat cagcagtgat aagaccatgc cattaagctt 2100  
 agcatgttga ccaagaaagc ttgtctgcat ctgaatccc gcggaacctg tgtcagcata 2160  
 ggttatggtt acgtgacag gccagcgaa agcatcattg gtgctatagc gcggcagttc 2220  
 aagttttccc gggtatgcaa accgaaatcc tcaattgaag agacggaagt tctgtttgta 2280  
  
 ttcatgggt acgatcgaa ggcccgtacg cacaattctt acaagctttc atcaaccttg 2340  
 accaacaattt atacaggttc cagactccac gaagccgat gt 2382  
  
 <210> 31  
 <211> 556  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> nsP3

<400> 31

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

1 5 10 15

Gly Val Ile Ile Asn Ala Ala Asn Ser Lys Gly Gln Pro Gly Gly Gly

20 25 30

Val Cys Gly Ala Leu Tyr Lys Lys Phe Pro Glu Ser Phe Asp Leu Gln

35 40 45

Pro Ile Glu Val Gly Lys Ala Arg Leu Val Lys Gly Ala Ala Lys His

50 55 60

Ile Ile His Ala Val Gly Pro Asn Phe Asn Lys Val Ser Glu Val Glu

65 70 75 80

Gly Asp Lys Gln Leu Ala Glu Ala Tyr Glu Ser Ile Ala Lys Ile Val

85 90 95

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

100 105 110

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

115 120 125

Leu Thr Ala Leu Asp Thr Thr Asp Ala Asp Val Ala Ile Tyr Cys Arg

130 135 140

Asp Lys Lys Trp Glu Met Thr Leu Lys Glu Ala Val Ala Arg Arg Glu

145 150 155 160

Ala Val Glu Glu Ile Cys Ile Ser Asp Asp Ser Ser Val Thr Glu Pro

165 170 175

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

180 185 190

Lys Gly Tyr Ser Thr Ser Asp Gly Lys Thr Phe Ser Tyr Leu Glu Gly

195 200 205

Thr Lys Phe His Gln Ala Ala Lys Asp Ile Ala Glu Ile Asn Ala Met

210 215 220

Trp Pro Val Ala Thr Glu Ala Asn Glu Gln Val Cys Met Tyr Ile Leu

225 230 235 240

Gly Glu Ser Met Ser Ser Ile Arg Ser Lys Cys Pro Val Glu Glu Ser

245 250 255

Glu Ala Ser Thr Pro Pro Ser Thr Leu Pro Cys Leu Cys Ile His Ala

260 265 270

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

275 280 285

Ile Thr Val Cys Ser Ser Phe Pro Leu Pro Lys Tyr Arg Ile Thr Gly

290 295 300

Val Gln Lys Ile Gln Cys Ser Gln Pro Ile Leu Phe Ser Pro Lys Val

305 310 315 320

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

325 330 335

Asp Glu Thr Pro Glu Pro Ser Ala Glu Asn Gln Ser Thr Glu Gly Thr

340 345 350

Pro Glu Gln Pro Pro Leu Ile Thr Glu Asp Glu Thr Arg Thr Arg Thr

355 360 365

Pro Glu Pro Ile Ile Ile Glu Glu Glu Glu Glu Asp Ser Ile Ser Leu

370 375 380

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

385 390 395 400

His Gly Pro Pro Ser Val Ser Ser Ser Ser Trp Ser Ile Pro His Ala

405 410 415

Ser Asp Phe Asp Val Asp Ser Leu Ser Ile Leu Asp Thr Leu Glu Gly

420 425 430

Ala Ser Val Thr Ser Gly Ala Thr Ser Ala Glu Thr Asn Ser Tyr Phe

435 440 445

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

450 455 460

Val Phe Arg Asn Pro Pro His Pro Ala Pro Arg Thr Arg Thr Pro Ser

465 470 475 480

Leu Ala Pro Ser Arg Ala Cys Ser Arg Thr Ser Leu Val Ser Thr Pro

485 490 495  
 Pro Gly Val Asn Arg Val Ile Thr Arg Glu Glu Leu Glu Ala Leu Thr  
 500 505 510  
 Pro Ser Arg Thr Pro Ser Arg Ser Val Ser Arg Thr Ser Leu Val Ser  
 515 520 525  
 Asn Pro Pro Gly Val Asn Arg Val Ile Thr Arg Glu Glu Phe Glu Ala

530 535 540  
 Phe Val Ala Gln Gln Gln Arg Phe Asp Ala Gly Ala

545 550 555

<210> 32

<211> 1671

<212> DNA

<213> Artificial Sequence

<220><223> nsP3 coding sequence

<400> 32

gcacctcat atcatgtgtt gcgaggggat attgccacgg ccaccgaagg agtgattata 60  
 aatgtctgcta acagcaaagg acaacctggc ggaggggtgt gcggagcgct gtataagaaa 120  
 ttcccggaaa gcttcgattt acagccgatc gaagtaggaa aagcgcgact ggtcaaaggt 180

gcagctaaac atatcattca tgccgtagga ccaaacttca acaaagtttc ggaggttgaa 240  
 ggtgacaaac agttggcaga ggcttatgag tccatcgcta agattgtcaa cgataacaat 300  
 tacaagtcag tagcgattcc actgttgtcc accggcatct tttccgggaa caaagatcga 360  
 ctaacccaat cattgaacca tttgttgaca gctttagaca ccactgatgc agatgtagcc 420  
 atatactgca gggacaagaa atgggaaatg actctcaagg aagcagtggc taggagagaa 480  
 gcagtggagg agatatgcat atccgacgac tcttcagtga cagaacctga tgcagagctg 540  
 gtgaggggtgc atccgaagag ttctttggct ggaaggaagg gctacagcac aagcgatggc 600

aaaactttct catatttga agggaccaag ttaccacagg cggccaagga tatagcagaa 660  
 attaatgcca tgtggccgtg tgcaacggag gccaatgagc aggtatgcat gtatatcctc 720  
 ggagaaaagca tgagcagtat taggtcgaaa tgccccgtcg aagagtcgga agcctccaca 780  
 ccacctagca cgctgccttg cttgtgcatc catgccatga ctccagaaag agtacagcgc 840  
 ctaaaagcct cacgtccaga acaattact gtgtgctcat cctttccatt gccgaagtat 900  
 agaatactg gtgtgcagaa gatccaatgc tcccagccta tattgttctc accgaaagtg 960

cctgcgtata ttcatccaag gaagtatctc gtggaaacac caccggtaga cgagactccg 1020

gagccatcgg cagagaacca atccacagag gggacacctg aacaaccacc acttataacc 1080

gaggatgaga ccaggactag aacgcctgag ccgatcatca tcgaagagga agaagaggat 1140

agcataagtt tgctgtcaga tggcccgacc caccaggtgc tgcaagtcga ggcagacatt 1200

cacgggccgc cctctgtatc tagtcatcc tggatcattc ctcattgcatc cgactttgat 1260

gtggacagtt tatccatact tgacacctg gagggagcta gcgtgaccag cggggcaacg 1320

tcagccgaga ctaactctta cttcgcaaag agtatggagt ttctggcgcg accggtgcct 1380

gcgcctcgaa cagtattcag gaacctcca catcccgtc cgcgcacaag aacaccgtca 1440

cttgaccca gcagggcctg ctcgagaacc agcctagttt ccaccccgcc aggcgtgaat 1500

agggatgatca ctagagagga gctcaggcg cttaccccg caccgactcc tagcaggtcg 1560

gtctcgagaa ccagcctggt ctccaaccg ccaggcgtaa atagggtgat tacaagagag 1620

gagtttgagg cgttcgtac acaacaaca tgacggtttg atgcgggtgc a 1671

<210> 33

<211> 607

<212> PRT

<213> Artificial Sequence

<220><223> nsP4

<400> 33

Tyr Ile Phe Ser Ser Asp Thr Gly Gln Gly His Leu Gln Gln Lys Ser

1 5 10 15

Val Arg Gln Thr Val Leu Ser Glu Val Val Leu Glu Arg Thr Glu Leu

20 25 30

Glu Ile Ser Tyr Ala Pro Arg Leu Asp Gln Glu Lys Glu Glu Leu Leu

35 40 45

Arg Lys Lys Leu Gln Leu Asn Pro Thr Pro Ala Asn Arg Ser Arg Tyr

50 55 60

Gln Ser Arg Lys Val Glu Asn Met Lys Ala Ile Thr Ala Arg Arg Ile

65 70 75 80

Leu Gln Gly Leu Gly His Tyr Leu Lys Ala Glu Gly Lys Val Glu Cys

85 90 95

Tyr Arg Thr Leu His Pro Val Pro Leu Tyr Ser Ser Ser Val Asn Arg

100 105 110  
 Ala Phe Ser Ser Pro Lys Val Ala Val Glu Ala Cys Asn Ala Met Leu  
 115 120 125  
 Lys Glu Asn Phe Pro Thr Val Ala Ser Tyr Cys Ile Ile Pro Glu Tyr  
 130 135 140  
 Asp Ala Tyr Leu Asp Met Val Asp Gly Ala Ser Cys Cys Leu Asp Thr  
 145 150 155 160  
  
 Ala Ser Phe Cys Pro Ala Lys Leu Arg Ser Phe Pro Lys Lys His Ser  
 165 170 175  
 Tyr Leu Glu Pro Thr Ile Arg Ser Ala Val Pro Ser Ala Ile Gln Asn  
 180 185 190  
 Thr Leu Gln Asn Val Leu Ala Ala Ala Thr Lys Arg Asn Cys Asn Val  
 195 200 205  
 Thr Gln Met Arg Glu Leu Pro Val Leu Asp Ser Ala Ala Phe Asn Val  
 210 215 220  
 Glu Cys Phe Lys Lys Tyr Ala Cys Asn Asn Glu Tyr Trp Glu Thr Phe  
  
 225 230 235 240  
 Lys Glu Asn Pro Ile Arg Leu Thr Glu Glu Asn Val Val Asn Tyr Ile  
 245 250 255  
 Thr Lys Leu Lys Gly Pro Lys Ala Ala Ala Leu Phe Ala Lys Thr His  
 260 265 270  
 Asn Leu Asn Met Leu Gln Asp Ile Pro Met Asp Arg Phe Val Met Asp  
 275 280 285  
 Leu Lys Arg Asp Val Lys Val Thr Pro Gly Thr Lys His Thr Glu Glu  
 290 295 300  
  
 Arg Pro Lys Val Gln Val Ile Gln Ala Ala Asp Pro Leu Ala Thr Ala  
 305 310 315 320  
 Tyr Leu Cys Gly Ile His Arg Glu Leu Val Arg Arg Leu Asn Ala Val  
 325 330 335  
 Leu Leu Pro Asn Ile His Thr Leu Phe Asp Met Ser Ala Glu Asp Phe  
 340 345 350  
 Asp Ala Ile Ile Ala Glu His Phe Gln Pro Gly Asp Cys Val Leu Glu

355                                      360                                      365  
 Thr Asp Ile Ala Ser Phe Asp Lys Ser Glu Asp Asp Ala Met Ala Leu

370                                      375                                      380  
 Thr Ala Leu Met Ile Leu Glu Asp Leu Gly Val Asp Ala Glu Leu Leu  
 385                                      390                                      395                                      400

Thr Leu Ile Glu Ala Ala Phe Gly Glu Ile Ser Ser Ile His Leu Pro  
 405                                      410                                      415

Thr Lys Thr Lys Phe Lys Phe Gly Ala Met Met Lys Ser Gly Met Phe  
 420                                      425                                      430

Leu Thr Leu Phe Val Asn Thr Val Ile Asn Ile Val Ile Ala Ser Arg  
 435                                      440                                      445

Val Leu Arg Glu Arg Leu Thr Gly Ser Pro Cys Ala Ala Phe Ile Gly  
 450                                      455                                      460

Asp Asp Asn Ile Val Lys Gly Val Lys Ser Asp Lys Leu Met Ala Asp  
 465                                      470                                      475                                      480

Arg Cys Ala Thr Trp Leu Asn Met Glu Val Lys Ile Ile Asp Ala Val  
 485                                      490                                      495

Val Gly Glu Lys Ala Pro Tyr Phe Cys Gly Gly Phe Ile Leu Cys Asp  
 500                                      505                                      510

Ser Val Thr Gly Thr Ala Cys Arg Val Ala Asp Pro Leu Lys Arg Leu

515                                      520                                      525  
 Phe Lys Leu Gly Lys Pro Leu Ala Ala Asp Asp Glu His Asp Asp Asp

530                                      535                                      540  
 Arg Arg Arg Ala Leu His Glu Glu Ser Thr Arg Trp Asn Arg Val Gly

545                                      550                                      555                                      560  
 Ile Leu Ser Glu Leu Cys Lys Ala Val Glu Ser Arg Tyr Glu Thr Val

565                                      570                                      575  
 Gly Thr Ser Ile Ile Val Met Ala Met Thr Thr Leu Ala Ser Ser Val

580                                      585                                      590

Lys Ser Phe Ser Tyr Leu Arg Gly Ala Pro Ile Thr Leu Tyr Gly

595                                      600                                      605

<210> 34  
 <211> 1821  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nsP4 encoding sequence

<400> 34

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tacatctttt cctccgacac cgggtcaaggg catttacaac aaaaatcagt aaggcaaacg | 60   |
| gtgctatccg aagtgggtgtt ggagaggacc gaattggaga ttctgtatgc cccgcgcctc | 120  |
| gaccaagaaa aagaagaatt actacgcaag aaattacagt taaatccac acctgctaac   | 180  |
| agaagcagat accagtccag gaaggtggag aacatgaaag ccataacagc tagacgtatt  | 240  |
| ctgcaaggcc tagggcatta tttgaaggca gaaggaaaag tggagtgcta ccgaaccctg  | 300  |
| catcctgttc ctttgtattc atctagtgtg aaccgtgcct tttcaagccc caaggtcgca  | 360  |
| gtggaagcct gtaacgcat gttgaaagag aactttccga ctgtggcttc ttactgtatt   | 420  |
| attccagagt acgatgccta tttggacatg gttgacggag cttcatgctg cttagacact  | 480  |
| gccagttttt gccctgcaaa gctgcgcagc tttccaaaga aacactccta tttggaaccc  | 540  |
| acaatacgat cggcagtgcc ttcagcgatc cagaacacgc tccagaacgt cctggcagct  | 600  |
| gccacaaaaa gaaattgcaa tgtcacgcaa atgagagaat tgcccgtatt ggattcggcg  | 660  |
| gcctttaatg tggaatgctt caagaaatat gcgtgtaata atgaatattg ggaaacgttt  | 720  |
| aaagaaaacc ccatcaggct tactgaagaa aacgtggtaa attacattac caaattaaaa  | 780  |
| ggacccaaaag ctgctgctct ttttgcaag acacataatt tgaatatgtt gcaggacata  | 840  |
| ccaatggaca ggittgtaat ggacttaaag agagacgtga aagtactcc aggaacaaaa   | 900  |
| catactgaag aacggcccaa ggtacagtg atccagctg ccgatccgt agcaacagcg     | 960  |
| tatctgtgcg gaatccaccg agagctgggt aggagattaa atgcggctct gcttccgaac  | 1020 |
| attcatacac tgittgatat gtcggtgaa gactttgacg ctattatagc cgagcacttc   | 1080 |
| cagcctgggg attgtgttct ggaaactgac atcgctcgt ttgataaaag tgaggacgac   | 1140 |
| gccatggctc tgaccgcgtt aatgattctg gaagacttag gtgtggacgc agagctgttg  | 1200 |
| acgtgattg aggcggcttt cggcgaaatt tcatcaatac atttcccac taaaactaaa    | 1260 |
| tttaaatcgc gagccatgat gaaatctgga atgttcctca cactgtttgt gaacacagtc  | 1320 |
| attaacattg taatcgcaag cagagtgttg agagaacggc taaccggatc accatgtgca  | 1380 |
| gcattcattg gagatgacaa tatcgtgaaa ggagtcaaat cggacaaatt aatggcagac  | 1440 |
| aggtgcgcca cctggttgaa tatggaagtc aagattatag atgctgtggt gggcgagaaa  | 1500 |

gcgcccttatt tctgtggagg gtttattttg tgtgactccg tgaccggcac agcgtgccgt 1560  
 gtggcagacc ccctaaaaag gctgtttaag ctgggcaaac ctctggcagc agacgatgaa 1620  
 catgatgatg acaggagaag ggcatgtcat gaagagtcaa cacgctggaa ccgagtgggt 1680  
 attctttcag agctgtgcaa ggcagtagaa tcaaggtatg aaaccgtagg aacttccatc 1740  
 atagttatgg ccatgactac tctagctagc agtggttaaat cattcagcta cctgagaggg 1800  
 gccctataa ctctctacgg c 1821

<210> 35

<211> 9561

<212> DNA

<213> Artificial Sequence

<220><223> recombinant vector

<220><221> misc\_feature

<222> (7614)..(7615)

<223> A GOI is present between positions 7614 and 7615

<400> 35

cgccagcaac gcgagctcta atacgactca ctatagatgg gcggcgcatg agagaagccc 60  
 agaccaatta cctaccctaa atggagaaaag ttcacgttga catcgaggaa gacagcccat 120  
 tcctcagagc ttigcagcgg agcttcccg agtttgaggt agaagccaag caggctactg 180  
 ataatgacca tgctaatacc agagcgtttt cgcactctggc ttcaaaactg atcgaaacgg 240  
 aggtggaccc atccgacacg atccttgaca ttggaagtgc gcccgcccgc agaattgtatt 300

ctaagcacia gtatcattgt atctgtccga tgagatgtgc ggaagatccg gacagattgt 360  
 ataagtatgc aactaagctg aagaaaaact gtaaggaaat aactgataag gaattggaca 420  
 agaaaatgaa ggagctggcc gccgtcatga gcgaccctga cctggaaact gagactatgt 480  
 gcctccagca cgacgagtcg tgctgctacg aagggaagt cgctgtttac caggatgtat 540  
 acgcggttga cggaccgaca agtctctatc accaagccaa taaggagatt agagtgcct 600  
 actggatagg ctttgacacc acccctttta tgtttaagaa cttggctgga gcataatcat 660  
 catactctac caactgggcc gacgaaaccg tgtaaacggc tcgtaacata ggcctatgca 720

gctctgacgt tatggagcgg tcacgtagag ggatgtccat tcttagaaaag aagtatttga 780  
 aaccatcaa caatgttcta ttctctgttg gctcgacat ctaccacgag aagagggact 840  
 tactgaggag ctggcacctg ccgtctgtat ttacttacg tggcaagcaa aattacacat 900  
 gtcggtgtga gactatagtt agttgcgacg ggtacgtcgt taaaagaata gctatcagtc 960  
 caggcctgta tggaagcct tcaggctatg ctgctacgat gcaccgcgag ggattcttgt 1020

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gctgcaaagt gacagacaca ttgaacgggg agagggtctc ttttccgtg tgcacgtatg   | 1080 |
| tgccagctac attgtgtgac caaatgactg gcatactggc aacagatgtc agtgcggacg  | 1140 |
| acgcgcaaaa actgctgggtt gggctcaacc agcgtatagt cgtcaacggt cgcacccaga | 1200 |
| gaaacaccaa taccatgaaa aattaccttt tgcccgtagt ggcccaggca ttgtctaggt  | 1260 |
| gggcaaagga atataaggaa gatcaagaag atgaaaggcc actaggacta cgagatagac  | 1320 |
| agttagtcat ggggtgttgt tgggttttta gaaggcaca gataacatct attataagc    | 1380 |
| gcccggatac ccaaacctac atcaaagtga acagcgattt ccactcattc gtgctgcca   | 1440 |
| ggataggcag taacacattg gagatcgggc tgagaacaag aatcaggaaa atgttagagg  | 1500 |
| agcacaagga gccgtcacct ctcatctaccg ccgaggacgt acaagaagct aagtgcgcag | 1560 |
| ccgatgaggc taaggaggtg cgtgaagccg aggagttgcg cgcagctcta ccacctttgg  | 1620 |
| cagctgatgt tgaggagccc actctggaag ccgatgtcga cttgatgtta caaggagctg  | 1680 |
| gggccggctc agtggagaca cctctgtggt tgataaagggt taccagctac gatggcgagg | 1740 |
| acaagatcgg ctcttacgt gtgctttctc cgcaggctgt actcaagagt gaaaaattat   | 1800 |
| cttgcatcca cctctcgt gaacaagtca tagtgataac aactctggc cgaaggggc      | 1860 |
| gttatgccgt ggaaccatac catggtaaag tagtgggtcc agaggacat gcaatacccg   | 1920 |
| tccaggactt tcaagctctg agtgaaagt ccaccattgt gtacaacgaa cgtgagttcg   | 1980 |
| taaacaggta cctgcacat attgccacac atggaggagc gctgaacact gatgaagaat   | 2040 |
| attacaaaac tgtcaagccc agcgagcacg acggcgaata cctgtacgac atcgacagga  | 2100 |
| aacagtgcgt caagaaagaa ctagtactg ggctagggt cacaggcgag ctggtggatc    | 2160 |
| ctcccttcca tgaattcgcc tacgagagtc tgagaacacg accagccgct cttaccaag   | 2220 |
| taccaacat aggggtgtat ggcgtgccag gatcaggcaa gtctggcatc attaaaagcg   | 2280 |
| cagtcaccaa aaaagatcta gtggtgagcg ccaagaaaga aaactgtgca gaaattataa  | 2340 |
| gggacgtcaa gaaaatgaaa gggctggacg tcaatgccag aactgtggac tcagtgtctt  | 2400 |
| tgaatggatg caaacacccc gtagagacc tgatatatga cgaagctttt gcttgtcatg   | 2460 |
| caggtactct cagagcgctc atagccatta taagacctaa aaaggcagtg ctctgcgggg  | 2520 |
| atcccaaaca gtgcggtttt tttaacatga tgtgcctgaa agtgcatttt aaccacgaga  | 2580 |
| tttgacaca agtcttcac aaaagcatct ctgccgttg cactaaatct gtgacttcgg     | 2640 |
| tcgtctcaac ctgttttac gacaaaaaaa tgagaacgac gaatccgaaa gagactaaga   | 2700 |
| ttgtgattga cactaccggc agtaccaaac ctaagcagga cgatctcatt ctcaattgtt  | 2760 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tcagagggtg ggtgaagcag ttgcaaatag attacaaagg caacgaaata atgacggcag  | 2820 |
| ctgcctctca agggctgacc cgtaaagggtg tgtatgccgt tcggtacaag gtgaatgaaa | 2880 |
| atcctctgta cgcaccacc tcaaacatg tgaacgtcct actgaccgc acggaggacc     | 2940 |
| gcatcgtgtg gaaaacacta gccggcgacc catggataaa aacactgact gccaagtacc  | 3000 |
| ctgggaattt cactgccacg atagaggagt ggcaagcaga gcatgatgcc atcatgaggc  | 3060 |
| acatcttggg gagaccggac cctaccgacg tcttcagaa taaggcaaac gtgtgttggg   | 3120 |
| ccaaggcttt agtgcgggtg ctgaagaccg ctggcataga catgaccact gaacaatgga  | 3180 |
| acactgtgga ttatittgaa acggacaaag ctactcagc agagatagta ttgaaccaac   | 3240 |
| tatgcgtgag gtcttttggg ctcatctgg actccgtct atttctgca cccactgttc     | 3300 |
| cgttatccat taggaataat cactgggata actccccgtc gcctaacaatg tacgggctga | 3360 |
| ataaagaagt ggtccgtcag ctctctcgca ggtaccaca actgcctcgg gcagttgcca   | 3420 |
| ctggaagagt ctatgacatg aacactggta cactgcgcaa ttatgatccg cgcataaacc  | 3480 |
| tagtacctgt aaacagaaga ctgcctcatg ctttagtcct ccaccataat gaacaccac   | 3540 |
| agagtgaact ttcttcattc gtcagcaaat tgaaggcgag aactgtcctg gtggtcgggg  | 3600 |
| aaaagttgtc cgtcccagcg aaaatgggtg actggtgtgc agaccggcct gaggctacct  | 3660 |
| tcagagctcg gctggattta ggcatcccag gtgatgtgcc caaatatgac ataatatitg  | 3720 |
| ttaatgtgag gaccccatat aaataccatc actatcagca gtgtgaagac catgccatta  | 3780 |
| agcttagcat gttgaccaag aaagcttgtc tgcactgaa tcccggcgga acctgtgtca   | 3840 |
| gcataggtta tggttacgct gacagggcca gcgaaagcat cattggtgct atagcgggc   | 3900 |
| agttcaagtt ttcccgggta tgcaaacga aatcctcact tgaagagacg gaagtictgt   | 3960 |
| ttgtattcat tgggtacgat cgcaaggccc gtacgcacaa ttcttacaag ctttcatcaa  | 4020 |
| ccttgaccaa catttataca ggttcagac tccacgaagc cggatgtgca cctcatatc    | 4080 |
| atgtggtgcg aggggatatt gccacggcca ccgaaggagt gattataaat gctgctaaca  | 4140 |
| gcaaaggaca acctggcgga ggggtgtgctg gagcgctgta taagaaattc ccgaaagct  | 4200 |
| tcgatttaca gccgatcgaa gtaggaaaag cgcgactggt caaagggtgca gctaaacata | 4260 |
| tcattcatgc ctaggacca aacttcaaca aagtttcgga ggttgaaggt gacaaacagt   | 4320 |
| tggcagaggc ttatgagtc atcgctaaga ttgtcaacga taacaattac aagtcagtag   | 4380 |
| cgattccact gtgtccacc ggcatctttt ccgggaacaa agatcgacta acccaatcat   | 4440 |
| tgaaccattt gctgacagct ttagacacca ctgatgcaga ttagaccata tactgcaggg  | 4500 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| acaagaaatg ggaaatgact ctcaaggaag cagtggctag gagagaagca gtggaggaga  | 4560 |
| tatgcatatc cgacgactct tcatgacag aacctgatgc agagctggtg aggggtgcatc  | 4620 |
| cgaagagtgc tttggctgga aggaagggct acagcacaag cgatggcaaa actttctcat  | 4680 |
| atttgggaagg gaccaagttt caccaggcgg ccaaggatat agcagaaatt aatgccatgt | 4740 |
| ggcccgttgc aacggaggcc aatgagcagg tatgcatgta taccctcgga gaaagcatga  | 4800 |
| gcagtattag gtcgaaatgc cccgtcgaag agtcggaagc ctccacacca cctagcacgc  | 4860 |
| tgcccttgctt gtgcatccat gccatgactc cagaaagagt acagcgccta aaagcctcac | 4920 |
|                                                                    |      |
| gtccagaaca aattactgtg tgcctatcct ttccattgcc gaagtataga atcactggtg  | 4980 |
| tgcagaagat ccaatgctcc cagcctatat tgttctcacc gaaagtgcct gcgtatattc  | 5040 |
| atccaaggaa gtatctctgt gaaacaccac cggtagacga gactccggag ccatcggcag  | 5100 |
| agaaccaatc cacagagggg acacctgaac aaccaccact tataaccgag gatgagacca  | 5160 |
| ggactagaac gcctgagccg atcatcatcg aagaggaaga agaggatagc ataagtttgc  | 5220 |
| tgtcagatgg cccgaccac cagggtgctgc aagtcgaggc agacattcac gggccgcctt  | 5280 |
| ctgtatctag ctcatcctgg tccattcttc atgcatccga ctttgatgtg gacagtttat  | 5340 |
|                                                                    |      |
| ccatacttga caccctggag ggagctagcg tgaccagcgg ggcaacgtca gccgagacta  | 5400 |
| actcttactt cgcaaagagt atggagtttc tggcgcgacc ggtgcctgcg cctcgaaacag | 5460 |
| tattcaggaa cctccacat cccgtccgc gcacaagaac accgtcactt gcaccagca     | 5520 |
| gggctgtctc gagaaccagc ctagtttcca ccccgccagg cgtgaatagg gtgatcacta  | 5580 |
| gagaggagct cgaggcgctt accccgtcac gcactcctag caggtcggtc tcgagaacca  | 5640 |
| gcctgggtctc caaccgccca ggcgtaaata gggtgattac aagagaggag tttagggcgt | 5700 |
| tcgtagcaca acaacaatga cggtttgatg cgggtgcata catcttttcc tccgacaccg  | 5760 |
|                                                                    |      |
| gtcaagggca tttaacaaca aatcagtaa ggcaaacggt gctatccgaa gtggtgttgg   | 5820 |
| agaggaccga attggagatt tcgtatgcc cgcgcctcga ccaagaaaa gaagaattac    | 5880 |
| tacgcaagaa attacagtta aatccacac ctgctaacag aagcagatac cagtccagga   | 5940 |
| aggtggagaa catgaaagcc ataacagcta gacgtattct gcaaggccta gggcattatt  | 6000 |
| tgaaggcaga aggaaaagtg gagtgtctacc gaacctgca tcctgttctt ttgtattcat  | 6060 |
| ctagtgtgaa ccgtgccttt tcaagcccca aggtcgcagt ggaagcctgt aacgccatgt  | 6120 |
| tgaaagagaa ctttccgact gtggcttctt actgtattat tccagagtac gatgcctatt  | 6180 |
|                                                                    |      |
| tggacatggt tgacggagct tcatgtgtct tagacactgc cagtttttgc cctgcaaagc  | 6240 |
| tgccgagctt tccaagaaa cactcctatt tggaaccac aatacgtcg gcagtgcctt     | 6300 |
| cagcgatcca gaacacgtc cagaacgtcc tggcagctgc caaaaaaga aattgcaatg    | 6360 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tcacgcaaat gagagaattg cccgtattgg attcggcggc ctttaattgt gaatgcttca  | 6420 |
| agaaatatgc gtgtaataat gaatatgtgg aaacgtttta agaaaacccc atcaggctta  | 6480 |
| ctgaagaaaa cgtggtaaat tacattacca aattaaaagg accaaaagct gctgctcttt  | 6540 |
| ttgcgaagac acataatttg aatatgttgc aggacatacc aatggacagg tttgtaattg  | 6600 |
|                                                                    |      |
| acttaagag agacgtgaaa gtgactccag gaacaaaaca tactgaagaa cggcccaagg   | 6660 |
| tacaggtgat ccaggctgcc gatccgctag caacagcgta tctgtgcgga atccaccgag  | 6720 |
| agctggttag gagattaaat gcggtcctgc ttccgaacat tcatacactg tttgatatgt  | 6780 |
| cggctgaaga ctttgacgct attatagccg agcacttcca gcctggggat tgtgttctgg  | 6840 |
| aaactgacat cgcgtcgttt gataaaagtg aggacgacgc catggctctg accgcgttaa  | 6900 |
| tgattctgga agacttaggt gtggacgcag agctgttgac gctgattgag gcggctttcg  | 6960 |
| gcgaaatttc atcaatacat ttgcccaacta aaactaaatt taaattcgga gccatgatga | 7020 |
|                                                                    |      |
| aatctggaat gticctcaca ctgtttgtga acacagtcac taacattgta atcgcaagca  | 7080 |
| gagtgttag agaacggcta accggatcac catgtgcagc attcattgga gatgacaata   | 7140 |
| tcgtgaaagg agtcaaactg gacaaattaa tggcagacag gtgcgccacc tggttgaata  | 7200 |
| tggaagtcaa gattatagat gctgtgggtg gcgagaaagc gccttatttc tgtggagggt  | 7260 |
| ttattttgtg tgactccgtg accggcacag cgtgccgtgt ggacagacccc ctaaaaaggc | 7320 |
| tgtttaagct tggcaaact ctggcagcag acgatgaaca tgatgatgac aggagaaggg   | 7380 |
| cattgatga agagtcaaca cgctggaacc gagtgggtat tctttcagag ctgtgcaagg   | 7440 |
|                                                                    |      |
| cagtagaatc aaggatgaa accgtaggaa ctccatcat agttatggcc atgactactc    | 7500 |
| tagctagcag tgtaaatca ttcagctacc tgagaggggc ccctataact ctctacggct   | 7560 |
| aacctgaatg gactacgaca tagtctagtc cgccaagtct agcatatgac cacctgatga  | 7620 |
| gcggccgga atttgcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa   | 7680 |
| aatttttatt ttatttttct tttcttttcc gaatcggatt ttgttttta tttttcaaaa   | 7740 |
| aaaaaaaaa aaaaaaaaaa aaaaaaaaaa acgcgtcgag gggaattaat tcttgaagac   | 7800 |
| gaaagggcca ggtggcactt ttcggggaaa tgtgcgcgga accctattt gtttattttt   | 7860 |
|                                                                    |      |
| ctaaatacat tcaataatgt atccgctcat gagacaataa ccctgataaa tgcttcaata  | 7920 |
| atattgaaaa aggaagagta tgagtattca acatttccgt gtcgccctta ttcccttttt  | 7980 |
| tgccgcattt tgctttctg tttttgctca ccagaaacg ctggtgaaag taaaagatgc    | 8040 |
| tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaagat  | 8100 |
| ccttgagagt tttgcctccg aagaacgttt tccaatgatg agcactttta aagttctgct  | 8160 |
| atgtggcgcg gtattatccc gtgttgacgc cgggcaagag caactcggtc gccgcataca  | 8220 |

ctattctcag aatgacttgg ttgagtactic accagtcaca gaaaagcatc ttacggatgg 8280

catgacagta agagaattat gcagtgtctgc cataaccatg agtgataaca ctgcggccaa 8340

cttacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 8400

ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga 8460

cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg 8520

cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaagt 8580

tgcaggacca ctctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg 8640

agccggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 8700

ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 8760

gatcgtctgag ataggtgcct cactgattaa gcatttgtaa ctgtcagacc aagtttactc 8820

atataactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 8880

cctttttgat aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc 8940

agaccccgta gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg 9000

ctgcttgcaa acaaaaaaac caccgtacc agcgggtggtt tgtttgccgg atcaagagct 9060

accaactctt ttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct 9120

tctagtgtag ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct 9180

cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg 9240

gttggaactca agacgatagt taccggataa ggcgacgagg tcgggctgaa cgggggggttc 9300

gtgcacacag ccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga 9360

gcattgagaa agcgccacgc ttcccgaagg gagaaaggcg gacaggatc cggtaagcgg 9420

cagggtcgga acaggagagc gcacgaggga gcttcagggg ggaaacgcct ggtatcttta 9480

tagtcctgtc gggtttcgc accctctgact tgagcgtcga tttttgtgat gctcgtcagg 9540

ggggcggagc ctatggaaaa a 9561

<210> 36

<211> 9963

<212> DNA

<213> Artificial Sequence

<220><223> vector comprising a nucleic acid sequence that encodes an RNA  
construct comprising PIV5

<220><221> misc\_feature

<222> (7614)..(7615)

<223> GOI is present between positions 7614 and 7615

<400> 36

|                                                                     |      |
|---------------------------------------------------------------------|------|
| cgccagcaac gcgagctcta atacgactca ctatagatgg gcggcgcatg agagaagccc   | 60   |
| agaccaatta cctacccaaa atggagaaaag ttcacgttga catcgaggaa gacagcccat  | 120  |
| tcctcagagc tttgcagcgg agcttcccgc agtttgaggt agaagccaag caggtcactg   | 180  |
| ataatgacca tgctaattgcc agagcgtttt cgcactctggc ttcaaaactg atcgaaacgg | 240  |
| aggtggaccc atccgacacg atccttgaca ttggaagtgc gcccggccgc agaattgatt   | 300  |
| ctaagcacia gtatcattgt atctgtccga tgagatgtgc ggaagatccg gacagattgt   | 360  |
| ataagtatgc aactaagctg aagaaaaact gtaaggaaat aactgataag gaattggaca   | 420  |
| agaaaatgaa ggagctggcc gccgtcatga gcgaccctga cctggaaact gagactatgt   | 480  |
| gcctccacga cgacgagtcg tgcgctacg aagggaagt cgctgtttac caggatgtat     | 540  |
| acgcggttga cggaccgaca agtctctatc accaagccaa taaggagtt agagtcgcct    | 600  |
| actggatagg ctttgacacc acccctttta tgtttaagaa cttggctgga gcatatccat   | 660  |
| catactctac caactgggcc gacgaaaccg tgttaacggc tcgtaacata ggcctatgca   | 720  |
| gctctgacgt taaggagcgg tcacgtagag ggatgtccat tcttagaaag aagtatttga   | 780  |
| aaccatccaa caatgttcta ttctctgttg gctcgacat ctaccacgag aagagggact    | 840  |
| tactgaggag ctggcacctg ccgtctgtat ttcacttacg tggcaagcaa aattacacat   | 900  |
| gtcgggttga gactatagtt agttgcgacg ggtacgtcgt taaaagaata gctatcagtc   | 960  |
| caggcctgta tgggaagcct tcaggctatg ctgctacgat gcaccgcgag ggattcttgt   | 1020 |
| gctgcaaagt gacagacaca ttgaacgggg agagggtctc ttttcccgtg tgcacgtatg   | 1080 |
| tgccagctac attgtgtgac caaatgactg gcatactggc aacagatgtc agtgcggacg   | 1140 |
| acgcgcaaaa actgctgggt gggctcaacc agcgtatagt cgtcaacggt cgcaccaga    | 1200 |
| gaaacaccaa taccatgaaa aattacctt tgcccgtagt ggcccaggca ttgctaggt     | 1260 |
| gggcaaagga atataaggaa gatcaagaag atgaaaggcc actaggacta cgagatagac   | 1320 |
| agttagtcat ggggtgttgt tgggctttta gaaggcacia gataacatct atttataagc   | 1380 |
| gcccggatac ccaaaccatc atcaaagtga acagcgattt ccactcattc gtgctgccc    | 1440 |
| ggataggcag taacacattg gagatcgggc tgagaacaag aatcaggaaa atgttagagg   | 1500 |
| agcacaagga gccgtcacct ctattaccg ccgaggacgt acaagaagct aagtgcgcag    | 1560 |
| ccgatgaggc taaggaggtg cgtgaagccg aggagttgcg cgcagctcta ccacctttgg   | 1620 |
| cagctgatgt tgaggagccc actctggaag ccgatgtcga cttgatgtta caaggagctg   | 1680 |
| gggccggctc agtggagaca cctcgtggct tgataaaggt taccagctac gatggcgagg   | 1740 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| acaagatcgg ctcttacgct gtgctttctc cgcaggctgt actcaagagt gaaaaattat  | 1800 |
|                                                                    |      |
| cttgcatcca cctctcgcgt gaacaagtca tagtgataac acactctggc cgaaaagggc  | 1860 |
| gttatgccgt ggaaccatac catggttaaag tagtggtgcc agaggacat gcaatacccg  | 1920 |
| tccaggactt tcaagctctg agtgaaagtg ccaccattgt gtacaacgaa cgtgagttcg  | 1980 |
| taaacaggta cctgcaccat attgccacac atggaggagc gctgaacact gatgaagaat  | 2040 |
| attacaaaac tgtcaagccc agcgagcacg acggcgaata cctgtacgac atcgacagga  | 2100 |
| aacagtgcgt caagaaagaa ctagtctctg ggctagggtc cacaggcgag ctggtggatc  | 2160 |
| ctcccttcca tgaattcgcc tacgagagtc tgagaacacg accagccgct ccttaccaag  | 2220 |
|                                                                    |      |
| taccaacat aggggtgtat ggcgtgccag gatcaggcaa gtctggcatc attaaaagcg   | 2280 |
| cagtcaccaa aaaagatcta gtggtgagcg ccaagaaaga aaactgtgca gaaattataa  | 2340 |
| gggacgtcaa gaaaatgaaa gggctggacg tcaatgccag aactgtggac tcagtctct   | 2400 |
| tgaatggatg caaacacccc gtagagacc tgtatatga cgaagctttt gcttgtcatg    | 2460 |
| caggtactct cagagcgctc atagccatta taagacctaa aaaggcagtg ctctgcgggg  | 2520 |
| atcccaaaca gtgcggtttt tttaacatga tgtgcctgaa agtgcatttt aaccacgaga  | 2580 |
| tttgacaca agtcttccac aaaagcatct ctgcggttg cactaaatct gtgacttcgg    | 2640 |
|                                                                    |      |
| tcgtctcaac cttgttttac gacaaaaaaa tgagaacgac gaatccgaaa gagactaaga  | 2700 |
| ttgtgattga cactaccggc agtaccaaac ctaagcagga cgatctcatt ctcaattgtt  | 2760 |
| tcagagggtg ggtgaagcag ttgcaaatag attacaaagg caacgaaata atgacggcag  | 2820 |
| ctgcctctca agggctgacc cgtaaagggtg tgtatgccgt tcggtacaag gtgaatgaaa | 2880 |
| atcctctgta cgcaccacc tcaaacatg tgaacgtcct actgaccgc acggaggacc     | 2940 |
| gcatcgtgtg gaaaacacta gccggcgacc catggataaa aacactgact gccaaatgacc | 3000 |
| ctgggaatth cactgccacg atagaggagt ggcaagcaga gcatgatgcc atcatgaggc  | 3060 |
|                                                                    |      |
| acatcttgga gagaccggac cctaccgacg tcttcagaa taaggcaaac gtgtgttggg   | 3120 |
| ccaaggcttt agtgcgggtg ctgaagaccg ctggcataga catgaccact gaacaatgga  | 3180 |
| acactgtgga ttatthtgaa acggacaaaag ctcactcagc agagatagta ttgaaccaac | 3240 |
| tatgcgtgag gttctthtga ctgatctgg actccggtct atthtctgca cccactgttc   | 3300 |
| cgthtatccat taggaataat cactgggata actccccgtc gcctaacatg tacgggctga | 3360 |
| ataaagaagt ggtccgtcag ctctctcgca ggtaccaca actgcctcgg gcagttgcca   | 3420 |
| ctggaagagt ctatgacatg aacactggta cactgcgcaa ttatgatccg cgcataaacc  | 3480 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| tagtacctgt aaacagaaga ctgcctcatg ctttagtcct ccaccataat gaacacccac | 3540 |
| agagtgactt ttcttcattc gtcagcaaat tgaagggcag aactgtcctg gtggtcgggg | 3600 |
| aaaagtgtc cgtcccaggc aaaatggttg actggttgtc agaccggcct gaggtacct   | 3660 |
| tcagagctcg gctggattta ggcatcccag gtgatgtgcc caaatatgac ataatatgtg | 3720 |
| ttaatgtgag gaccccatat aaataccatc actatcagca gtgtgaagac catgccatta | 3780 |
| agcttagcat gttgaccaag aaagcttgtc tgcactgaa tcccggcgga acctgtgtca  | 3840 |
| gcataggtta tggttacgct gacagggcca gcgaaagcat cattggtgct atagcgcggc | 3900 |
|                                                                   |      |
| agttcaagtt ttcccgggta tgcaaaccga aatcctcact tgaagagacg gaagttctgt | 3960 |
| ttgtattcat tgggtacgat cgcaaggccc gtacgcacaa ttcttacaag ctttcatcaa | 4020 |
| ccttgaccaa catttataca ggttccagac tccacgaagc cggatgtgca cctcatatc  | 4080 |
| atgtggtgcg aggggatatt gccacggcca ccgaaggagt gattataaat gctgctaaca | 4140 |
| gcaaaggaca acctggcgga ggggtgtgcg gagcgctgta taagaaattc ccgaaagct  | 4200 |
| tcgatttaca gccgatcgaa gtaggaaaag cgcgactggt caaaggtgca gctaacata  | 4260 |
| tcattcatgc cgtaggacca aacttcaaca aagtttcgga ggttgaaggt gacaaacagt | 4320 |
|                                                                   |      |
| tggcagaggc ttatgagtc atcgctaaga ttgtcaacga taacaattac aagtcagtag  | 4380 |
| cgattccact gttgtccacc ggcatctttt ccgggaacaa agatcgacta acccaatcat | 4440 |
| tgaaccattt gctgacagct ttagacacca ctgatgcaga ttagccata tactgcaggg  | 4500 |
| acaagaaatg ggaatgact ctcaaggaag cagtggctag gagagaagca gtggaggaga  | 4560 |
| tatgcatatc cgacgactct tcagtacag aacctgatgc agagctggtg aggggtgcatc | 4620 |
| cgaagagttc ttggctgga aggaagggt acagcacaag cgatggcaaa acttttcat    | 4680 |
| atttgaagg gaccaagttt caccaggcgg ccaaggatat agcagaaatt aatgccatgt  | 4740 |
|                                                                   |      |
| ggcccgttgc aacggaggcc aatgagcagg tatgcatgta tatcctcgga gaaagcatga | 4800 |
| gcagtattag gtcgaaatgc ccgctgaag agtcggaagc ctccacacca ctagcacgc   | 4860 |
| tgcttgcctt gtgcatccat gccatgactc cagaaagagt acagcgcta aaagcctcac  | 4920 |
| gtccagaaca aattactgtg tgctcactct ttccattgcc gaagtataga atcactggtg | 4980 |
| tgcagaagat ccaatgctcc cagcctatat tgttctcacc gaaagtcct gcgtatattc  | 5040 |
| atccaaggaa gtatctctg gaaacaccac cggtagacga gactccggag ccatcggcag  | 5100 |
| agaaccaatc cacagagggg acacctgaac aaccaccact tataaccgag gatgagacca | 5160 |
|                                                                   |      |
| ggactagaac gcctgagccg atcatcatcg aagaggaaga agaggatagc ataagtttgc | 5220 |
| tgtcagatgg cccgaccac caggtgctgc aagtcaggc agacattcac gggccgcct    | 5280 |
| ctgtatctag ctcatcctgg tccattctc atgcatccga ctttgatgtg gacagtttat  | 5340 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| ccatacttga caccctggag ggagctagcg tgaccagcgg ggcaacgtca gccgagacta | 5400 |
| actcttactt cgaaagagt atggagtttc tggcgcgacc ggtgcctgcg cctcgaacag  | 5460 |
| tattcaggaa cctccacat cccgtccgc gcacaagaac accgtcactt gcacccagca   | 5520 |
| gggcctgctc gagaaccagc ctagtttcca ccccgccagg cgtgaatagg gtgatcacta | 5580 |
|                                                                   |      |
| gagaggagct cgaggcgctt accccgtcac gcactcctag caggtcggtc tcgagaacca | 5640 |
| gcctggtctc caaccgcca ggcgtaaata gggtgattac aagagaggag tttagggcgt  | 5700 |
| tcgtagcaca acaacaatga cggtttgatg cgggtgcata catcttttcc tccgacaccg | 5760 |
| gtcaagggca tttaacaaca aaatcagtaa ggcaaacggt gctatccgaa gtggtgttgg | 5820 |
| agaggaccga attggagatt tcgtatgcc cgcgcctcga ccaagaaaaa gaagaattac  | 5880 |
| tacgcaagaa attacagtta aatccacac ctgctaacag aagcagatac cagtccagga  | 5940 |
| aggtggagaa catgaaagcc ataacagcta gacgtattct gcaaggccta gggcattatt | 6000 |
|                                                                   |      |
| tgaaggcaga aggaaaagtg gagtgtacc gaaccctgca tccgttcctt ttgtattcat  | 6060 |
| ctagtgtgaa ccgtgccttt tcaagcccca aggtcgcagt ggaagcctgt aacgccatgt | 6120 |
| tgaaagagaa ctttccgact gtggcttctt actgtattat tccagagtac gatgcctatt | 6180 |
| tggacatggt tgacggagct tcatgtctgt tagacactgc cagtttttgc cctgcaaagc | 6240 |
| tgcgcagctt tccaaagaaa cactcctatt tggaaccac aatacgatcg gcagtgcctt  | 6300 |
| cagcgatcca gaacacgtc cagaacgtcc tggcagctgc cacaaaaaga aattgcaatg  | 6360 |
| tcacgcaaat gagagaattg cccgtattgg attcggcggc ctttaatgtg gaatgcttca | 6420 |
|                                                                   |      |
| agaaatatgc gtgtaataat gaatatggg aaacgtttaa agaaaacccc atcaggetta  | 6480 |
| ctgaagaaaa cgtggtaaat tacattacca aattaaagg accaaaagct gctgctcttt  | 6540 |
| ttgcgaagac acataatttg aatatgttgc aggacatacc aatggacagg ttgtaatgg  | 6600 |
| acttaaagag agacgtgaaa gtgactccag gaacaaaaca tactgaagaa cggcccaagg | 6660 |
| tacaggtgat ccaggctgcc gatccgctag caacacgta tctgtgcgga atccaccgag  | 6720 |
| agctggttag gagattaaat gcggtcctgc ttccgaacat tcataactg ttigtatatgt | 6780 |
| cggctgaaga ctttgacgt attatagccg agcacttcca gcctggggat tgtgttctgg  | 6840 |
|                                                                   |      |
| aaactgacat cgcgtcgttt gataaaagtg aggacgacgc catggctctg accgcgttaa | 6900 |
| tgattctgga agacttaggt ttggacgcag agctgttgac gctgattgag gcggctttcg | 6960 |
| gcgaaatttc atcaatacat ttgccacta aaactaaatt taaattcgga gccatgatga  | 7020 |
| aatctggaat gttcctcaca ctgtttgtga acacagtcac taacattgta atcgcaagca | 7080 |
| gagtgttag agaacggcta accggatcac catgtgcagc attcattgga gatgacaata  | 7140 |
| tcgtgaaagg agtcaaatcg gacaaattaa tggcagacag gtgcgccacc tggttgaata | 7200 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tggaagtcaa gattatagat gctgtggtgg gcgagaaagc gccttatttc tgtggagggt  | 7260 |
|                                                                    |      |
| ttatittgtg tgactccgtg accggcacag cgtgccgtgt ggcagacccc ctaaaaaggc  | 7320 |
| tgtttaagct tggcaaacct ctggcagcag acgatgaaca tgatgatgac aggagaaggg  | 7380 |
| cattgcatga agagtcaaca cgctggaacc gagtgggtat tctttcagag ctgtgcaagg  | 7440 |
| cagtagaatc aaggtatgaa accgtaggaa ctcccatcat agttatggcc atgactactc  | 7500 |
| tagctagcag tgttaaatca ttcagctacc tgagaggggc ccctataact ctctacggct  | 7560 |
| aacctgaatg gactacgaca tagtctagtc cgccaagtct agcatatggc caccgggaga  | 7620 |
| aagagaggct ctggcgaagg cagaggcagc ctgcttacat gtggcgacgt ggaagagaac  | 7680 |
|                                                                    |      |
| cccggaccta tggactacgt gtccctgctg aaccagattt ggcagaagta cctgaacagc  | 7740 |
| ccctacacca cctgtctgta catccccaag cctaccgcca agtacacacc tctcgtgggc  | 7800 |
| acatctctgc accccgtgct gtggaattgc cagctgagct ttgccggcta caccgagtct  | 7860 |
| gccgtgaaca gcacaaagcg cctggccaaa caggacgccg ctgagagaat tgcctggctg  | 7920 |
| ctgcacaagg atggcggcat ccctgatggc ttagcctgt acctgagaca cagcagcctg   | 7980 |
| ttcgcccaga gcgaggaaga ggaatccttc agcaactgat gagcggccgc gaattggcaa  | 8040 |
| gctgcttaca tagaactcgc ggcgattggc atgccgcctt aaaattttta ttttattttt  | 8100 |
|                                                                    |      |
| cttttctttt ccgaatcgga ttttgttttt aatatttcaa aaaaaaaaaa aaaaaaaaaa  | 8160 |
| aaaaaaaaaa aaacgcgtcg aggggaattt attcttgaag acgaaaggcg caggtggcac  | 8220 |
| ttttcgggga aatgtgcgcg gaaccctat ttgtttattt ttctaatac attcaaatat    | 8280 |
| gtatccgctc atgagacaat aacctgata aatgcttcaa taatattgaa aaaggaagag   | 8340 |
| tatgagtatt caacatttcc gtgtcgccct tattcccttt ttgctggcat ttgccttcc   | 8400 |
| tgtttttgc taccagaaa cgctgggtgaa agtaaaagat gctgaagatc agttgggtgc   | 8460 |
| acgagtgggt tacatcgaac tggatctcaa cagcggttaag atccttgaga gttttcgccc | 8520 |
|                                                                    |      |
| cgaagaacgt ttccaatga tgagcacttt taaagtcttg ctatgtggcg cggtattatc   | 8580 |
| ccgtgtgtac gccgggcaag agcaactcgg tcgccgata cactattctc agaatgactt   | 8640 |
| ggttgagtac tcaccagtc cagaaaagca tcttacggat ggcatgacag taagagaatt   | 8700 |
| atgcagtgt gccataacca tgagtataaa cactgcggcc aacttacttc tgacaacgat   | 8760 |
| cggaggaccg aaggagctaa ccgttttttt gcacaacatg ggggatcatg taactcgctt  | 8820 |
| tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat  | 8880 |
| gcctgtagca atggcaacaa cgttgcgcaa actattaact ggcgaaactac ttactctagc | 8940 |

ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg 9000  
ctcggccctt cggctggct ggtttattgc tgataaatct ggagccgtg agcgtgggtc 9060  
tcgcggtatc attgcagcac tggggccaga tgtaagccc tccgtatcg tagttatcta 9120  
cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg agatagggtc 9180  
ctcactgatt aagcattggc aactgtcaga ccaagtttac tcatatatac tttagattga 9240  
tttaaaactt catttttaat ttaaaaggat ctaggtgaag atcctttttg ataatctcat 9300  
gacccaaatc cctaacgtg agttttcgtt ccactgagcg tcagaccccg tagaaaagat 9360

caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa 9420  
accaccgcta ccagcgggtg tttgtttgcc ggatcaagag ctaccaactc tttttccgaa 9480  
ggtaactggc ttcagcagag cgcagatacc aaatactgtc cttctagtgt agccgtagtt 9540  
aggccaccac ttcaagaact ctgtagcacc gcctacatac ctgctctgc taatcctgtt 9600  
accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata 9660  
gttaccgat aaggcgcagc ggtcgggctg aacggggggt tcgtgcacac agcccagctt 9720  
ggagcgaacg acctacaccg aactgagata cctacagcgt gagcattgag aaagcggcac 9780

gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg gaacaggaga 9840  
gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg 9900  
ccaccttga cttgagcgtc gatttttgtg atgctcgtca ggggggcgga gcctatggaa 9960  
aaa 9963

<210> 37

<211> 10302

<212> DNA

<213> Artificial Sequence

<220><223> vector comprising a nucleic acid sequence that encodes an RNA  
construct comprising PIV5

<220>

<221> misc\_feature

<222> (7614)..(7615)

<223> GOI present between positions 7614 and 7615

<400> 37

cgccagcaac gcgagctcta atacgactca ctatagatgg gcggcgcgtg agagaagccc 60  
agaccaatta cctacccaaa atggagaaaag ttcacgttga catcgaggaa gacagcccat 120  
tcctcagagc tttgcagcgg agcttccgc agtttgaggt agaagccaag caggtcactg 180  
ataatgacca tgctaagcc agagcgtttt cgcactctggc ttcaaaactg atcgaaacgg 240

|                                                                    |      |
|--------------------------------------------------------------------|------|
| agggtggaccc atccgacacg atccttgaca ttggaagtgc gcccgcgccg agaatgtatt | 300  |
|                                                                    |      |
| ctaagcacia gtatcattgt atctgtccga tgagatgtgc ggaagatccg gacagattgt  | 360  |
| ataagtatgc aactaagctg aagaaaaact gtaaggaaat aactgataag gaattggaca  | 420  |
| agaaaatgaa ggagctggcc gccgtcatga gcgaccctga cctggaaact gagactatgt  | 480  |
| gcctccacga cgacgagtcg tgtcgtacg aagggaagt cgctgtttac caggatgtat    | 540  |
| acgcggttga cggaccgaca agtctctatc accaagccaa taaggagatt agagtcgcct  | 600  |
| actggatagg ctttgacacc acccctttta tgtttaagaa cttggctgga gcataatccat | 660  |
| catactctac caactgggcc gacgaaaccg tgttaacggc tcgtaacata ggcctatgca  | 720  |
|                                                                    |      |
| gctctgacgt taiggagcgg tcacgtagag ggatgtccat tcttagaaag aagtatttga  | 780  |
| aaccatccaa caatgttcta ttctctgttg gctcgacat ctaccacgag aagagggact   | 840  |
| tactgaggag ctggcacctg ccgtctgtat ttactttacg tggcaagcaa aattacacat  | 900  |
| gtcgggttga gactatagtt agttgcgacg ggtacgtcgt taaaagaata gctatcagtc  | 960  |
| caggcctgta tgggaagcct tcaggctatg ctgctacgat gcaccgcgag ggattcttgt  | 1020 |
| gctgcaaagt gacagacaca ttgaacgggg agagggtctc ttttccgtg tgcacgtatg   | 1080 |
| tgccagctac attgtgtgac caaatgactg gcatactggc aacagatgtc agtgcggacg  | 1140 |
|                                                                    |      |
| acgcgcaaaa actgctggtt gggctcaacc agcgtatagt cgtcaacggt cgcaccacga  | 1200 |
| gaaacaccaa taccatgaaa aattaccttt tgcccgtagt ggcccaggca ttgtctaggt  | 1260 |
| gggcaaagga atataaggaa gatcaagaag atgaaaggcc actaggacta cgagatagac  | 1320 |
| agttagtcat ggggtgttgt tgggctttta gaaggcacia gataacatct attataagc   | 1380 |
| gcccggatag ccaaaccatc atcaaagtga acagcgattt ccactcattc gtgctgcccc  | 1440 |
| ggataggcag taacacattg gagatcgggc tgagaacaag aatcaggaaa atgttagagg  | 1500 |
| agcacaagga gccgtcacct ctcatcaccg ccgaggacgt acaagaagct aagtgcgcag  | 1560 |
|                                                                    |      |
| ccgatgaggc taaggaggctg cgtgaagccg aggagttgcg cgcagctcta ccacctttgg | 1620 |
| cagctgatgt tgaggagccc actctggaag ccgatgtcga cttgatgtta caaggagctg  | 1680 |
| gggccggctc agtggagaca cctcgtggct tgataaaggt taccagctac gatggcgagg  | 1740 |
| acaagatcgg ctcttacgt gtgctttctc cgcaggctgt actcaagagt gaaaaattat   | 1800 |
| cttgcatcca cctctcgtg gaacaagtca tagtgataac acactctggc cgaaggggc    | 1860 |
| gttatgccgt ggaaccatac catggttaaag tagtggtgcc agaggacat gcaataccg   | 1920 |
| tccaggactt tcaagctctg agtgaaagt ccaccattgt gtacaacgaa cgtgagttcg   | 1980 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| taaacaggta cctgcacat attgccacac atggaggagc gctgaacact gatgaagaat   | 2040 |
| attacaaaac tgtcaagccc agcgagcacg acggcgaata cctgtacgac atcgacagga  | 2100 |
| aacagtgcgt caagaaagaa ctagtcactg ggctagggtt cacaggcgag ctggtggatc  | 2160 |
| ctcccttcca tgaattcgcc tacgagagtc tgagaacacg accagccgct ccttaccaag  | 2220 |
| taccaacat aggggtgtat ggcgtgccag gatcaggcaa gtctggcatc attaaaagcg   | 2280 |
| cagtcaccaa aaaagatcta gtggtgagcg ccaagaaaga aaactgtgca gaaattataa  | 2340 |
| gggacgtcaa gaaaatgaaa gggctggacg tcaatgccag aactgtggac tcagtgtctt  | 2400 |
|                                                                    |      |
| tgaatggatg caaacacccc gtagagacc tgtatatga cgaagctttt gcttgtcatg    | 2460 |
| caggtactct cagagcgctc atagccattt taagacctaa aaaggcagtg ctctgcgggg  | 2520 |
| atcccaaaca gtgcggtttt tttaacatga tgtgcctgaa agtgcatttt aaccacgaga  | 2580 |
| tttgacaca agtcttccac aaaagcatct ctgccgttg cactaaatct gtgacttcgg    | 2640 |
| tcgtctcaac ctgtttttac gacaaaaaaa tgagaacgac gaatccgaaa gagactaaga  | 2700 |
| ttgtgattga cactaccggc agtaccaaac ctaagcagga cgatctcatt ctcaattgtt  | 2760 |
| tcagagggtg ggtgaagcag ttgcaaatag attacaaagg caacgaaata atgacggcag  | 2820 |
|                                                                    |      |
| ctgcctctca agggctgacc cgtaaagggtg tgtatgccgt tcggtacaag gtgaatgaaa | 2880 |
| atcctctgta cgcaccacc tcagaacatg tgaacgtcct actgaccgc acggaggacc    | 2940 |
| gcatcgtgtg gaaaacacta gccggcgacc catggataaa aacctgact gccaaatacc   | 3000 |
| ctgggaattt cactgccacg atagaggagt ggcaagcaga gcatgatgcc atcatgaggc  | 3060 |
| acatcttga gagaccggac cctaccgacg tcttcagaa taaggcaaac gtgtgttggg    | 3120 |
| ccaaggcttt agtgcgggtg ctgaagaccg ctggcataga catgaccact gaacaatgga  | 3180 |
| acactgtgga ttattttgaa acggacaaag ctcaatcagc agagatagta ttgaaccaac  | 3240 |
|                                                                    |      |
| tatgcgtgag gttcttttga ctgatctgg actccgtct attttctgca cccactgttc    | 3300 |
| cgttatccat taggaataat cactgggata actccccgtc gcctaacatg tacgggctga  | 3360 |
| ataaagaagt ggtccgtcag ctctctcgca ggtaccaca actgcctcgg gcagttgcca   | 3420 |
| ctggaagagt ctatgacatg aacctggta cactgcgcaa ttatgatccg cgcataaacc   | 3480 |
| tagtacctgt aaacagaaga ctgcctcatg ctttagtcct ccaccataat gaacaccac   | 3540 |
| agagtgaatt ttcttcttc gtcagcaaat tgaaggcgag aactgtcctg gtggtcgggg   | 3600 |
| aaaagtgtc cgtcccagcg aaaatggttg actggtgtg agaccggcct gaggtacct     | 3660 |
|                                                                    |      |
| tcagagctcg gctggattta ggcacccag gtgatgtgcc caaatatgac ataatatgtg   | 3720 |
| ttaatgtgag gacccatatt aaataccatc actatcagca gtgtgaagac catgccatta  | 3780 |
| agcttagcat gttgaccaag aaagcttgtc tgcactgaa tcccggcgga acctgtgtca   | 3840 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| gcataggtta tggttacgct gacagggcca gcgaaagcat cattggtgct atagcgcggc | 3900 |
| agttcaagtt ttcccggga tgcaaacga aatcctcact tgaagagacg gaagtctgt    | 3960 |
| ttgtattcat tgggtacgat cgcaaggccc gtacgcacaa ttcttacaag ctttcatcaa | 4020 |
| ccctgaccaa catttataca gggtccagac tccacgaagc cggatgtgca ccctcatatc | 4080 |
|                                                                   |      |
| atgtggtgcg aggggatatt gccacggcca ccgaaggagt gattataaat gctgctaaca | 4140 |
| gcaaaggaca acctggcgga ggggtgtgcg gagcgctgta taagaaattc ccggaagct  | 4200 |
| tcgatttaca gccgatcgaa gtaggaaaag cgcgactggt caaagggtgca gctaacata | 4260 |
| tcattcatgc cgtaggacca aacttcaaca aagtttcgga ggttgaaggt gacaaacagt | 4320 |
| tggcagaggg ttatgagtc atcgctaaga ttgtcaacga taacaattac aagtcagtag  | 4380 |
| cgattccact gttgtccacc ggcatctttt ccgggaacaa agatcgacta acccaatcat | 4440 |
| tgaaccattt gctgacagct ttagacacca ctgatgcaga ttagccata tactgcaggg  | 4500 |
|                                                                   |      |
| acaagaaatg ggaaatgact ctcaaggaag cagtggctag gagagaagca gtggaggaga | 4560 |
| tatgcatatc cgacgactct tcagtacag aacctgatgc agagctggtg aggggtgcatc | 4620 |
| cgaagagttc tttggctgga aggaagggt acagcacaa cgatggcaaa actttctcat   | 4680 |
| atttggaagg gaccaagttt caccagggcg ccaaggatat agcagaaatt aatgccatgt | 4740 |
| ggcccgttgc aacggaggcc aatgagcagg tatgcatgta tctctcgga gaaagcatga  | 4800 |
| gcagtattag gtcgaaatgc cccgtcgaag agtcggaagc ctccacacca ctagcacgc  | 4860 |
| tgcccttgctt gtgcatccat gccatgactc cagaaagagt acagcgcta aaagcctcac | 4920 |
|                                                                   |      |
| gtccagaaca aattactgtg tgctcatcct ttccattgcc gaagtataga atcactggtg | 4980 |
| tgcagaagat ccaatgctcc cagcctatat tgttctcacc gaaagtgcct gcgtatattc | 5040 |
| atccaaggaa gtatctctg gaaacaccac cggtagacga gactccggag ccacggcag   | 5100 |
| agaaccaatc cacagagggg acacctgaac aaccaccact tataaccgag gatgagacca | 5160 |
| ggactagaac gcctgagccg atcatcatcg aagaggaaga agaggatagc ataagtttgc | 5220 |
| tgtcagatgg cccgaccac caggtgctgc aagtcgagc agacattcac gggccgcct    | 5280 |
| ctgtatctag ctcatcctgg tccattctc atgcatccga ctttgatgtg gacagtttat  | 5340 |
|                                                                   |      |
| ccatacttga caccctggag ggagctagcg tgaccagcgg ggcaacgtca gccgagacta | 5400 |
| actcttactt cgaaagagt atggagtttc tggcgcgacc ggtgcctgcg cctcgaacag  | 5460 |
| tattcaggaa cctccacat cccgtccgc gcacaagaac accgtcactt gcaccagca    | 5520 |
| gggcctgctc gagaaccagc ctagtttcca cccgccagg cgtgaatagg gtgatcacta  | 5580 |
| gagaggagct cgaggcgctt accccgtcac gcactcctag caggtcggtc tcgagaacca | 5640 |
| gcctggtctc caaccgccca ggcgtaaata ggttgattac aagagaggag tttgaggcgt | 5700 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| tcgtagcaca acaacaatga cggtttgatg cgggtgcata catcttttcc tccgacaccg | 5760 |
|                                                                   |      |
| gtcaagggca tttaacaaca aaatcagtaa ggcaaacggt gctatccgaa gtggtgttgg | 5820 |
| agaggaccga attggagatt tcgtatgcc cgcgcctcga ccaagaaaaa gaagaattac  | 5880 |
| tacgcaagaa attacagtta aatccacac ctgctaacag aagcagatac cagtccagga  | 5940 |
| aggtggagaa catgaaagcc ataacagta gacgtattct gcaaggccta gggcattatt  | 6000 |
| tgaaggcaga aggaaaagtg gagtgtacc gaaccctgca tctgttctt ttgtattcat   | 6060 |
| ctagtgtgaa ccgtgccttt tcaagcccca aggtcgagc ggaagcctgt aacgccatgt  | 6120 |
| tgaagagaa ctttcgact gtggcttctt actgtattat tccagagtac gatgcctatt   | 6180 |
|                                                                   |      |
| tggacatggt tgacggagct tcatgtgct tagacactgc cagtttttgc cctgcaaagc  | 6240 |
| tgcgcagctt tccaaagaaa cactcctatt tggaaccac aatacgatcg gcagtgcctt  | 6300 |
| cagcgatcca gaacacgtc cagaacgtcc tggcagctgc cacaaaaaga aattgcaatg  | 6360 |
| tcacgcaaat gagagaattg cccgtattgg attcggcggc ctttaattgt gaatgcttca | 6420 |
| agaaatatgc gtgtaataat gaatatggg aaacgtttaa agaaaacccc atcaggctta  | 6480 |
| ctgaagaaaa cgttgtaaat tacattacca aattaaaagg accaaaagct gctgctcttt | 6540 |
| ttgcaagac acataatttg aatatgttg aggacatacc aatggacagg tttgtaatgg   | 6600 |
|                                                                   |      |
| acttaaagag agacgtgaaa gtgactccag gaacaaaaca tactgaagaa cggccaagg  | 6660 |
| tacaggtgat ccaggctgcc gatccgctag caacacgta tctgtgcgga atccaccgag  | 6720 |
| agctggttag gagattaaat gcggtcctgc ttccgaacat tcataactg tttgatattg  | 6780 |
| cggctgaaga ctttgacgt attatagccg agcacttcca gcctggggat tgtgttctgg  | 6840 |
| aaactgacat cgcgtcgttt gataaaagt aggacgacgc catggctctg accgcgttaa  | 6900 |
| tgattctgga agacttaggt gtggacgcag agctgttgac gctgattgag gcggctttcg | 6960 |
| gcgaaatttc atcaatacat ttgccacta aaactaaatt taaattcgga gccatgatga  | 7020 |
|                                                                   |      |
| aatctggaat gttcctcaca ctgtttgtga acacagtc taacattgta atcgcaagca   | 7080 |
| gagtgttag agaacggcta accggatcac catgtgcagc attcattgga gatgacaata  | 7140 |
| tcgtgaaagg agtcaaactg gacaaattaa tggcagacag gtgcgccacc tggttgaata | 7200 |
| tggaagtcaa gattatagat gctgtggtgg gcgagaaagc gccttatttc tgtggagggt | 7260 |
| ttattttgtg tgactccgtg accggcacag cgtgccgtgt ggcagacccc ctaaaaaggc | 7320 |
| tgtttaagct tggcaaacct ctggcagcag acgatgaaca tgatgatgac aggagaaggg | 7380 |
| cattgcatga agagtcaaca cgctggaacc gagtgggtat tctttcagag ctgtgcaagg | 7440 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| cagtagaatc aaggtatgaa accgtaggaa cttccatcat agttatggcc atgactactc | 7500 |
| tagctagcag tgttaaatca ttcagctacc tgagaggggc ccctataact ctctacggct | 7560 |
| aacctgaatg gactacgaca tagtctagtc cgccaagtct agcatatggc caccgggaga | 7620 |
| aagagaggct ctggcgaagg cagaggcagc ctgcttacat gtggcgacgt ggaagagaac | 7680 |
| cccgaccta tggaccctac cgacctgagc ttcagccccg acgagatcaa caagctgac   | 7740 |
| gagacaggcc tgaacaccgt ggaatacttc accagccagc aagtgaccgg cacaagcagc | 7800 |
| ctgggcaaga acacaattcc tccaggcgtg accggcctgc tgacaaatgc tgccgaggcc | 7860 |
|                                                                   |      |
| aagatccaag agagcaccaa ccaccagaag ggctctgttg gaggcggagc caagcctaag | 7920 |
| aagcccagac ctaagatcgc catcgtgccc gccgacgata agacagtgcc tggcaagccc | 7980 |
| attcctaate ctctgtggg cctcgacagc acccctagca cacagacagt gctggatctg  | 8040 |
| agcggcaaga cactgcctag cggcagctat aagggcgtga agctggccaa gttcggcaaa | 8100 |
| gaaaacctga tgaccgggtt catcgaggaa cccagagaga accctatcgc caccagctct | 8160 |
| cccctgact tcaagagagg cagagacacc ggcggttcc acagaagaga gtacagcatt   | 8220 |
| ggctgggtcg gagatgaagt gaaagtgacc gagtgggtga accccagctg cagccctatt | 8280 |
|                                                                   |      |
| acagccgccc ctagaagatt cgagtgcacc tgcaccagt gtcctgtgac ctgtagcgag  | 8340 |
| tgcgagcggg acacatgatg agcgcccgcg aattggcaag ctgcttacat agaactcgcg | 8400 |
| gcgattggca tgccgcctta aaatttttat tttatttttc ttttcttttc cgaatcgat  | 8460 |
| tttgttttta atatttcaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aacgcgtcga  | 8520 |
| ggggaattaa ttcttgaaga cgaaagggcc aggtggcact tttcggggaa atgtgcgcgg | 8580 |
| aaccctatt tgittatttt tctaaataca ttcaaataatg tatccgctca tgagacaata | 8640 |
| accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg | 8700 |
|                                                                   |      |
| tgtegccctt attccctttt ttgcggcatt ttgccttctt gtttttgctc acccagaaac | 8760 |
| gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact | 8820 |
| ggatctcaac agcggtgaaga tcttgagag ttttcgcccc gaagaacgtt ttccaatgat | 8880 |
| gagcactttt aaagtctgc tatgtggcgc ggtattatcc cgtgttgacg ccgggcaaga  | 8940 |
| gcaactcggc cgccgcatc actattctca gaatgacttg gttgagtact caccagtcac  | 9000 |
| agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgtg ccataacat   | 9060 |
| gagtataac actgcggcca acttacttct gacaacgac ggaggaccga aggagctaac   | 9120 |
|                                                                   |      |
| cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct | 9180 |
| gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgtagcaa tggcaacaac | 9240 |
| gttcgcaaaa ctattaactg gcgaactact tactctagct tcccggcaac aattaataga | 9300 |

ctggatggag gcgataaag ttgcaggacc acttctgcgc tcggcccttc cggctggctg 9360  
gtttattgct gataaatctg gagccggtga gctgggtct cgcggtatca ttgcagcact 9420  
ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac 9480  
tatggatgaa cgaaatagac agatcgctga gatagggtgcc tctactgatta agcattggta 9540

actgtcagac caagtttact catatatact ttagattgat ttaaaacttc atttttaatt 9600  
taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc cttaacgtga 9660  
gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt cttgagatcc 9720  
ttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgtac cagcgggtgt 9780  
ttgtttgccg gatcaagagc taccaactct tttccgaag gtaactggct tcagcagagc 9840  
gcagatacca aatactgtcc ttctagtgt gccgtagtta ggccaccact tcaagaactc 9900  
tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg 9960

cgataagtcg tgccttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg 10020  
gtcgggctga acggggggtt cgtgcacaca gccagcttg gagcgaacga cctacaccga 10080  
actgagatac ctacagcgtg agcattgaga aagcgccacg cttcccgaag ggagaaaggc 10140  
ggacaggtat ccggtaacgc gcagggtcgg aacaggagag cgcacgaggg agcttccagg 10200  
gggaaacgcc tggatcttt atagtctgt cgggtttcgc cacctctgac ttgagcgtcg 10260  
atthttgtga tgctcgtcag gggggcggag cctatggaaa aa 10302

<210> 38

<211> 10217

<212> RNA

<213> Artificial Sequence

<220><223> RNA construct

<400> 38

augggcggcg caugagagaa gccagacca auuaccuacc caaaauggag aaaguucacg 60  
uugacaucga ggaagacagc ccuuccuca gagcuuugca gcggagcuuc ccgcaguug 120  
agguagaagc caagcagguc acugauaaug accaugcuua ugccagagcg uuuucgcauc 180  
uggcuuacaa acugaucgaa acggaggugg acccauccga cacgauccuu gacauuggaa 240  
gugcgccgc ccgcagaaug uauucuaagc acaaguauca uuguauucgu ccgaugagau 300  
gugcggaaga uccggacaga uuguauaagu augcaacuaa gcugaagaaa aacuguaagg 360

aaauaacuga uaaggaauug gacaagaaaa ugaaggagcu ggccgcccgc augagcgacc 420  
cugaccugga aacugagacu augugccucc acgacgacga gucgugucgc uacgaagggc 480

|                                                                   |      |
|-------------------------------------------------------------------|------|
| aagucgcugu uuaccaggau guauacgcgg uugacggacc gacaagucuc uaucaccaag | 540  |
| ccaauaaggg aguuagaguc gccuacugga uaggcuuuga caccaccccu uuuauguuuu | 600  |
| agaacuuggc uggagcauau ccaucauacu cuaccaacug ggccgacgaa accguguuaa | 660  |
| cggcucguaa cauaggccua ugcagcucug acguuaugga gcggucacgu agagggaugu | 720  |
| ccauucuuag aaagaaguau uugaaaccu ccaacaugu ucuauucucu guuggcucga   | 780  |
| ccaucuaacca cgagaagg gacuuacuga ggagcuggca ccugccgucu guauuucacu  | 840  |
| uacguggcaa gcaaaauuac acaugucggu gugagacuau aguuaguugc gacggguacg | 900  |
| ucguuaaaag aauagcuauc aguccaggcc uguaugggaa gccuucaggc uaucgucua  | 960  |
| cgaugcaccg cgagggauuc uugucgucga aagugacaga cacauugaac ggggagaggg | 1020 |
| ucucuuuucc cgugucacg uaugugccag cuacauugug ugacaaaug acuggcacuac  | 1080 |
| uggcaacaga ugucagucg gacgacgcgc aaaaacugcu gguugggcuc aaccagcgua  | 1140 |
| uagucgucaa cggucgcacc cagagaaaca ccaauaccu gaaaaauuac cuuuugcccg  | 1200 |
| uagugccca ggcauuugcu agguaggcaa aggaauuaa ggaagauca gaagaugaaa    | 1260 |
| ggccacuagg acuacgagau agacaguua ucauggggug uuguugggcu uuuaagaggc  | 1320 |
| acaagauaac aucuauuuu aagcgcccgg auacccaaac caucaucaa gugaacagcg   | 1380 |
| auuuccacuc auucgugcug cccaggauag gcaguaacac auuggagauc gggcugagaa | 1440 |
| caagaucag gaaauguua gaggagcaca aggagccguc accucuuu accgccgagg     | 1500 |
| acguacaaga agcuaguc gcagccgaug aggcuaagga ggugcgugaa gccgaggagu   | 1560 |
| ugcgcgacg ucuaccacu uuggcagcug auguugagga gccacucug gaagccgaug    | 1620 |
| ucgacuugau guuacaagag gcuggggccg gcucagugga gacaccucgu ggcuuuauaa | 1680 |
| agguuaccag cuacgaugc gaggacaaga ucggcucua cgugugcuu ucuccgagg     | 1740 |
| cuguacucaa gagugaaaa uuauucugca uccaccucu cgcuagaaca gucauaguga   | 1800 |
| uaacacacuc ugccgaaaa gggcguaug ccguggaacc auaccauggu aaaguagugg   | 1860 |
| ugccagaggg acaugcaua cccguccagg acuuucaagc ucugagugaa agugccacca  | 1920 |
| uuguguacaa cgaacgugag uucguaaaca gguaccugca ccuuuugcc acacauggag  | 1980 |
| gagcgugaa cacugauga gaauuuuaca aaacuguca gcccgagcag cagcagggcg    | 2040 |
| aauaccugua cgacaucgac aggaaacagu gcgucaagaa agaacuaguc acugggcuag | 2100 |
| ggcucacagg cgagcuggug gaucucccu uccaugaaau cgccuacgag agucugagaa  | 2160 |
| cacgaccagc cgcuccuuac caaguacaa ccuaggggu guauggcgug ccaggauacg   | 2220 |
| gcaagucugg cauuuuuua agcgaguca ccaaaaaaga ucuaguggug agcgccaaga   | 2280 |
| aagaaaacug ugcagaaau auaggggacg ucaagaaau gaaagggcug gacgucaug    | 2340 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ccagaacugu ggacucagug cucuugaauug gaugcaaaca ccccguaagag acccuguaua | 2400 |
| uugacgaagc uuugcuugu caugcaggua cucucagagc gcucauagcc auuauaagac    | 2460 |
| cuaaaaaggc agugcucugc ggggauccea aacagugcgg uuuuuuaac augaugugcc    | 2520 |
| ugaaagugca uuuuaccac gagauuugca cacaagucuu ccacaaaagc aucucucgcc    | 2580 |
| guugcacuaa aucugugacu ucggucgucu caaccuuguu uuacgacaaa aaaaugagaa   | 2640 |
| cgacgaaucc gaaagagacu aagauuguga uugacacuaac cggcaguacc aaaccuaagc  | 2700 |
| aggacgaucu cauucucacu uguuucagag ggugggugaa gcaguugcaa auagauuaca   | 2760 |
| aaggcaacga aauaagacg gcagcugccu cucaaggguu gacccguaaa gguguguaug    | 2820 |
| ccguucggua caaggugaau gaaaauccuc uguacgcacc caccucagaa caugugaacg   | 2880 |
| uccuacugac ccgcacggag gaccgcaucg uguggaaaac acuagccggc gacccaugga   | 2940 |
| uaaaaacacu gacugccaag uaccucggga auuucacugc cagcauagag gaguggcaag   | 3000 |
| cagagcauga ugccaucag aggcacacu ugagagagacc ggaccuacc gacgucuucc     | 3060 |
| agaauaaggc aaacgugugu ugggccaagg cuuuagugcc ggugcugaag accgcuggca   | 3120 |
| uagacaugac cacugaacaa uggaacacug uggaauuuu ugaacaggac aaagcucacu    | 3180 |
| cagcagagau aguauugaac caacuaugcg ugagguucuu uggaucugau cuggacuccg   | 3240 |
| gucuauuuuc ugcaccacu guuccguuau ccuuaggaa uauacacugg gaaacuccc      | 3300 |
| cgucgccuaa cauguacggg cugaauaaag aagugguccg ucagcucucu cgcagguacc   | 3360 |
| cacaacugcc ucgggcaguu gccacuggaa gagucuauga caugaacacu gguacacugc   | 3420 |
| gcauuuuga uccgcgcaua aaccuaguac cuguaaacag aagacugccu caugcuuag     | 3480 |
| uccuccacca uaugaacac ccacagagug acuuuucuu auucgucagc aaauugaagg     | 3540 |
| gcagaacugu ccuggugugc ggggaaaagu uguccgucc aggcaaaug guugacuggu     | 3600 |
| ugucagaccg gccugaggu accuucagag cucggcugga uuuaggcauc ccaggugaug    | 3660 |
| ugcccaaaua ugacauaaua uuuguuaug ugaggacccc auauaaauac caucacuauc    | 3720 |
| agcaguguga agaccaugcc auuaagcuua gcauguugac caagaaagcu ugucugcauc   | 3780 |
| ugaaucccg cggaaccugu gucagcauag guuauugguu cgucgacagg gccagcgaau    | 3840 |
| gcaucauugg ugcuaugcg cggcaguua aguuuucccg gguaugcaaa ccgaaauccu     | 3900 |
| cacuugaaga gacggaagu cuguuuguau ucauugggua cgaucgcaag gcccgucgc     | 3960 |
| acaauucua caagcuuca ucaaccuuga ccaacauua uacagguucc agacuccacg      | 4020 |
| aagccggaug ugcaccuca uaucaugugg ugcgagggga uauugccacg gccaccgaag    | 4080 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gagugauuau aaaugcugcu aacagcaaag gacaaccugg cggaggggug ugcggagcgc  | 4140 |
| uguauaagaa auucccgaa agcuucgauu uacagccgau cgaaguagga aaagcgcgac   | 4200 |
| uggucaaaagg ugcagcuaaa cauaucauuc augccguagg accaaacuuc aacaaaguuu | 4260 |
| cggagguuga aggugacaaa caguuggcag aggcuuuga guccaucgcu aagauuguca   | 4320 |
| acgauaaca uuacaaguca guagcgauuc cacuguuguc caccggcauc uuuuccggga   | 4380 |
| acaaagaucg acuaacccaa ucauugaacc auuugcugac agcuuagac accacugaug   | 4440 |
| cagauguagc cauauacugc agggacaaga aaugggaaau gacucucaag gaagcagugg  | 4500 |
| cuaggagaga agcaguggag gagauaugca uauccgacga cucuucagug acagaaccug  | 4560 |
| augcagagcu ggugagggug cauccgaaga guucuuggc uggaaggaag ggcuacagca   | 4620 |
| caagcgaugg caaaacuuc ucauauuugg aagggaccaa guuucaccag gcggccaagg   | 4680 |
| auauagcaga aauuaugcc auguggcccg uugcaacgga ggccaauagag cagguaugca  | 4740 |
| uguauauccu cggagaaagc augagcagua uuaggucgaa augccccguc gaagagucgg  | 4800 |
| aagccuccac accaccuagc acgcugccuu gcuugugcau ccaugccaug acuccagaaa  | 4860 |
| gaguacagcg ccuaaaagcc ucacguccag acaaaauuac ugugugcuca uccuuuccau  | 4920 |
| ugccgaagua uagaauacu ggugugcaga agauccaaug cucccagccu auauuguucu   | 4980 |
| caccgaaagu gccugcgauu auucauccaa ggaaguaucu cguggaaaca ccaccgguag  | 5040 |
| acgagacucc ggagccaucg gcagagaacc aauccacaga ggggacaccu gaacaaccac  | 5100 |
| cacuuuaac cgaggaugag accaggacua gaacgccuga gccgauauc aucgaagagg    | 5160 |
| aagaagagga uagcauaagu uugcugucag augggccgac ccaccaggug cugcaagucg  | 5220 |
| aggcagacau ucacgggccg ccucugua uagcucauc cugguccauu ccuauugcau     | 5280 |
| ccgacuuuga uguggacagu uuauccauac uugacaccu ggagggagcu agcgugacca   | 5340 |
| gcggggcaac gucagccgag acuaacucu acuucgcaa gaguauggag uuucuggcgc    | 5400 |
| gaccggugcc ugcgccugca acaguaauca ggaaccucc acauccgcu ccgcgcacaa    | 5460 |
| gaacaccguc acuugcacc agcagggccu gcucgagaac cagccuaguu uccacccgc    | 5520 |
| caggcgugaa uagggugauc acuagagagg agcucgaggc gcuuaccccg ucacgcacuc  | 5580 |
| cuagcagguc ggucucgaga accagccugg ucuccaacc gccaggcgua aaugagguga   | 5640 |
| uuacaagaga ggaguuugag gcguucguag cacaacaaca augacgguu gaugcgggug   | 5700 |
| cauacauuu uuccuccgac accggucaag ggcauuuaca aaaaaauca guaaggcaaa    | 5760 |
| cggugcuauc cgaaguggug uuggagagga ccgaauugga gauuucguau gccccgcgc   | 5820 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ucgaccaaga aaaagaagaa uuacuacgca agaaaauaca guaaaauccc acaccugcua   | 5880 |
| acagaagcag auaccagucc aggaaggugg agaacaugaa agccauaaca gcuagacgua   | 5940 |
| uucugcaagg ccuagggc auuuugaagg cagaaggaaa aguggagugc uaccgaaccc     | 6000 |
| ugcauccugu uccuuuguau ucaucuagug ugaaccgugc cuuuucaagc cccaaggugc   | 6060 |
| caguggaagc cuguaacgcc auguugaaag agaacuucc gacuguggcu ucuuacugua    | 6120 |
| uuauuccaga guacgaugcc uauuuggaca ugguugacgg agcuucaugc ugcuuagaca   | 6180 |
| cugccaguuu uugcccugca aagcugcgca gcuuuccaaa gaaacacucc uauuuggaac   | 6240 |
| ccacaauacg aucggcagug ccuucagcga uccagaacac gcuccagaac guccuggcag   | 6300 |
| cugccacaaa aagaaaugc aaugucacgc aaaugagaga auugcccgua uuggauucgg    | 6360 |
| cggccuuuaa uguggaauagc uucaagaaau augcguguaa uaaugaauau ugaggaaacgu | 6420 |
| uuaaagaaaa ccccaucagg cuuacugaag aaaacguggu aaauuacauu accaaaauaa   | 6480 |
| aaggacaaa agcugcugcu cuuuuugcga agacacauaa uuugaauaug ugcaggaca     | 6540 |
| uaccaaugga cagguuugua auggacuuaa agagagacgu gaaagugacu ccaggaacaa   | 6600 |
| aacauacuga agaacggccc aagguacagg ugauccaggc ugccgaucgg cuagcaacag   | 6660 |
| cguaucugug cggaauccac cgagagcugg uuaggagauu aaugcgguc cugcuuccga    | 6720 |
| acauucauac acuguuugau augucggcug aagacuuuga cgcuauuua gccgagcacu    | 6780 |
| uccagccugg ggauuguguu cuggaaacug acaucgcguc guuugauaaa agugaggacg   | 6840 |
| acgccauggc ucugaccgug uuaaugauuc uggaagacuu agguuggac gcagagcugu    | 6900 |
| ugacgcugau ugaggcgguu uucggcgaaa uuucaucaau acuuuugccc acuaaaacua   | 6960 |
| aaauuaaaau cggagccaug augaaaucug gaauguuccu cacacuguuu gugaacacag   | 7020 |
| ucauuacau uguaaucgca agcagagugu ugagagaacg gcuaccgga ucaccaugug     | 7080 |
| cagcauuc auaggauagc aaauucguga aaggagucua aucggacaaa uuaauggcag     | 7140 |
| acaggugcgc caccugguug aaauuggaag ucaagauuuu agaugcugug gugggcgaga   | 7200 |
| aagcgccuuu uuucugugga ggguuuuuuu ugugugacuc cgugaccggc acagcgugcc   | 7260 |
| guguggcaga ccccuaaaa aggcuguuuu agcuuggcaa accucuggca gcagacgaug    | 7320 |
| aacaugauga ugacaggaga agggcauugc augaagaguc aacacgcugg aaccgagugg   | 7380 |
| guauucuuu agagcugugc aaggcaguag aaucaggua ugaaccgua ggaacuucca      | 7440 |
| ucauaguau ggccaugacu acucuagcua gcaguguuaa aucauucagc uaccugagag    | 7500 |
| gggccccuau aacucucuac ggcuaccug aauggacuac gacauagucu aguccgcaa     | 7560 |
| gucuagcau uguauggcca ccauggaaga ugccaagaac aucaagaagg gccugcccc     | 7620 |
| auucuacccc cuggaagaug gaacagccgg cgagcagcug cacaaggcca ugaagagaua   | 7680 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| cgcccuggug cccggcaca ucgccuucac cgaugccac aucgaggugg acauccua     | 7740 |
| cgccgaguac uucgagauga gcgugcggu ggccgaagcu augaagcgcu acggccugaa  | 7800 |
| caccaaccac cggaucgucg ugugcagcga gaacagccug caguucuua ugcccugcu   | 7860 |
| gggcgcccug uuuaucggag uggcuguggc ccugccaac gacauca acgagcgca      | 7920 |
| gcugcugaac agcaugggca ucagccagcc caccguggug uucgugucca agaaggacu  | 7980 |
| gcagaaauc cugaacguc agaagaagcu gcccuauc cagaaaauca ucaucaugga     | 8040 |
| cagcaagacc gacuaccagg gcuuccagag cauguacacc uucgugacca gccaucugcc | 8100 |
| cccugguuc aacgaguac acuuugugc cgagagcuuc gaccgggaca agacaaucc     | 8160 |
| ccugaucaug aacagcagcg gcagaccgg acugccuaa ggcguggccc ugccucacag   | 8220 |
| aacugccugc gugcgguua gccacgccg ggaccuauc uucggcaacc agaucaucc     | 8280 |
| cgacaccgcc auccugagcg ugguccuuu ccaccaggc uucggcaugu ucaccaccu    | 8340 |
| gggcuaccug aucugcggu uccggguggu gcugauguac agauucgagg aagaacuguu  | 8400 |
| ccugcgagc cugcaggacu acaagaacca gagcgcccug cuggugccua ccuguucag   | 8460 |
| cuucuugcc aagagcacc ugaucauua guacgaccug agcaaccugc acgagaucg     | 8520 |
| cucugcgga gcccccuu cuaaagaagu ggagaggcc guggccaagc gguuccaucu     | 8580 |
| gccuggauc agacagggu auggccugac cgagacaacc agcgccauuc ugaucacccc   | 8640 |
| cgaggcgac gauaagccug gcgcccuggg aaagguggug ccuuuucg aggccaaggu    | 8700 |
| gguggaccug gacaccgca agacacuggg cgugaaccag agggcggaac ugugugcg    | 8760 |
| gggaccuau aucaugagcg gcuacugaa caaccccgag gccaccaacg ccugauuga    | 8820 |
| caaggauaggc uggcugcaca gcggcgacau ugccuacugg gacgaggacg agcacuuuu | 8880 |
| caucgugac cgcgugaagu ccugauca guacaagggc uaccaggugg cccagccga     | 8940 |
| gcuggaau cuccugcugc agcacccaa cauuucgau gccggcgugg caggacugcc     | 9000 |
| cgaugaugau gcuggcgaac ugccagccg ugugguggug cuggaacacg gaaagaccu   | 9060 |
| gaccgagaaa gaaucgugg acuaugugc cagccaagug accaccgca agaacugag     | 9120 |
| aggcgcgug guuuuugg acgaggugcc aaaggccug acaggcaagc uggacgccc      | 9180 |
| gaagauca gagaucuga uuaaggcca gaaaggcggc aagaucgccg uggaucggag     | 9240 |
| aaagagaggc ucuggcgaag gcagaggcag ccugcuuaca ugugcgacg uggaagaga   | 9300 |
| ccccgaccu auggaccua ccgaccugag cuucagccc gacgagauca acaagcugau    | 9360 |
| cgagacaggc cugaacaccg uggaauacu caccagccag caagugaccg gcacaagcag  | 9420 |
| ccugggcaag aacacaau cuccaggcgu gaccggccug cugacaaug cugccgaggc    | 9480 |
| caagaucac gagagcacca accaccaga gggcucugu ggaggcgag ccaagccua      | 9540 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| gaagcccaga ccuaagaucg ccaucgugcc cgccgacgau aagacagugc cuggcaagcc  | 9600  |
| cauuccuaau ccucugcugg gccucgacag caccuccuagc acacagacag ugcuggaucu | 9660  |
| gagcggcaag acacugccua gcggcagcua uaagggcgug aagcuggcca aguucggcaa  | 9720  |
| agaaaaccug augaccgggu ucaucgagga acccagagag aaccuauucg ccaccagcuc  | 9780  |
| ucccaucgac uucaagagag gcagagacac cggcggcuuc cacagaagag aguacagcau  | 9840  |
| uggcuggguc ggagaugaag ugaaagugac cgaguggugc aaccccagcu gcagcccuau  | 9900  |
| uacagccgcc gcuagaagau ucgagugcac cugucaccag uguccuguga ccuguagcga  | 9960  |
| gugcgagcgg gacacaugau gagcggccgc gaauuggcaa gcugcuuaca uagaacucgc  | 10020 |
| ggcgauuggc augccgccuu aaaaauuuua uuuuauuuu cuuuucuuu ccgaauucgga   | 10080 |
| uuuuguuuuu aaauuuuca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   | 10140 |
| aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa  | 10200 |
| aaaaaaaaaa aaaaaa                                                  | 10217 |
| <210> 39                                                           |       |
| <211> 9878                                                         |       |
| <212> RNA                                                          |       |
| <213> Artificial Sequence                                          |       |
| <220><223> the RNA construct                                       |       |
| <400> 39                                                           |       |
| augggcggcg caugagagaa gccagacca auuaccuacc caaaauggag aaaguucacg   | 60    |
| uugacaucga ggaagacagc ccuuccuca gagcuuugca gcggagcuuc ccgcaguug    | 120   |
| agguagaagc caagcagguc acugauaaug accaugcuua ugccagagcg uuuucgcauc  | 180   |
| uggcuucaa acugaucgaa acggaggugg acccauccga cacgauccuu gacauuggaa   | 240   |
| gugcgccgc ccgcagaaug uauucuaagc acaaguauca uuguauucgu ccgaugagau   | 300   |
| gugcggaaga uccggacaga uuguauaagu augcaacuaa gcugaagaaa aacuguaagg  | 360   |
| aaauaacuga uaaggaauug gacaagaaaa ugaaggagcu ggccgccguc augagcgacc  | 420   |
| cugaccugga aacugagacu augugccucc acgacgacga gucgugucgc uacgaagggc  | 480   |
| aagucgcugu uuaccaggau guauacgcgg uugacggacc gacaagucuc uaucaccaag  | 540   |
| ccaauaaggg aguuagaguc gccuacugga uaggcuuuga caccacccu uuuauuuua    | 600   |
| agaacuuggc uggagcauau ccaucauacu cuaccaacug ggccgacgaa accguguuaa  | 660   |
| cggcucguua cauaggccua ugcagcucug acguuaugga gcggucacgu agagggaugu  | 720   |
| ccauucuuag aaagaaguau uugaaaccu ccaacaugu ucuauucucu guuggcucga    | 780   |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ccaucuacca cgagaagagg gacuuacuga ggagcuggca ccugccgucu guauuucacu  | 840  |
| uacguggcaa gcaaaaauac acaugucggu gugagacuau aguuaguugc gacggguacg  | 900  |
| ucguuuuag aauagcuau aguccaggcc uguauuggaa gccuucaggc uaugcugcua    | 960  |
| cgaugcaccg cgagggauuc uugugcugca aagugacaga cacauugaac ggggagaggg  | 1020 |
| ucucuuuucc cgugugcagc uaugugccag cuacauugug ugacaaaug acuggcuauc   | 1080 |
| uggcaacaga ugucagugcg gacgacgcgc aaaaacugcu gguugggcuc aaccagcgua  | 1140 |
| uagucgucaa cgugcgcacc cagagaaaca ccaauaccau gaaaaauac cuuuugcccg   | 1200 |
| uagugggcca ggcauuugcu agguugggcaa aggaauauaa ggaagaucaa gaagaugaaa | 1260 |
| ggccacuagg acuacgagau agacaguua ucauggggug uuguugggcu uuuagaaggc   | 1320 |
| acaagauaac aucuauuuau aagcgcccgg auacccaaac caucaucaa gugaacagcg   | 1380 |
| auuuuccacuc auucgugcug cccaggauag gcaguaacac auuggagauc gggcugagaa | 1440 |
| caagaauacg gaaaauuuu gaggagcaca aggagccguc accucucauu accgccgagg   | 1500 |
| acguacaaga agcuagugc gcagccgaug aggcuaagga ggugcgugaa gccgaggagu   | 1560 |
| ugcgcgcagc ucuaccacu uuggcagcug auguugagga gcccacucug gaagccgaug   | 1620 |
| ucgacuugau guuacaagag gcugggggcg gcucagugga gacaccucgu ggcuugauaa  | 1680 |
| agguuaccag cuacgauggc gaggacaaga ucggcucuua cgcugugcuu ucuccgcagg  | 1740 |
| cuguacucaa gagugaaaaa uuaucuugca uccaccucu cgcugaacaa gucauaguga   | 1800 |
| uaacacacuc ugccgaaaaa gggcguuau ccguggaacc auaccauggu aaaguagugg   | 1860 |
| ugccagaggg acaugcaaua cccguccagg acuuucaagc ucugagugaa agugccacca  | 1920 |
| uuguguacaa cgaacgugag uucguaaaca gguaccugca ccuuauugcc acacauggag  | 1980 |
| gagcgugaa cacugaugaa gaauuuuaca aaacugucua gccagcgag cacgacggcg    | 2040 |
| aauaccugua cgacaugcag aggaacacagu gcgucaagaa agaacuaguc acugggcuag | 2100 |
| ggcucacagg cgagcuggug gauccuccu uccaugaauu cgccuacgag agucugagaa   | 2160 |
| cacgaccagc cgcuccuuac caaguaccaa ccuaggggu guauggcgug ccaggauacg   | 2220 |
| gcaagucugg caucauuaaa agcgcaguca ccaaaaaaga ucuaguggug agcgccaaga  | 2280 |
| aagaaaacug ugcagaaaau auaagggacg ucaagaaaau gaaagggcug gacgucaaug  | 2340 |
| ccagaacugu ggacucagug cucuugaau augaacaaca ccccuagag acccuuaua     | 2400 |
| uugacgaagc uuuugcuugu caugcaggua cucucagagc gcucauagcc auuauaagac  | 2460 |
| cuaaaaaggc agugcucugc ggggauccca aacagugcgg uuuuuuuac augaugugcc   | 2520 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ugaaagugca uuuuaaccac gagauuugca cacaagucuu ccacaaaagc aucucucgcc  | 2580 |
|                                                                    |      |
| guugcacuaa aucugugacu ucggucgucu caaccuuguu uuacgacaaa aaaaugagaa  | 2640 |
| cgacgaaucc gaaagagacu aagauuguga uugacacuaac cggcaguacc aaaccuaagc | 2700 |
| aggacgaucu cauucucacu uguuucagag ggugggugaa gcaguugcaa auagauuaca  | 2760 |
| aaggcaacga aauaauagacg gcagcugccu cucaagggcu gacccguaaa gguguguaug | 2820 |
| ccguucggua caaggugaau gaaaauccuc uguacgcacc caccucagaa caugugaacg  | 2880 |
| uccuacugac ccgcacggag gaccgcaucg uguggaaaac acuagccggc gacccaugga  | 2940 |
| uaaaaacacu gacugccaag uaccucggga auuucacugc cacgauagag gaguggcaag  | 3000 |
|                                                                    |      |
| cagagcauga ugccaucaug aggcacauuc ugagagagacc ggacccuacc gacgucuucc | 3060 |
| agaauaaggc aaacgugugu ugaggccaagg cuuuagugcc ggugcugaag accgcuggca | 3120 |
| uagacaugac cacugaacaa uggaacacug uggaauuuuu ugaaacggac aaagcucacu  | 3180 |
| cagcagagau aguauugaac caacuaucg ugagguucuu uggaucugau cuggacuccg   | 3240 |
| gucuauuuuc ugcacccacu guuccguuau ccuuuaggaa uauucacugg gauaacuccc  | 3300 |
| cgucgccuaa cauguacggg cugaauaaaag aagugguccg ucagcucucu cgcagguacc | 3360 |
| cacaacugcc ucgggcaguu gccacuggaa gagucuaua caugaacacu gguacacugc   | 3420 |
|                                                                    |      |
| gcauuuaua uccgcgcaua aaccuaguac cuguaaacag aagacugccu caugcuuuag   | 3480 |
| uccuccacca uauagaacac ccacagagug acuuuucuu auucgucagc aaauugaagg   | 3540 |
| gcagaacugu ccuggugugc ggggaaaagu uguccgucc aggcaaaug guugacuggu    | 3600 |
| ugucagaccg gccugaggu accuucagag cucggcugga uuuaggcauc ccaggugaug   | 3660 |
| ugcccaaaua ugacauaua uuuguuaaug ugaggacccc auauaaaauac caucacuauc  | 3720 |
| agcaguguga agaccaugcc auuaagcuua gcauguugac caagaaagcu ugucugcauc  | 3780 |
| ugaaucccg cggaaccugu gucagcauag guuauggua cgucgacagg gccagcgaaa    | 3840 |
|                                                                    |      |
| gcaucauugg ugcuaugcgc cggcaguua aguuuucccg gguaugcaaa ccgaaauccu   | 3900 |
| cacuugaaga gacggaaguu cuguuuguau ucauugggua cgaucgcaag gcccguaagc  | 3960 |
| acaauucua caagcuuua ucaaccuuga ccaacauua uacagguucc agacuccacg     | 4020 |
| aagccggaug ugcacccua uaucaugugg ugcgagggga uauugccacg gccaccgaag   | 4080 |
| gagugauuu aaaugcugc aacagcaaag gacaaccugg cggaggggug ugcggagcgc    | 4140 |
| uguauaagaa auucccgga agcuucgau uacagccgau cgaaguagga aaagcgcgac    | 4200 |
| uggucaaagg ugcagcuaaa cauaucauuc augccguagg accaaacuuc aacaaaguuu  | 4260 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| cggagguuga agguacaaaa caguuggcag aggcuuuuga guccaucgcu aagauuguca  | 4320 |
| acgauaaca uuacaaguc guagcgauuc cacuguuguc caccggcauc uuuuccggga    | 4380 |
| acaaagaucg acuaacccaa ucauugaacc auuugcugac agcuuuagac accacugaug  | 4440 |
| cagauguagc cauauacugc agggacaaga aaugggaaau gacucucaag gaagcagugg  | 4500 |
| cuaggagaga agcaguggag gagauaugca uauccgacga cucuucagug acagaaccug  | 4560 |
| augcagagcu ggugagggug cauccgaaga guucuuuggc uggaaggaag ggcucacagca | 4620 |
| caagcgaugg caaaacuuc ucauauuugg aagggaacca guuucaccag gcggccaagg   | 4680 |
| auauagcaga aauuaugcc auguggcccg uugcaacgga ggccaauagag cagguaugca  | 4740 |
| uguauauccu cggagaaagc augagcagua uuaggucgaa augccccguc gaagagucgg  | 4800 |
| aagccuccac accaccuagc acgcugccuu gcuugugcau ccaugccaug acuccagaaa  | 4860 |
| gaguacagcg ccuaaaagcc ucacguccag aacaaauuac ugugugcuca uccuuuccau  | 4920 |
| ugccgaagua uagaauacu ggugugcaga agauccaaug cuccagccu auauuguucu    | 4980 |
| caccgaaagu gccugcgau auucauccaa ggaaguauc cguggaaaca ccaccgguag    | 5040 |
| acgagacucc ggagccaugc gcagagaacc aauccacaga ggggacaccu gaacaaccac  | 5100 |
| cacuuuaac cgaggauag accaggacua gaacgccuga gccgaucauc aucgaagagg    | 5160 |
| aagaagagga uagcauaagu uugcugucag auggcccgac ccaccaggug cugcaagucg  | 5220 |
| aggcagacau ucacgggccg ccucucuguau cuagcucauc cugguccauu ccuauugcau | 5280 |
| ccgacuuuga uguggacagu uuauccauac uugacacccu ggaggggagcu agcgugacca | 5340 |
| gcggggcaac gucagccgag acuaacucuu acuuugcaaa gaguauaggag uuucuggcgc | 5400 |
| gaccggugcc ugcgccuca acaguaauca ggaaccucc acaucccgcu ccgcgcacaa    | 5460 |
| gaacaccguc acuuacacc agcagggccu gcucgagaac cagccuaguu uccaccccg    | 5520 |
| caggcgugaa uaggguagc acuaagagagg agcucgaggc gcuuaccccg ucacgcacuc  | 5580 |
| cuagcagguc ggucucgaga accagccug ucuccaacc gccaggcgua aaaggguaga    | 5640 |
| uuacaagaga ggaguuugag gcguucguag cacaacaaca augacgguuu gaugcgggug  | 5700 |
| cauacauuu uuccuccgac accggucaag ggcauuuaca acaaaaauca guaaggcaaa   | 5760 |
| cggugcuau cgaaguggug uuggagagga ccgaauugga gauuucguau gccccgcgc    | 5820 |
| ucgaccaaga aaaagaaga uuacuacgca agaaauuaca guuaauucc acaccugua     | 5880 |
| acagaagcag auaccagucc aggaaggugg agaacaugaa agccauaaca gcuagacgua  | 5940 |
| uucugcaagg ccuagggcgu uauuuagaagg cagaagaaa aguggagugc uaccgaacc   | 6000 |
| ugcauccugu uccuuuguau ucaucuagug ugaaccgugc cuuuucaagc cccaaggugc  | 6060 |
| caguggaagc cuguacgcc auguugaaag agaacuucc gacugggcu ucuuacugua     | 6120 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| uuauuccaga guacgaugcc uauuuggaca ugguugacgg agcuucaugc ugcuuagaca  | 6180 |
| cugccaguuu uugcccugca aagcugcgca gcuuuccaaa gaaacacucc uauuuggaac  | 6240 |
| ccacaauacg aucggcagug ccuucagcga uccagaacac gcuccagaac guccuggcag  | 6300 |
| cugccacaaa aagaaaauugc aaugucacgc aaaugagaga auugcccgua uuggauucgg | 6360 |
| cggccuuuaa uguggaaugc uucaagaaau augcguguaa uaaugaauau ugaggaaacgu | 6420 |
| uuaaagaaaa ccccaucagg cuuacugaag aaaacguggu aaauuacauu accaaaauaa  | 6480 |
| aaggacaaaa agcugcugcu cuuuuugcga agacacauaa uuugaauaug ugcaggaca   | 6540 |
| uaccaaugga cagguuuuga auggacuuaa agagagacgu gaaagugacu ccaggaacaa  | 6600 |
| aacauacuga agaacggccc aagguacagg ugauccaggc ugccgaucgg cuagcaacag  | 6660 |
| cguaucugug cggaauccac cgagagcugg uuaggagauu aaaugcgguc cugcuuccga  | 6720 |
| acauucauac acuguuugau augucggcug aagacuuuga cgcuauuuaa gccgagcacu  | 6780 |
| uccagccugg ggauuguguu cuggaaacug acaucgcguc guuugauaaa agugaggacg  | 6840 |
| acgccauggc ucugaccgcg uuaaugauuc uggaagacuu agguugggac gcagagcugu  | 6900 |
| ugacgcugau ugaggcggcu uucggcgaaa uuucaucaau acauuugccc acuaaaacua  | 6960 |
| aaauuaaaau cgagccaug augaaaucug gaauguuccu cacacuguuu gugaacacag   | 7020 |
| ucauuacau uguaaucgca agcagagugu ugagagaacg gcuaaccgga ucaccaugug   | 7080 |
| cagcauucau uggagaugac aaauucguga aaggagucua aucggacaaa uuaauggcag  | 7140 |
| acaggugcgc caccugguug aaauuggaag ucaagauuau agaugcugug gugggcgaga  | 7200 |
| aagcgccuua uuucugugga ggguuuuuu ugugugacuc cgugaccggc acagcgugcc   | 7260 |
| guguggcaga ccccuaaaa aggcuguuuu agcuuggcaa accucuggca gcagacgaug   | 7320 |
| aacaugauga ugacaggaga agggcauugc augaagaguc aacacgcugg aaccgagugg  | 7380 |
| guauucuuuc agagcugugc aaggcaguag aaucaggua ugaaaccgua ggaacuucca   | 7440 |
| ucauaguau ggccaugacu acucuagcua gcaguguuaa aucauucagc uaccugagag   | 7500 |
| gggccccuau aacucucua ggcuaaccug aauggacuac gacauagucu aguccgcaa    | 7560 |
| gucuagcaua uguauggcca ccauggaaga ugccaagaac aucaagaagg gccugcccc   | 7620 |
| auucuacccc cuggaagaug gaacagccgg cgagcagcug cacaaggcca ugaagagaua  | 7680 |
| cgccugguug cccggcaca ucgccuucac cgaugccac aucgagguug acauaccua     | 7740 |
| cgccgaguac uucgagauga gcgugcggu ggccgaagcu augaagcgu acggccugaa    | 7800 |
| caccaaccac cggaucguc ugugcagcga gaacagccug caguucuua ugcccugcu     | 7860 |
| gggcgcccug uuuaucggag ugucguggc ccugccaac gacauuaca acgagcgca      | 7920 |
| gcugcugaac agcaugggca ucagccagcc caccguggug uucgugucca agaagggacu  | 7980 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gcagaaaauc cugaacgugc agaagaagcu gcccauauuc cagaaaauca ucaucaugga  | 8040 |
| cagcaagacc gacuaccagg gcuuccagag cauguacacc uucgugacca gccaucugcc  | 8100 |
| cccuggcuuu aacagaguacg acuucgugcc cgagagcuuc gaccgggaca agacaaucgc | 8160 |
| ccugaucaug aacagcagcg gcagcaccgg acugccuaaa ggcguggccc ugccucacag  | 8220 |
| aacugccugc gugcgguuua gccacgcccg ggaccuauuc uucggcaacc agaucauucc  | 8280 |
| cgacaccgcc auccugagcg uggugccuuu ccaccacggc uucggcaugu ucaccacccu  | 8340 |
| gggcuaccug aucugcgguu uccggguggu gcugauguac agauucgagg aagaacuguu  | 8400 |
| ccugcggagc cugcaggacu acaagaacca gagcgcccug cuggugccua ccuguuucag  | 8460 |
| cuucuuugcc aagagcaccu ugaucgauaa guacgaccug agcaaccugc acgagaucgc  | 8520 |
| cucuggcgga gccccccgu cuaaagaagu gggagaggcc guggccaagc gguuccaucu   | 8580 |
| gccuggcauc agacaggguu auggccugac cgagacaacc agcgccauuc ugaucacccc  | 8640 |
| cgaggcgac gauaagccug gcgcccuggg aaagguggug ccuucuuucg aggccaaaggu  | 8700 |
| gguggaccug gacaccggca agacacuggg cgugaaccag aggggcgaac ugugugugcg  | 8760 |
| gggaccuauu aucaugagcg gcuacgugaa caaccccag gccaccaacg ccugauuga    | 8820 |
| caaggauugc uggcugcaca gcggcgacau ugccuacugg gacgaggacg agcacuucuu  | 8880 |
| caucguggac cggcugaagu ccugaucaa guacaagggc uaccaggugg ccccagccga   | 8940 |
| gcuggaauuc auccugcugc agcaccccaa caucuucgau gccggcgugg caggacugcc  | 9000 |
| cgaugaugau gcuggcgaac ugccagccgc uguggguggug cuggaacacg gaaagaccu  | 9060 |
| gaccgagaaa gaaauugugg acuacuggc cagccaagug accaccgcca agaaacugag   | 9120 |
| aggcggcgug guguuugugg acgaggugcc aaagggccug acaggcaagc uggacgcccg  | 9180 |
| gaagaucaga gagauccga uuaaggccaa gaaaggcggc aagaucgccc uggauaggag   | 9240 |
| aaagagaggc ucuggcgaag gcagaggcag ccugcuuaca uguggcgacg uggaagagaa  | 9300 |
| ccccggaccu auggacuacg ugucccugcu gaaccagauu uggcagaagu accugaacag  | 9360 |
| cccuacacc accugucugu acauccccaa gccuaccgcc aaguacacac cucucuggg    | 9420 |
| cacauucug cccccgugc uguggaauug ccagcugagc uuugccggcu acaccgaguc    | 9480 |
| ugccgugaac agcacaaagg ccugggccaa acaggacgcc gcucagagaa uugccuggu   | 9540 |
| gcugcacaag gauggcggca ucccugaugg cuguagccug uaccugagac acagcagccu  | 9600 |
| guucgcccag agcgaggag aggaauccuu cagcaacuga ugagcggccg cgaauuggca   | 9660 |
| agcugcuuac auagaacug cggcgauugg caugccgccu uaaaauuuuu auuuuuuuu    | 9720 |

ucuuuuuuu uccgaaucgg auuuuuuuu uauuuuuu aaaaaaaaaa aaaaaaaaaa 9780

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 9840

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 9878

<210> 40

<211> 10217

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence encoding the RNA construct

<400> 40

atgggcggcg catgagagaa gccagacca attacctacc caaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaa actgatcgaa acggaggtgg acccatccga cagcatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccgacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct ggccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcggt ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccaccctt tttatgttta 600

agaacttggc tggagcatat ccatcactat ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaacat ccaacaatgt tctattctct gttagctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttact 840

tacgtggcaa gcaaattac acatgtcgtt gtgagactat agttagtgtc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgtgcta 960

cgatgcaccg cgagggatc ttgtgctgca aagtgcaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgacaaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccc 1200

tagtggccca ggcatttgc aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatgggtg ttgttgggt tttagaaggc 1320

|                                                                   |      |
|-------------------------------------------------------------------|------|
| acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg | 1380 |
|                                                                   |      |
| atttcactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa  | 1440 |
| caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg | 1500 |
| acgtacaaga agctaagtc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt  | 1560 |
| tgcgcgcagc tctaccacct ttggcagctg atgttgagga gccactctg gaagccgatg  | 1620 |
| tcgacttgat gttacaagag gctggggccg gctcagtga gacacctctg ggcttgataa  | 1680 |
| aggttaccag ctacgatggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg | 1740 |
| ctgtactcaa gagtgaaaaa ttatcttgca tccacctctt cgctgaacaa gtcatagtga | 1800 |
|                                                                   |      |
| taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg | 1860 |
| tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtga agtgccacca  | 1920 |
| ttgtgtacaa cgaacgtgag ttctgaaaca ggtacctgca ccatattgcc acacatggag | 1980 |
| gagcgtgaa cactgatgaa gaattattaca aaactgtcaa gccagcgag cacgacggcg  | 2040 |
| aatacctgta cgacatcgac aggaacagc gcgtcaagaa agaactagtc actgggctag  | 2100 |
| ggctcacagg cgagctggtg gatctctcct tccatgaatt cgctacgag agtctgagaa  | 2160 |
| cacgaccagc cgctccttac caagtacaa ccataggggt gtatggcgtg ccaggatcag  | 2220 |
|                                                                   |      |
| gcaagtctgg catcattaaa agcgcagtca caaaaaaga tctagtgtg agcgccaaga   | 2280 |
| aagaaaactg tgcagaaatt ataaggacg tcaagaaat gaaagggtg gacgtcaatg    | 2340 |
| ccagaactgt ggactcagt ctcttgaatg gatgcaaca cccgtagag accctgtata    | 2400 |
| ttgacgaagc ttttgcctgt catgcaggta ctctcagagc gctcatagcc attataagac | 2460 |
| ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc  | 2520 |
| tgaagtgtca ttttaaccac gagatttga cacaagtctt ccacaaaagc atctctcgcc  | 2580 |
| gttgactaa atctgtgact tcggctgtct caaccttgtt ttacgacaaa aaaatgagaa  | 2640 |
|                                                                   |      |
| cgacgaatcc gaaagagact aagattgtga ttgacctac cggcagtacc aaacctaagc  | 2700 |
| aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca | 2760 |
| aaggcaacga aataatgacg gcagctgcct ctcaagggt gaccgtaaa ggtgtgtatg   | 2820 |
| ccgttcggtg caaggtgaat gaaaatctc tgtacgcacc cactcagaa catgtgaacg   | 2880 |
| tcctactgac ccgcacggag gaccgatcg tgtggaaaac actagccggc gaccatgga   | 2940 |
| taaaaacact gactgccaag taccctggga atttactgc cacgatagag gagggtgcaag | 3000 |
| cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc | 3060 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca   | 3120 |
| tagacatgac cactgaacaa tggaacactg tggattatit tgaacaggac aaagctcact   | 3180 |
| cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg   | 3240 |
| gtctatitit tgcaccact gttccgttat ccattaggaa taatcactgg gataactccc    | 3300 |
| cgtcgctaa catgtacggg ctgaataaag aagtggctcg tcagctctct cgcaggtacc    | 3360 |
| cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc   | 3420 |
| gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag   | 3480 |
|                                                                     |      |
| tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg   | 3540 |
| gcagaactgt cctgggtggtc ggggaaaagt tgcctgtccc aggcaaaatg gttgactggt  | 3600 |
| tgtcagaccg gcctgaggct accttcagag ctgggtgga tttaggcatc ccaggatgatg   | 3660 |
| tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatc catcactatc    | 3720 |
| agcagtgtga agacatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc    | 3780 |
| tgaatcccgg cggaacctgt gtcagcatag gttatggta cgctgacagg gccagcgaaa    | 3840 |
| gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct   | 3900 |
|                                                                     |      |
| cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc   | 3960 |
| acaattctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg   | 4020 |
| aagccggatg tgcacctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag    | 4080 |
| gagtgtat taatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc     | 4140 |
| tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac   | 4200 |
| tgggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaaacttc aacaaagttt | 4260 |
| cggaggttga aggtgacaaa cagttggcag aggccttatga gtccatcgct aagattgtca  | 4320 |
|                                                                     |      |
| acgataacaa ttacaagtca gtagcgattc cactgttgte caccggcatc ttttcggga    | 4380 |
| acaaagatcg actaaccaa tcattgaacc atttgctgac agctttagac accactgatg    | 4440 |
| cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg   | 4500 |
| ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagt acagaacctg    | 4560 |
| atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca   | 4620 |
| caagcgatgg caaaacttct tcatatttgg aagggaacaa gtttcaccag gcggccaagg   | 4680 |
| atatagcaga aattaatgcc atgtggcccc ttgcaacgga ggccaatgag caggtatgca   | 4740 |
|                                                                     |      |
| tgtatatacct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg  | 4800 |
| aagcctccac accacctage acgtgcctt gcttgtgcat ccatgccatg actccagaaa    | 4860 |
| gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgttgtctca tcctttccat   | 4920 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct  | 4980 |
| caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccgtag   | 5040 |
| acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac  | 5100 |
| cattataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg   | 5160 |
|                                                                    |      |
| aagaagagga tagcataagt ttgctgtcag atggcccgcac ccaccaggtg ctgcaagtcg | 5220 |
| aggcagacat tcacgggccc ccctctgtat ctagctcatc ctggtccatt cctcatgcat  | 5280 |
| ccgactttga tgtggacagt ttatccatac ttgacacctt ggaggagct agcgtgacca   | 5340 |
| gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc  | 5400 |
| gaccggtgcc tgcgcctcga acagtattca ggaacctcc acatcccgt ccgcgcacaa    | 5460 |
| gaacaccgtc acttgacccc agcagggcct gctcgagaac cagcctagtt tccacccgc   | 5520 |
| caggcgtgaa tagggtgac actagagagg agctcgaggc gcttaccocg tcacgcactc   | 5580 |
|                                                                    |      |
| ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga  | 5640 |
| ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg  | 5700 |
| catacatctt ttctccgac accggtcaag ggcatctaca acaaaaatca gtaaggcaaa   | 5760 |
| cgggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgc | 5820 |
| tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta  | 5880 |
| acagaagcag ataccagtc aggaagggtg agaacatgaa agccataaca gctagacgta   | 5940 |
| ttctgcaagg cctaggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc   | 6000 |
|                                                                    |      |
| tgcatcctgt tctttgtat tcatctagt tgaaccgtgc cttttcaagc cccaaggctg    | 6060 |
| cagtgaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta   | 6120 |
| ttattccaga gtacgatgcc tatttgaca tggttgacgg agcttcatgc tgcttagaca   | 6180 |
| ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac  | 6240 |
| ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag  | 6300 |
| ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg  | 6360 |
| cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt  | 6420 |
|                                                                    |      |
| ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa  | 6480 |
| aaggacaaa agctgctgct ctttttgca agacacataa tttgaatatg ttgcaggaca    | 6540 |
| taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa  | 6600 |
| aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag  | 6660 |
| cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttcgca  | 6720 |
| acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact  | 6780 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| tccagcctgg ggatttgttt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg | 6840 |
| acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt | 6900 |
| tgacgtgat tgaggcggct ttcggcgaaa ttcatcaat acatttgccc actaaaacta   | 6960 |
| aatttaaatt cggagccatg atgaaatctg gaatgttctt cacactgttt gtgaacacag | 7020 |
| tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcacatgtg  | 7080 |
| cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag | 7140 |
| acaggtgcgc cacctgggtg aatatggaag tcaagattat agatgctgtg gtgggcgaga | 7200 |
| aagcgctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc  | 7260 |
| gtgtggcaga cccctaaaa aggcgtgtta agcttggcaa acctctggca gcagacgatg  | 7320 |
| aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg | 7380 |
| gtattctttt agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca | 7440 |
| tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag | 7500 |
| ggggccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgcaa  | 7560 |
| gtctagcata tgtatggcca ccatggaaga tgccaagaac atcaagaagg gccctgcccc | 7620 |
| attctacccc ctggaagatg gaacagccgg cgagcagctg cacaaggcca tgaagagata | 7680 |
| cgccctggtg cccggcacia tgccttcac cgatgcccac atcgaggtgg acatcaccta  | 7740 |
| cgccgagtac ttcgagatga gcgtgctggc ggccgaagct atgaagcgt acggcctgaa  | 7800 |
| caccaaccac cggatcgtcg tgtgcagcga gaacagcctg cagtcttcca tgcccgtgct | 7860 |
| gggcgcctg tttatcggag tggctgtggc cctgccaac gacatctaca acgagcgca    | 7920 |
| gctgtgaac agcatgggca tcagccagcc caccgtggtg ttcgtgtcca agaagggact  | 7980 |
| gcagaaaatc ctgaacgtgc agaagaagct gcccatcacc cagaaaatca tcatcatgga | 8040 |
| cagcaagacc gactaccagg gcttcagag catgtacacc ttcgtgacca gccatctgcc  | 8100 |
| ccctggcttc aacgagtacg acttcgtgcc cgagagcttc gaccgggaca agacaatcgc | 8160 |
| cctgatcatg aacagcagcg gcagaccgg actgcctaaa ggctgtggcc tgccctcacag | 8220 |
| aactgcctgc gtgcggttta gccacgcccg ggaccctatc ttcggcaacc agatcatccc | 8280 |
| cgacaccgcc atcctgagcg tggcgtctt ccaccacggc ttcggcatgt tcaccacct   | 8340 |
| gggctacctg atctgcggct tccgggtggt gctgatgtac agattcgagg aagaactgtt | 8400 |
| cctgcggagc ctgcaggact acaagatcca gagcgccctg ctggtgccta ccctgttcag | 8460 |
| cttctttgcc aagagcacc tgatcgataa gtacgacctg agcaacctgc acgagatcgc  | 8520 |

ctctggcgga gccccctgt ctaaagaagt gggagaggcc gtggccaagc ggttccatct 8580  
gcctggcatc agacagggt atggcctgac cgagacaacc agcgccattc tgatacccc 8640  
cgaggcgac gataagcctg gcgcctggg aaaggtggtg ccattcttcg aggccaaggt 8700  
ggtggacctg gacaccggca agacactggg cgtgaaccag aggggcgaac tgtgtgtgcg 8760  
gggacctatg atcatgagcg gctactgaa caaccccgag gccaccaacg ccctgattga 8820  
caaggatggc tggtgcaca gcggcgacat tgcctactgg gacgaggacg agcacttctt 8880  
catcgtggac cggctgaagt ccctgatcaa gtacaagggc taccaggtgg cccagccga 8940  
  
gctggaatct atcctgtgc agcaccccaa catcttcgat gccggcgtgg caggactgcc 9000  
cgatgatgat gctggcgaac tgccagccgc tgtggtggtg ctggaacacg gaaagacat 9060  
gaccgagaaa gaaatcgtgg actacgtggc cagccaagt accaccgcca agaaactgag 9120  
aggcgcgctg gtgtttgtgg acgagtgcc aaagggcctg acaggcaagc tggacgccc 9180  
gaagatcaga gagatcctga ttaaggccaa gaaaggcggc aagatcgccg tggatcgag 9240  
aaagagaggc tctggcgaag gcagaggcag cctgcttaca tgtggcgacg tggaagagaa 9300  
ccccggacct atggacccta cggacctgag ctacagcccc gacgagatca acaagctgat 9360  
  
cgagacaggc ctgaacaccg tggaatactt caccagccag caagtgaccg gcacaagcag 9420  
cctgggcaag aacacaattc ctccaggcgt gaccggcctg ctgacaaatg ctgccaggc 9480  
caagatccaa gagagcacca accaccagaa gggctctgtt ggaggcggag ccaagcctaa 9540  
gaagcccaga ctaagatcg ccacgtgcc cgccgacgat aagacagtgc ctggcaagcc 9600  
cattccta at cctctgtgg gcctcgacag caccctagc acacagacag tgctggatct 9660  
gagcggcaag acactgccta gcggcagcta taagggcgtg aagctggcca agttcggcaa 9720  
agaaaacctg atgaccgggt tcatcgagga acccagagag aacctatcg ccaccagctc 9780  
  
tcccatcgac ttcaagagag gcagagacac cggcggttc cacagaagag agtacagcat 9840  
tggtgggtc ggagatgaag tgaaagtac cgagtgtgc aacccagct gcagccctat 9900  
tacagccgc gctagaagat tctagtgac ctgtcaccag tgtctgtga cctgtagcga 9960  
gtgcgagcgg gacacatgat gagcgccgc gaattggcaa gctgcttaca tagaactcgc 10020  
ggcgattggc atgccgcctt aaaattttta tttattttt ctttctttt ccgaatcgga 10080  
ttttgtttt aatatttcaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 10140  
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 10200  
  
aaaaaaaaa aaaaaaa 10217

<210> 41

<211> 9878

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence encoding RNA construct

<400> 41

|                                                                   |      |
|-------------------------------------------------------------------|------|
| atgggcggcg catgagagaa gccagacca attacctacc caaaatggag aaagttcacg  | 60   |
| ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg | 120  |
| aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc | 180  |
| tggtttcaaa actgatcgaa acggaggtgg acccatccga cagatcctt gacattggaa  | 240  |
|                                                                   |      |
| gtgcgccccg ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat | 300  |
| gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg | 360  |
| aaataactga taaggaattg gacaagaaaa tgaaggagct ggccgccgtc atgagcgacc | 420  |
| ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc | 480  |
| aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag | 540  |
| ccaataaggg agttagagtc gcctactgga taggtttga caccaccctt tttatgttta  | 600  |
| agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa | 660  |
|                                                                   |      |
| cggtctgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt | 720  |
| ccattcttag aaagaagtat ttgaaacat ccaacaatgt tctattctct gttggctcga  | 780  |
| ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttact  | 840  |
| tacgtggcaa gcaaaattac acatgtcggg gtgagactat agttagtgtc gacgggtacg | 900  |
| tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgtctcta | 960  |
| cgatgcaccg cgagggttc ttgtctgca aagtgcaga cacattgaac ggggagaggg    | 1020 |
| tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaag actggcatac  | 1080 |
|                                                                   |      |
| tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta | 1140 |
| tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgccc  | 1200 |
| tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa | 1260 |
| ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc | 1320 |
| acaagataac atctatttat aagcggccgg ataccctaac catcatcaaa gtgaacagcg | 1380 |
| atttcactc attcgtctg cccaggatag gcagtaacac attggagatc gggctgagaa   | 1440 |
| caagaatcag gaaaatgta gaggagcaca aggagccgtc acctctcatt accgccgagg  | 1500 |
|                                                                   |      |
| acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt | 1560 |
| tgcgcgacg tctaccacct ttggcagctg atgttgagga gccactctg gaagccgatg   | 1620 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa | 1680 |
| aggttaccag ctacgatggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg | 1740 |
| ctgtactcaa gaggtaaaaa ttatcttgca tccacctctt cgctgaacaa gtcatagtga | 1800 |
| taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg | 1860 |
| tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtga agtgccacca  | 1920 |
| ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag | 1980 |
| gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gccacgcgag cagcagggcg | 2040 |
| aatacctgta cgacatcgac aggaacacgt gcgtcaagaa agaactagtc actgggctag | 2100 |
| ggctcacagg cgagctgggt gatcctccct tccatgaatt cgcctacgag agtctgagaa | 2160 |
| cacgaccagc cgctccttac caagtacaa ccataggggt gtatggcgtg ccaggatcag  | 2220 |
| gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtgtg agcgccaaga  | 2280 |
| aagaaaaactg tgcagaaatt ataaggacg tcaagaaat gaaagggtg gacgtcaatg   | 2340 |
| ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca cccgtagag accctgtata  | 2400 |
| ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac | 2460 |
| ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc  | 2520 |
| tgaagtga ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc   | 2580 |
| gttgactaa atctgtgact tcggctgtct caaccttgtt ttacgacaaa aaaatgagaa  | 2640 |
| cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaacg | 2700 |
| aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca | 2760 |
| aaggcaacga aataatgacg gcagctgcct ctcaagggt gaccgtaaa ggtgtgtatg   | 2820 |
| ccgttcggta caaggtgaat gaaaatctc tgtacgcacc cacctcagaa catgtgaacg  | 2880 |
| tcctactgac ccgcacggag gaccgatcg tgtggaaaac actagccggc gacccatgga  | 2940 |
| taaaaacact gactgccaag taccctggga atttactgc cacgatagag gagggtcaag  | 3000 |
| cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc | 3060 |
| agaataaggc aaacgttgtt tgggccaagg ctttagtgcc ggtgtgaag accgttgga   | 3120 |
| tagacatgac cactgaacaa tggaacactg tggattatgt tgaacggac aaagctcact  | 3180 |
| cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg | 3240 |
| gtctatcttc tgcaccact gtccgttat ccattaggaa taatcactgg gataactccc   | 3300 |
| cgtcgcctaa catgtacggg ctgaataaag aagtgtccg tcagctctct cgcaggtacc  | 3360 |
| cacaactgcc tcgggcagtt gccactggaa gactctatga catgaacact ggtacactgc | 3420 |
| gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag | 3480 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg  | 3540 |
| gcagaactgt cctgggtggc ggggaaaagt tgtccgtccc aggcaaatg gttgactggt   | 3600 |
|                                                                    |      |
| tgtcagaccg gcctgaggct accttcagag ctccgctgga tttaggcatc ccaggatgatg | 3660 |
| tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatc catcactatc   | 3720 |
| agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc  | 3780 |
| tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa  | 3840 |
| gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct  | 3900 |
| catttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtagc   | 3960 |
| acaattctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg  | 4020 |
|                                                                    |      |
| aagccggatg tgcacctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag   | 4080 |
| gagtatttat aaatgctgct aacagcaaaag gacaacctgg cggaggggtg tgcggagcgc | 4140 |
| tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac  | 4200 |
| tggtaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt   | 4260 |
| cggaggttga aggtgacaaa cagttggcag aggccttatga gtccatcgct aagattgtca | 4320 |
| acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttcggga   | 4380 |
| acaaagatcg actaaccaa tcattgaacc atttctgac agctttagac accactgatg    | 4440 |
|                                                                    |      |
| cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg  | 4500 |
| ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagt acagaacctg   | 4560 |
| atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaaggaag ggctacagca | 4620 |
| caagcgtatg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg  | 4680 |
| atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca  | 4740 |
| tgtatatact cgagaaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg  | 4800 |
| aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa  | 4860 |
|                                                                    |      |
| gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgtca tcctttccat   | 4920 |
| tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct  | 4980 |
| caccgaaagt gcctgcgtat attcatcaa ggaagtatct cgtggaaaca ccaccgtag    | 5040 |
| acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac  | 5100 |
| catttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg  | 5160 |
| aagaagagga tagcataagt ttgtgtcag atggcccgac ccaccaggtg ctgcaagtcg   | 5220 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat  | 5280 |
|                                                                    |      |
| ccgactttga tgtggacagt ttatccatac ttgacaccct ggaggagct agcgtgacca   | 5340 |
| gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc  | 5400 |
| gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgtt ccgcgcacaa  | 5460 |
| gaacaccgtc attgcaccc agcagggcct gctcgagaac cagcctagtt tccacccgc    | 5520 |
| caggcgtgaa tagggtgac actagagagg agctcgaggc gcttaccctg tcacgcactc   | 5580 |
| ctagcaggtc ggtctcgaga accagcctgg tctccaacc gccaggcgta aatagggtga   | 5640 |
| ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg  | 5700 |
|                                                                    |      |
| catacatctt ttctccgac accggtcaag ggcatctaca acaaaaatca gtaaggcaaa   | 5760 |
| cgggtgctatc cgaagtgggtg ttggagagga ccgaattgga gatttcgtat gccccgcgc | 5820 |
| tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaattccc acacctgcta | 5880 |
| acagaagcag ataccagtcc aggaagggtg agaacatgaa agccataaca gctagacgta  | 5940 |
| ttctgcaagg cctagggcatt tatttgaagg cagaaggaaa agtggagtgc taccgaacc  | 6000 |
| tgcatactgt tcttttgtat tcatctagt tgaacctgac cttttcaagc cccaaggctg   | 6060 |
| cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta  | 6120 |
|                                                                    |      |
| ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca  | 6180 |
| ctgccagttt ttgcccgcga aagctgcgca gctttccaaa gaaacactcc tatttggaa   | 6240 |
| ccacaatac atcggcagtg ccttcagcga tccagaacac gctccagaac gtctggcag    | 6300 |
| ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgtt ttggattcgg  | 6360 |
| cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaactg  | 6420 |
| ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa  | 6480 |
| aaggacaaa agctgctgct ctttttgca agacacataa tttgaatatg ttgcaggaca    | 6540 |
|                                                                    |      |
| taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa  | 6600 |
| aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag  | 6660 |
| cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttcga   | 6720 |
| acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact  | 6780 |
| tccagcctgg ggatttgtgt ctggaaactg acatcgctc gtttgataaa agtgaggacg   | 6840 |
| acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt  | 6900 |
| tgacgtgat tgaggcggtt ttcggcgaaa tttcatcaat acatttgccc actaaaacta   | 6960 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag | 7020 |
| tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcacatgtg  | 7080 |
| cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag | 7140 |
| acaggcgcgc cacctgggtg aatatggaag tcaagattat agatgctgtg gtgggcgaga | 7200 |
| aagcgctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc  | 7260 |
| gtgtggcaga cccctaaaa aggctgttta agcttggaac acctctggca gcagacgatg  | 7320 |
| aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg | 7380 |
|                                                                   |      |
| gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca | 7440 |
| tcatagttaa ggccatgact actctagctg gcagtgttaa atcattcagc tacctgagag | 7500 |
| gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgcaa  | 7560 |
| gtctagcata tgtatggcca ccatggaaga tgccaagaac atcaagaagg gccctgcccc | 7620 |
| attctacccc ctggaagatg gaacagccgg cgagcagctg cacaaggcca tgaagagata | 7680 |
| cggcctgggt cccggcacia tgccttcac cgatgcccac atcgaggtgg acatcaccta  | 7740 |
| cgccgagtac ttcgagatga gcgtgcggct ggccgaagct atgaagcgt acggcctgaa  | 7800 |
|                                                                   |      |
| caccaaccac cggatcgtcg tgtgcagcga gaacagcctg cagtcttcca tgcccgtgct | 7860 |
| gggcgcctg tttatcggag tggctgtggc ccctgccaac gacatctaca acgagcgca   | 7920 |
| gtctgtgaac agcatgggca tcagccagcc caccgtggtg ttcgtgtcca agaagggact | 7980 |
| gcagaaaatc ctgaacgtgc agaagaagct gcccatcacc cagaaaatca tcatcatgga | 8040 |
| cagcaagacc gactaccagg gcttcagag catgtacacc ttcgtgacca gccatctgcc  | 8100 |
| ccctggcttc aacgagtag acttcgtgcc cgagagcttc gaccgggaca agacaatcg   | 8160 |
| cctgatcatg aacagcagcg gcagcaccgg actgcctaaa ggctggccc tgcctcacag  | 8220 |
|                                                                   |      |
| aactgcctgc gtgcggttta gccacgcccg ggaccctatc ttcggcaacc agatcatccc | 8280 |
| cgacaccgcc atctgagcg tggcgccttt ccaccacggc ttcggcatgt tcaccacct   | 8340 |
| gggctacctg atctgcggct tccgggtggt gctgatgtac agattcgagg aagaactgtt | 8400 |
| cctgcggagc ctgcaggact acaagateca gagcgccctg ctggtgccta ccctgttcag | 8460 |
| cttctttgcc aagagcacc tgatcgataa gtacgacctg agcaacctgc acgagatcg   | 8520 |
| ctctggcgga gccccctgt ctaaagaagt gggagaggcc gtggccaagc ggttccatct  | 8580 |
| gcctggcatc agacagggt atggcctgac cgagacaacc agcgccattc tgatcacccc  | 8640 |
|                                                                   |      |
| cgaggcgac gataagcctg gcgccgtggg aaaggtggtg ccattcttcg aggccaaggt  | 8700 |
| ggtggacctg gacaccggca agacactggg cgtgaaccag agggcggaac tgtgtgtgcg | 8760 |
| gggacctatg atcatgagcg gctacgtgaa caaccccgag gccaccaacg ccctgattga | 8820 |

caaggatggc tggctgcaca gcggcgacat tgcctactgg gacgaggacg agcacttctt 8880  
catcgtggac cggttgaagt ccctgatcaa gtacaagggc taccaggtgg cccagccga 8940  
gctggaatct atcctgtgc agcaccccaa catcttcgat gccggcgtgg caggactgcc 9000  
cgatgatgat gctggcgaac tgccagccgc tgtggtggtg ctggaacacg gaaagacat 9060

gaccgagaaa gaaatcgtgg actacgtggc cagccaagt accaccgcca agaaactgag 9120  
aggcggcgtg gtgttttgg acgaggtgcc aaaggcctg acaggcaagc tggacgccc 9180  
gaagatcaga gagatcctga ttaaggccaa gaaaggcggc aagatcgccg tggatcggag 9240  
aaagagaggc tctggcgaag gcagaggcag cctgcttaca tgtggcgacg tggagagaa 9300  
ccccggacct atggactacg tgtccctgct gaaccagatt tggcagaagt acctgaacag 9360  
cccctacacc acctgtctgt acatcccaa gcctaccgcc aagtacacac ctctcgtggg 9420  
cacatctctg caccctgtgc tgtggaattg ccagctgagc ttgcccgtt acaccgagtc 9480

tgccgtgaac agcacaagg ccctggccaa acaggacgcc gctcagagaa ttgcctggct 9540  
gctgcacaag gatggcgga tccctgatgg ctgtagcctg tacctgagac acagcagcct 9600  
gttcgcccag agcgaggaag aggaatcctt cagcaactga tgagcggcgg cgaattggca 9660  
agctgcttac atagaactcg cggcgattgg catgccgct taaaattttt attttatttt 9720  
tcttttcttt tccgaatcgg attttgtttt taatatttca aaaaaaaaa aaaaaaaaa 9780  
aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa 9840  
aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaa 9878

<210> 42

<211> 588

<212> RNA

<213> human herpesvirus 2

<400> 42

gugagagacu gcuaccugau gggcuacugc cggacaagac ugaggaccuag acauaggggc 60  
agacugcugc agaucagcgg cggaacaugg gaugugcggc ugagaaacgc caucagagag 120  
guggaagccc acuucgagcc ugccgcugaa ccugugugug aacugcccug ucugaacgu 180  
agaagauacg gccugagug cgacgugggc aaccuggaaa caaauggcgg cagcaccagc 240  
gacgacgaga uuuccgaugc caccgacagc gacgauacac ugccagcca cagcgauaca 300  
gaaggcggac caucuccugc cggaagagag aaucugagu cugccucugg cggagccaau 360

gccgcuagac uggaauugca guucggcacc uucgacugga caagcgagga aggcucucag 420  
ccuuggcugu cugcuguggu ggccgauaca agcucugccg agagaagugg acuuccugcu 480

ccuggcgccu guagagcuac agaggcuccu gaaagagagg acggcugcag aaagaugcgg 540

uucccugccg ccuguccuua uccuugcggc cacacauuuc ugcggccc 588

<210> 43

<211> 1191

<212> RNA

<213> human herpesvirus 1

<400> 43

gugagagacu gcuaccugau gggcuacugc cggacaagac ugggaccuag aacauggggc 60

agacugcugc agaucagcgg cggaacaugg gaugugcggc ugagaaacgc caucagagag 120

guggaagccc acuucgagcc ugccgcugaa ccugugugug aacugcccug ucugaacgcu 180

agaagauacg gccugagug cgcgugggc aaccuggaaa caaauggcgg cagcaccagc 240

gacgacgaga uuuccgaugc caccgacagc gacgauacac uggccagcca cagcgauaca 300

gaaggcggac caucuccugc cggaagagag aauccugagu cugccucugg cggagccauu 360

gccgcuagac uggaauugca guucggcacc uucgacugga caagcgagga aggcucucag 420

ccuuggcugu cugcuguggu ggccgauauc agagacugcu accugauggg cuacugccgg 480

gcuagacugg ccccuagaac auggugcaga cugcugcaag ugucuggcgg cacauggggc 540

augcaccuga gaaacaccau cagagaggug gaagccagau ucgacgccac agccgagccu 600

gugugcaagc ugccuugucu ggaaacucgg agauacggcc ccgagugcga ccugagcaau 660

cuggaaaaucc accugagcgc caccagcgc gacgagauuu cugaugccac cgaccuggaa 720

gccgccggau cugaucuuac acuggccagc cagagcgaca ccgaggauug uccaucucca 780

gugacucugg aaacccuga gccuagagga ucucuggccg ugcgacugga agaугaguuc 840

ggcgaguucg acuggacccc ucaagaggga ucucagccuu ggcuugucg cgugguggcc 900

gauacaagca gcguggaaag acccggaaccu agcgauucug gugcuggcag agccgccgag 960

gauagaaaгу gccuggaugg cugccggaag augcgguucu cuaccgccug uccaуauccu 1020

ugcagcgaca ccuuccugc gccuugauaa acaagcucug ccgagagaag uggacuuccu 1080

gcuccuggcg ccuguagagc uacagaggcu ccugaaagag aggacggcug cagaaaug 1140

cgguucccug ccgccugucc uуauccuugc ggccacacau uucugcgcc c 1191

<210> 44

<211> 483

<212> RNA

<213> human herpesvirus 1

<400> 44

augagccaga cacagccucc agcuccaguu ggaccuggcg acccugaugu guaucugaag 60  
ggcgugccaa gcgccggcau gcauccuaga gguguucaug ccccuagagg acaccccaga 120

augaucucug gccuccuca gagaggcgac aacgaucagg cugcuggaca guguggcgau 180  
agcggacugc ugagaguggg cgccgauacc acaaucagca agccaucuga ggcugugcgg 240  
ccuccuacaa ucccagaac accuagagug ccccgcgagc caagagugcc uagaccuccu 300  
agagagccca gagaaccag agugccaagg gcucccagag auccuagagu ccucgggac 360  
ccuagggacc caagacaacc ugaucaccc agagagccuc ggagcccaag agagccaaga 420  
agcccuaggg aacccggac accaagaaca cccagggaac cuagaaccgc cagaggcagc 480  
gug 483

<210> 45  
<211> 549  
<212> RNA  
<213> Orf virus  
<400> 45

auggccugug aaugcgccag ccugauccug gaacugcuga gaaagagcga cgacaagcug 60  
cccgccaagc agaucgcca agagcugggc aucucuaagc acgaggcca cggcgagcug 120  
uaccggcugc uggaauucuga cgaagugugc ugcgaggagc gcaauccucc ucguugguuc 180  
guggaaugug cccuagcgc ucccaccgaa gaggacgaga auagcgacac cgagccuau 240  
gaaaccgagg ccggcugcga uacacuguuu ggcgagagaca ucgacaucau gaccagagc 300  
gccgugaucc ggcuagaugc ccugaauccu guguccgccg ugaacgaguu cugcaugaug 360

accacccggc cucuggaauu uugcgagaca agagccggcg gagaggauca cugcccaga 420  
uucaccugua ccaucaccau cagcggcaag guggugcug uugccgaugg cgccucuaag 480  
aaacuggcca gacacaccgc cuguagcagc gccugacaa uccugauca caacugcggc 540  
aucagcuuc 549

<210> 46  
<211> 504  
<212> RNA  
<213> pestivirus type 1  
<400> 46

auggaacuga ucaccaacga gcugcuguac aagaccuaca agcagaaacc cgugggcguc 60  
gaggaacccg uguaugauca agcuggcgac ccucuguuug gcgagagagg cgcuuucac 120

ccucagagca cacugaagcu gccccacaag cggggcgaaa gagaugugcc uaccaaccug 180  
gccagccugc cuaagagagg cgauugcaga accggcaaua gcagaggccc uguguccggc 240  
aucuaccuga aaccuggacc acuguucuaac caggacuaca agggaccgcu guaccacaga 300  
gccccucugg aacuguuuga agagggcagc augugcgaaa ccaccaagcg gaucggaaga 360  
gugaccggcu cugacggcaa gcuguaccac aucuacgugu gcaucgacgg cugcaucauc 420  
aucaagagcg ccaccagauc cuaccagcgg guguucagau gggugcacia cagacuggac 480  
ugcccucugu gggucaccag cugc 504

<210> 47

<211> 666

<212> RNA

<213> Simian parainfluenza virus 5

<400> 47

auggaccua cggaccugag cuucagcccc gacgagauca acaagcugau cgagacaggc 60  
cugaacaccg uggaauacuu caccagccag caagugaccg gcacaagcag ccugggcaag 120  
aacacaauc cuccaggcgu gaccggccug cugacaaaug cugccgaggc caagaucuaa 180  
gagagcacca accaccagaa gggcucuguu ggaggcggag ccaagccuaa gaagcccaga 240  
ccuaagaucg ccaucgugcc cgccgacgau aagacagugc cuggcaagcc cauuccuaau 300  
ccucugcugg gccucgacag caccuuagc acacagacag ugcuggaucu gagcggaag 360

acacugccua gcggcagcua uaagggcgug aagcuggcca aguucggcaa agaaaaccug 420  
augacccggu ucaucgagga acccagagag aaccuauug ccaccagcuc ucccaucgac 480  
uucaagagag gcagagacac cggcggcuuc cacagaagag aguacagcau uggcuggguc 540  
ggagaugaag ugaagugac cgaguggugc aacccagcu gcagcccuau uacagccgcc 600  
gcuagaagau ucgagugcac cugucaccag uguccuguga ccuguagcga gugcgagcgg 660  
gacaca 666

<210> 48

<211> 327

<212> RNA

<213> Middle East respiratory syndrome related coronavirus

<400> 48

auggacuacg uguccugcu gaaccagauu uggcagaagu accugaacag cccuacacc 60  
accugucugu acaucccaa gccuaccgcc aaguacacac cucucguggg cacaucucug 120  
caccgcgugc uguggaaug ccagcugagc uuugccggcu acaccgaguc ugccgugaac 180

agcacaaagg ccucggccaa acaggacgcc gcucagagaa uugccuggcu gcugcacaag 240  
gauggcggca ucccugaugg cuguagccug uaccugagac acagcagccu guucgcccag 300  
agcgaggaag aggaauccuu cagcaac 327

<210> 49  
<211> 1143  
<212> RNA  
<213> Langat virus  
<400> 49

guguucaagg acaaggugga caccaaggcu caagagccuc agccuggcac caagaucauc 60  
augagagccg ugaacgacug gcugcuggaa cggcugguca agaaaagcag accccggaug 120  
ugcagccggg aagaguuuuu cgccaaagug cggagcaaug ccgcucucgg agcuuggagu 180  
gacgagcaga acaaguggaa guccgccaga gaagccuggg aagaucacga guuuuggagc 240  
cugguggaag ccgagagaga gaggcaucug cagggaagau gugccacug cguguacaac 300  
augauaggga agagagagaa gaagcugggc gaguucggag uggccaaagg cagcagagcc 360

aucugguaua uugggcuggg cagccgcuuc cuggaauuug agggccuggg cuuccugaac 420  
gaggaucau gggcuagcag agccucuucu ggugcuggcg uggaaggcau cagccugaau 480  
uauucggcu ggcaccugaa gaaacuggcc ucucugucug gcggccuguu cuacgccgau 540  
gauacagccg gaugggacac aaagauacc aacgccgacc uggacgacga ggaacagauc 600  
cugagauaua uggacggcga ccacaaaaag cuggccgcca ccgugcugag aaaggccuau 660  
cacgccaagg ucgucagagu ggccagaccu aguagagaag gcggcugcgu gauggacauc 720  
aucaccagaa gggaccagcg cggcucuggc cagguuguga cauacgccc gaacaccauc 780

accaacauc aggugcagcu cgugcggaug auggaaggcg agggcgugau cgaaguggcc 840  
gacagccaua auccucggcu gcugagagug gaaaaguggc uggaagaaca cggcgaagaa 900  
cggcugagca gaaugcuggu guccggcgac gauuuguug ugcggcccgu ggacgacaga 960  
uucagcaagg ccuguacuu ucugaaugac auggccaaga ccagaaagga caccggcgag 1020  
ugggagccuu cuacaggcuu ugccagcugg gaagaagugc cuuucugcag ccaccacuuc 1080  
cacgagcugg ucaugaagga uggcagagcc cuggugugc ccugcagaga ucaggacgaa 1140  
cug 1143

<210> 50  
<211> 1605  
<212> RNA

<213> Artificial Sequence

<220><223> nsP1 RNA sequence

<400> 50

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| auggagaaag uucacguuga caucgaggaa gacagcccau uccucagagc uuugcagcgg | 60  |
| agcuucccg aguuugaggu agaagccaag caggucacug auaaugacca ugcuaaugcc  | 120 |
| agagcguuuu cgcaucuggc uucaaaacug aucgaaacgg agguggaccc auccgacacg | 180 |
| auccuugaca uuggaagugc gcccggccgc agaauuguau cuaagcacia guaucuugu  | 240 |
| aucuguccga ugagaugugc ggaagaucg gacagauugu auaaguauc aacuaagcug   | 300 |

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| aagaaaaacu gaaaggaaau aacugauaag gaauuggaca agaaaaugaa ggagcuggcc | 360 |
| gccgucauga gcgaccuga ccuggaaacu gagacuaugu gccuccacga cgacgagucg  | 420 |
| ugucgcuacg aagggaagu cgcuuuuac caggauuau acgcggguaga cggaccgaca   | 480 |
| agucucuauc accaagccaa uaaggaguu agagucgcu acuggauagg cuuugacacc   | 540 |
| acccuuuuu uguuaagaa cuuggcugga gcuaucgau cauacucua caacugggcc     | 600 |
| gacgaaaccg uguuaacggc ucguaacuaa ggccuaugca gcucugacgu uauggagcgg | 660 |
| ucacguagag ggauguccau ucuuagaaag aaguauuga aaccauccaa caauguucua  | 720 |

|                                                                  |      |
|------------------------------------------------------------------|------|
| uucucuguug gcucgaccu cuaccacgag aagagggacu uacugaggag cuggcaccug | 780  |
| ccgucuguau uucacuuacg uggaagcaa aauuacacau gucgguguga gacuauuguu | 840  |
| aguugcgacg gguacgcu uaaaagaaua gcuaucaguc caggccguga ugggaagccu  | 900  |
| ucaggcuau cugcuacgau gcaccgag ggauucuugu gcugcaaagu gacagacaca   | 960  |
| uugaacgggg agaggguc uuuuccgug ugcacguaug ugccagcuac auugugugac   | 1020 |
| caaaugacug gcuaucugc aacagaugc agugcgagc acgcgcaaaa acugcugguu   | 1080 |
| gggcuaacc agcguauagu cgucaacggu cgacccaga gaaacacca uaccaugaaa   | 1140 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| aauuaccuu ugcccguagu ggcccaggca uuugcuaggu gggcaaagga auuaaggaa   | 1200 |
| gaucaagaag augaaaggcc acuaggacua cgagauagac aguuagucgu gggguuguu  | 1260 |
| ugggcuuuu gaaggcacia gauaacauc auuuauaagc gcccggauac ccaaaccuac   | 1320 |
| aucaaaguga acagcgauuu ccacucauuc gugcugccca ggauaggcag uaacacauug | 1380 |
| gagaucgggc ugagaacaag aaucaggaaa auguuagagg agcacaagga gccgucaccu | 1440 |
| cucuuaccg ccgaggacgu acaagaagcu aagugcgag ccgaugaggc uaaggaggug   | 1500 |
| cgugaagccg aggaguugc cgacgucua ccaccuuugg cagcugaugu ugaggagccc   | 1560 |

|                                                   |      |
|---------------------------------------------------|------|
| acucuggaag ccgaugucga cuugauguua caagaggcug gggcc | 1605 |
|---------------------------------------------------|------|

<210> 51

<211> 2382

<212> RNA

<213> Artificial Sequence

<220><223> nsP2 RNA sequence

<400> 51

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ggcucagugg agacaccucg uggcuugaua aagguuacca gcuacgaugg cgaggacaag  | 60   |
| aucggcucuu acgcugugcu uucuccgcag gcuguacuca agagugaaaa auuauucuugc | 120  |
| auccaccuc ucgcugaaca agucauagug auaacacacu cuggccgaaa agggcguuau   | 180  |
| gccguggaac cauaccaugg uaaaguagug gugccagagg gacaugcaau acccguccag  | 240  |
| gacuaucaag cucugaguga aagugccacc auuguguaca acgaacguga guucguaaac  | 300  |
| agguaccucg accauauugc cacacaugga ggagcgcuca acacugauga agaauauuac  | 360  |
| aaaacuguca agcccagcga gcacgacggc gaauaccugu acgacaucga caggaaacag  | 420  |
| ugcguaaga aagaacuagu cacugggcua gggcucacag gcgagcuggu ggauccuccc   | 480  |
| uuccaugaa ucgccuacga gagucugaga acacgaccag ccgcuccuua ccaaguacca   | 540  |
| accuagggg uguauggcgu gccaggauca ggcaagucug gcaucauuua aagcgaguc    | 600  |
| acaaaaaag aucuaguggu gagcgccaag aaagaaaacu gugcagaaau uauaaggac    | 660  |
| gucaagaaaa ugaagggcgu ggacgucaau gccagaacug uggacucagu gcucuugaau  | 720  |
| ggaugcaaac accccguaga gaccucguau auugacgaag cuuuugcuug ucaugcaggu  | 780  |
| acucucagag cgcucauagc cauuaaaga ccuaaaaagg cagugcucug cggggauccc   | 840  |
| aaacagugcg guuuuuuuua caugaugugc cugaaaguc auuuuaacca cgagauuugc   | 900  |
| acacaagucu uccacaaaag caucucucgc cguugcacua aaucugugac uucggucguc  | 960  |
| ucaaccuugu uuucgacaa aaaaugaga acgacgauc cgaaagagac uaagauugug     | 1020 |
| auugacacua ccggcaguac caaaccuaag caggacgauc ucauucucac uuguuucaga  | 1080 |
| ggguggguga agcaguugca aaauaguuac aaaggcaacg aaauaagac ggcagcugcc   | 1140 |
| ucucaagggc ugacccguaa agguguguau gccguucggu acaaggugaa ugaaaauccu  | 1200 |
| cugucgcac ccaccucaga acaugugaac guccuacuga cccgcacgga ggaccgcauc   | 1260 |
| guguggaaaa cacuagccgg cgacccaugg auaaaaacac ugacugccaa guaccuggg   | 1320 |
| aaauucacug ccacgauaga ggaguggcaa gcagagcaug augccaucau gaggcacauc  | 1380 |
| uuggagagac cggacccuac cgacgucuuc cagaauaagg caaacgugug uugggccaag  | 1440 |
| gcuuuagugc cggugcugaa gaccgcuggc auagacauga ccacugaaca auggaacacu  | 1500 |
| guggauuuu uugaaacgga caaagcucac ucagcagaga uaguaaugaa ccaacuaugc   | 1560 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gugagguucu uuggacucga ucuggacucc ggucuaauuuu cugcaccac uguuccguua  | 1620 |
| uccauuagga auaaucacug ggauaacucc ccgucgccua acauguacgg gcugaauaaa  | 1680 |
| gaaguggucc gucagcucuc ucgcagguac ccacaacugc cucgggcagu ugccacugga  | 1740 |
| agagucuaug acaugaacac ugguacacug cgcaauuauug auccgcgcgu aaaccuagua | 1800 |
| ccuguaaaca gaagacugcc ucaugcuuuu guccuccacc auaaugaaca cccacagagu  | 1860 |
| gacuuuucu cauucgucag caaaugaag ggagaaacug uccugguggu cggggaaaag    | 1920 |
| uuguccgucc caggcaaaa gguugacugg uugucagacc ggccugaggc uaccuucaga   | 1980 |
| gcucggcugg auuuaggcau cccaggugau gugcccaau augacauau auuuguuaau    | 2040 |
| gugaggacc cauaaaaua ccaucacuau cagcagugug aagaccaugc cauaaagcuu    | 2100 |
| agcauguuga ccaagaaagc uugucugcau cugaaucccg gcggaaccug ugucagcaua  | 2160 |
| gguuaugguu acgcugacag ggccagcgaa agcaucauug gugcuauagc gcggcaguuc  | 2220 |
| aaguuuuccc ggguaugcaa accgaaaucc ucacuugaag agacggaagu ucuguuugua  | 2280 |
| uucuuugggu acgaucgcaa ggcccguacg cacaauucu acaagcuuc aucaaccuug    | 2340 |
| accaacauuu auacagguuc cagacuccac gaagccggau gu                     | 2382 |
| <210> 52                                                           |      |
| <211> 1671                                                         |      |
| <212> RNA                                                          |      |
| <213> Artificial Sequence                                          |      |
| <220><223> nsP3 RNA sequence                                       |      |
| <400> 52                                                           |      |
| gcaccucua aucaugguu gcgaggggau auugccacgg ccaccgaagg agugauuaa     | 60   |
| aaugcugcua acagcaaagg acaaccuggc ggaggggugu gcggagcgcu guauaagaaa  | 120  |
| uucccgaaa gcuucgauuu acagccgauc gaaguaggaa aagcgcgacu ggucaaaggu   | 180  |
| gcagcuuaac auaucauua ugccguagga ccaacuua acaaaguuc ggagguugaa      | 240  |
| ggugacaaac aguuggcaga ggcuuauag uccaucgcu agauuguaa cgauaacaau     | 300  |
| uacaagucag uagcgauucc acuguugucc accggcaucu uuuccgggaa caaagaucga  | 360  |
| cuaacccaau cauugaacca uuugcugaca gcuuuagaca ccacugaugc agauguagcc  | 420  |
| auauacugca gggacaagaa augggaaaug acucucaagg aagcaguggc uaggagagaa  | 480  |
| gcaguggagg agauaugcau auccgacgac ucuucaguga cagaaccuga ugcagagcug  | 540  |
| gugagggugc uuccgaagag uucuuuggcu ggaagggaagg gcuacgcac aagcgauggc  | 600  |
| aaaacuucu cauauuugga agggaccaag uuucaccagg cggccaagga uauagcagaa   | 660  |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| auuaaugcca uguggcccgug ugaacggag gccaaugagc agguaugcau guauauccuc  | 720  |
| ggagaaagca ugagcaguau uaggucgaaa ugccccgucg aagagucgga agccuccaca  | 780  |
| ccaccuagca cgcugccuug cuugugcauc caugccauga cuccagaaag aguacagcgc  | 840  |
| cuaaaagccu caguccaga acaaaauacu gugugcucau ccuuuccauu gccgaaguau   | 900  |
| agaauacug gugugcagaa gauccaauug ucccagccua uauuguucuc accgaaagug   | 960  |
| ccugcguaua uucauccaag gaaguaucuc guggaaacac caccgguaga cgagacuccg  | 1020 |
| gagccaucgg cagagaacca auccacagag gggacaccug aacaaccacc acuuauaacc  | 1080 |
|                                                                    |      |
| gaggauagaga ccaggacuag aacgccugag ccgaucauca ucgaagagga agaagaggau | 1140 |
| agcauaaguu ugcugucaga ugccccgacc caccaggugc ugcaagucga ggcagacauu  | 1200 |
| cacgggccgc ccucuguauc uagcucaucc ugguccauuc cucaugcauc cgacuuugau  | 1260 |
| guggacaguu uauccauacu ugacaccucg gaggagcua gcgugaccag cggggcaacg   | 1320 |
| ucagccgaga cuaacucuua cuucgcaaag aguauggagu uucuggcgcg accggugccu  | 1380 |
| gcgccucgaa caguauucag gaaccucca caucccguc cgcgcacaag aacaccguca    | 1440 |
| cuugcaccca gcagggccug cucgagaacc agccuaguuu ccaccccgcc aggcgugaau  | 1500 |
|                                                                    |      |
| aggguagauc cuagagagga gcucgaggcg cuuaccccg cagcacucc uagcaggucg    | 1560 |
| gucucgagaa ccagccuggu cuccaaccg ccaggcguaa auaggugau uacaagagag    | 1620 |
| gaguugagg cguucguagc acaacaaca ugacgguuug augcgggugc a             | 1671 |
| <210> 53                                                           |      |
| <211> 1821                                                         |      |
| <212> RNA                                                          |      |
| <213> Artificial Sequence                                          |      |
| <220><223> nsP4 RNA sequence                                       |      |
| <400> 53                                                           |      |
| uacauuuuu ccuccgacac cgucaaggg cauuuacaac aaaaucagu aaggcaaacg     | 60   |
| gugcuauccg aagugguguu ggagaggacc gaauuggaga uuucguaugc cccgcgccuc  | 120  |
|                                                                    |      |
| gaccaagaaa aagaagaauu acuacgcaag aaauuacagu uaaauccac accugcuaac   | 180  |
| agaagcagau accaguccag gaagguggag aacaugaaag ccuaaacagc uagacguauu  | 240  |
| cugcaaggcc uagggcuaau uuugaaggca gaaggaaaag uggagugcua ccgaaccug   | 300  |
| cauccguuc cuuuguauuc aucuagugug aaccgugccu uuucaagccc caaggucgca   | 360  |
| guggaagccu gaaacgcau guugaaagag aacuuuccga cuguggcuuc uuacugauuu   | 420  |
| auuccagagu acgaugccua uuuggacaug guugacggag cuucaugcug cuuagacacu  | 480  |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| gccaguuuuu gccucgcaaa gcugcgagc uuuccaaaga aacacuccua uuuggaaccc    | 540  |
| acaauacgau cggcagugcc uucagcgauc cagaacacgc uccagaacgu ccuggcagcu   | 600  |
| gccacaaaaa gaaaauugcaa ugucacgcaa augagagaaug ugcccguuuu ggauucggcg | 660  |
| gccuuuaaug uggaauugcuu caagaaauau gcguguaaua augaauauug ggaaacguuu  | 720  |
| aaagaaaacc ccaucaggcu uacugaagaa aacgugguaa auuacauuac caaaauaaaa   | 780  |
| ggaccaaag cugcugcucu uuugcggaag acacauuuu ugaauauguu gcaggacaua     | 840  |
| ccaauaggaca gguuuguaau ggacuuaaag agagacguga aagugacucc aggaacaaaa  | 900  |
| cauacugaag aacggcccaa gguacaggug auccaggcug ccgauccgcu agcaacagcg   | 960  |
| uauugugcg gaauccaccg agagcugguu aggagauuaa augcgguccu gcuuccgaac    | 1020 |
| auucauacac uguuugauau gucggcugaa gacuuugacg cuuuuauagc cgagcacuuc   | 1080 |
| cagccugggg auuguguucu ggaaacugac aucgcgucgu uugauaaaag ugaggacgac   | 1140 |
| gccauggcuc ugaccgcuu aaugauucug gaagacuuag guguggacgc agagcuguug    | 1200 |
| acgcugauug aggcggcuuu cggcgaaauu ucaucaauac auuugccac uaaaacuaaa    | 1260 |
| uuuaauucg gagccaugau gaaucugga auguuccua cacuguuugu gaacacaguc      | 1320 |
| auuaacauug uaaucgcaag cagaguguug agagaacggc uaaccggauc accaugugca   | 1380 |
| gcuuuauug gagaugacaa uaucugaaa ggagucuuu cggacaaau aauggcagac       | 1440 |
| aggugcgcca ccugguugaa uauggaaguc aagauuauag augcuguggu gggcgagaaa   | 1500 |
| gcgccuuuu ucuguggagg guuuuuuuug ugugacuccg ugaccggcac agcgugccgu    | 1560 |
| guggcagacc ccuaaaaaag gcuguuuuag cuuggcaaac cucuggcagc agacgaugaa   | 1620 |
| caugaugaug acaggagaag ggcauugcau gaagagucac cacgcuggaa ccgagugggu   | 1680 |
| auucuuucag agcugugcaa ggcaguagaa ucaagguaug aaaccguagg aacuuccauc   | 1740 |
| auaguuaugg ccaugacuac ucuagcuagc aguguuuuuu cauucagcua ccugagaggg   | 1800 |
| gccccuauaa cucucuacgg c                                             | 1821 |
| <210> 54                                                            |      |
| <211> 44                                                            |      |
| <212> RNA                                                           |      |
| <213> Artificial Sequence                                           |      |
| <220><223> 5' UTR conserved sequence element                        |      |
| <400> 54                                                            |      |
| augggcgcg caugagagaa gccagacca auuaccuacc caaa                      | 44   |
| <210> 55                                                            |      |

|                                                                    |                                                 |     |
|--------------------------------------------------------------------|-------------------------------------------------|-----|
| <211>                                                              | 237                                             |     |
| <212>                                                              | DNA                                             |     |
| <213>                                                              | Severe acute respiratory syndrome coronavirus 2 |     |
| <400>                                                              | 55                                              |     |
| atgatgcaa ctattttctt tgctggcata ctaattgtta cgactattgt ataccttaca   |                                                 | 60  |
| atagtgaac ttcttcaatt gtcattactt caggtgatgg cacaacaagt cctattttctg  |                                                 | 120 |
| aacatgacta ccagattggg ggttatactg aaaaatggga atctggagca aaagactgtg  |                                                 | 180 |
| ttgtattaca cagttacttc acttcagact attaccagct gtactcaact caattga     |                                                 | 237 |
| <210>                                                              | 56                                              |     |
| <211>                                                              | 237                                             |     |
| <212>                                                              | RNA                                             |     |
| <213>                                                              | Severe acute respiratory syndrome coronavirus 2 |     |
| <400>                                                              | 56                                              |     |
| augaugccaa cuauuuucuu ugcuggcaua cuauuuguua cgacuauugu auaccuuaca  |                                                 | 60  |
| auagugcaac uucuucaauu gucauuacuu caggugaugg cacaacaagu ccuauuucug  |                                                 | 120 |
| aacaugacua ccagauuggu gguuauacug aaaaauaggga aucuggagca aaagacugug |                                                 | 180 |
| uuguauuaca caguuacuuc acuucagacu auuaccagcu guacucaacu caauuga     |                                                 | 237 |
| <210>                                                              | 57                                              |     |
| <211>                                                              | 45                                              |     |
| <212>                                                              | DNA                                             |     |
| <213>                                                              | Artificial Sequence                             |     |
| <220><223>                                                         | 26S promoter                                    |     |
| <400>                                                              | 57                                              |     |
| gggcccttat aactctctac ggctaacctg aatggactac gacat                  |                                                 | 45  |