(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0074314
(43) 공개일자 2021년06월21일

(51) 국제특허분류(Int. Cl.)
C07K 14/005 (2006.01) *A61K 39/145* (2006.01)
A61P 31/16 (2006.01) *C12N 15/86* (2006.01)
(52) CPC특허분류
C07K 14/005 (2013.01)
A61K 39/145 (2013.01)
(21) 출원번호 10-2021-7013479
(22) 출원일자(국제) 2019년10월08일
심사청구일자 없음
(85) 번역문제출일자 2021년05월04일
(86) 국제출원번호 PCT/US2019/055125
(87) 국제공개번호 WO 2020/076775
국제공개일자 2020년04월16일
(30) 우선권주장
62/742,868 2018년10월08일 미국(US)

(71) 출원인
얀센 파마슈티칼즈, 인코포레이티드
미국 뉴저지 08560 타이터스빌 트렌턴-하버톤 로드 1125
(72) 발명자
디하트, 제이슨, 엘.
미국 캘리포니아 92121, 샌디에고, 메리필드 로우 3210
왕, 내태니얼, 스테픈
미국 캘리포니아 92121, 샌디에고, 메리필드 로우 3210
(뒷면에 계속)
(74) 대리인
강철중

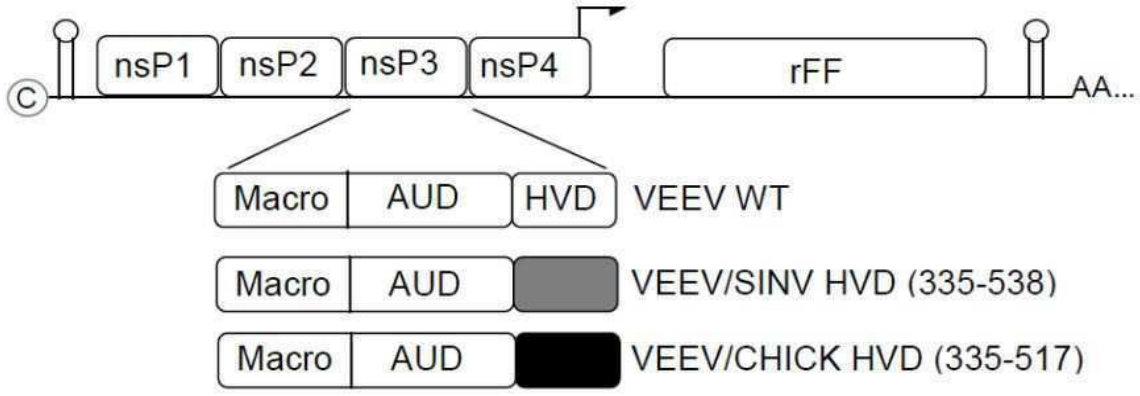
전체 청구항 수 : 총 39 항

(54) 발명의 명칭 **생체치료제의 투여를 위한 알파바이러스 기반 레플리콘**

(57) 요 약

본 발명은 포유류 내에 이종성 단백질 또는 펩티드를 투여하고 포유류로부터 감소된 면역 반응을 유발하거나 또는 면역 반응을 유발하지 않는 것에 유용한 RNA 레플리콘을 제공한다. 상기 RNA 레플리콘은 이종성 단백질 또는 펩티드, 뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2, 및 nsP4; 및 알파바이러스 nP3 단백질 매크로 도메인, 중앙 도메인, 및 초가변 도메인을 인코딩하는 RNA 서열을 가진다. 인코딩된 초바면 도메인은 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 아미노산 서열을 가질 수 있거나; 또는 뉴 월드 알파바이러스 nsP3 초가변 도메인의 부분으로부터 유래된 아미노산 서열 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 다른 부분을 가질 수 있다.

대 표 도 - 도2a



(52) CPC특허분류
A61P 31/16 (2018.01)
C12N 15/86 (2013.01)
C12N 2760/16134 (2013.01)
C12N 2770/36122 (2013.01)
C12N 2770/36143 (2013.01)
(72) 발명자
알리아흐마드, 파리나즈
미국 캘리포니아 92121, 샌디에고, 메리필드 로우 3210
미야케-스토너, 시게키
미국 캘리포니아 92037, 라 졸라, 노스 토레이 파인 로드 11149
캄루드, 커트, 이버
미국 66061, 캔자스, 올라트, 에스. 메스콰이트 스트리트 12243

명 세 서

청구범위

청구항 1

다음을 포함하는 RNA 레플리콘:

이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열;

5' 및 3' 알파바이러스 비번역 영역;

뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2, 및 nsP4로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및

알파바이러스 nsP3 매크로 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열;

알파바이러스 nsP3 중앙 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및

알파바이러스 nsP3 초가변 도메인(HVD; hypervariable domain)으로부터 유래된 다음을 포함하는 아미노산 서열을 인코딩하는 RNA 서열:

a. 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 아미노산 서열; 또는

b. 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분, 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분을 포함하는 아미노산 서열.

청구항 2

다음을 5' 부터 3' 말단까지의 순서로 포함하는, 알파바이러스 레플리콘 RNA:

(1) 알파바이러스 레플리콘의 복제를 지시하는 알파바이러스 5' 비번역 서열;

(2) 알파바이러스 비구조 단백질 nsP1, nsP2, nsP3, 및 nsP4을 인코딩하는 RNA 서열;

(3) 알파바이러스 서브게놈 프로모터 서열,

(4) 하나 이상의 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열, 및

(5) 알파바이러스 3' 비번역 서열,

여기서 nsP1, nsP2 및 nsP4는 하나 이상의 뉴 월드 알파바이러스로부터 유래하고,

상기 nsP3는 아미노 말단으로부터 카르복실 말단까지의 순서로, 알파바이러스 nsP3 매크로 도메인, 알파바이러스 nsP3 중앙 도메인 및 알파바이러스 nsP3 초가변 도메인(HVD)를 포함하고, 여기서

상기 매크로 도메인 및 중앙 도메인은 하나 이상의 뉴 월드 알파바이러스 및/또는 올드 월드 알파바이러스로부터 유래하고; 및

상기 HVD는 올드 월드 알파바이러스의 HVD를 포함하거나, 또는 HVD는 뉴 월드 알파바이러스로부터 유래된 HVD의 부분 및 올드 월드 알파바이러스로부터 유래된 HVD의 부분을 포함한다.

청구항 3

제1항 또는 제2항에 있어서,

상기 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 뉴 월드 알파바이러스로부터 유래된 것인, RNA 레플리콘.

청구항 4

제1항 또는 제2항에 있어서,

상기 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 올드 월드 알파바이러스로부터 유

래된 것인, RNA 레플리콘.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 알파바이러스 nsP3은 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 아미노산 또는 올드 월드 알파바이러스 nsP3 초가변 도메인의 아미노산을 포함하는, RNA 레플리콘.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 올드 월드 알파바이러스는 CHIKV, SINV, 및 SFV로 이루어진 군에서 선택되는 것인, RNA 레플리콘.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis, VEEV)인, RNA 레플리콘.

청구항 8

제1항 내지 제6항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis virus, VEEV), 웨스턴 이콰인 뇌염 바이러스(western equine encephalitis virus, WEEV), 및 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV)로 이루어진 군에서 선택되는 것인, RNA 레플리콘.

청구항 9

제8항에 있어서,

상기 올드 월드 알파바이러스는 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV), 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV), 로스 리버 바이러스(Ross River Virus, RRV), 사기야마 바이러스(Sagiyama virus, SAGV), 게타 바이러스(Getah virus, GETV), 미들버그 바이러스(Middleburg virus, MIDV), 베바루 바이러스(Bebaru virus, BEBV), 오뇽뇽 바이러스(O'nyong nyong virus, ONNV), 엔두무 바이러스(Ndumu, NDUV), 및 바르마 포레스트 바이러스(Barmah Forest virus, BFV)로 이루어진 군에서 선택되는 것인, RNA 레플리콘.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서,

상기 초가변 도메인은 FGDF(서열번호 18) 및 FGSF(서열번호 19)로 이루어진 군에서 선택되는 모티프를 가지는 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분을 포함하는, RNA 레플리콘.

청구항 11

제10항에 있어서,

상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 FGDF/FGDF(서열번호 20) 반복, FGSF/FGSF(서열번호 21) 반복, FGDF/FGSF(서열번호 22) 반복, 및 FGSF/FGDF(서열번호 23) 반복으로 이루어진 군에서 선택되는 반복을 포함하고; 및 추가로 상기 반복 서열은 적어도 10 및 25 이하의 아미노산에 의해 분리되는, RNA 레플리콘.

청구항 12

제11항에 있어서,

상기 반복 서열은 NEGEIESLSSELLT(서열번호 6), SDGEIDELSRRVTTESEPVL(서열번호 7), 및 DEHEVDALASGIT(서열번

호 8)로 이루어진 군에서 유래된 아미노산 서열에 의해 분리되는, RNA 레플리콘.

청구항 13

제10항에 있어서,

상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함하는 RNA 레플리콘

a. CHIKV nsP3 HVD의 아미노산 479-482 또는 497-500 또는 479-500 또는 335-517; 또는

b. SFV nsP3 HVD의 아미노산 451-454 또는 468-471 또는 451-471; 또는

c. SINV nsP3 HVD의 아미노산 490-493 또는 513-516 또는 490-516 또는 335-538.

청구항 14

제11항에 있어서,

상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함하는 RNA 레플리콘

a. CHIKV nsP3 HVD의 아미노산 479-500 또는 335-517; 또는

b. SFV nsP3 HVD의 아미노산 451-471; 또는

c. SINV nsP3 HVD의 아미노산 490-516.

청구항 15

제10항 내지 제14항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 VEEV nsP3 초가변 도메인의 아미노산 478-518을 포함하지 않는, RNA 레플리콘.

청구항 16

제10항 내지 제14항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 VEEV nsP3 초가변 도메인의 아미노산 478-545를 포함하지 않는, RNA 레플리콘.

청구항 17

제10항 내지 제14항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않는, RNA 레플리콘.

청구항 18

제17항에 있어서,

상기 올드 월드 알파바이러스는 CHIKV이고, 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 CHIKV의 아미노산 335-517을 포함하는, RNA 레플리콘.

청구항 19

제17항에 있어서,

상기 올드 월드 알파바이러스는 SINV이고, 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 SINV의 아미노산 335-538을 포함하는, RNA 레플리콘.

청구항 20

제10항 내지 제14항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 EEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 EEEV nsP3

초가변 도메인의 아미노산 531-547을 포함하지 않는, RNA 레플리콘.

청구항 21

제20항에 있어서,

상기 뉴 월드 알파바이러스는 EEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 EEEV nsP3 초가변 도메인의 아미노산 531-547을 포함하지 않고, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함하는, RNA 레플리콘:

a. CHIKV nsP3 HVD의 아미노산 479-500;

b. SFV nsP3 HVD의 아미노산 451-471; 또는

c. SINV nsP3 HVD의 아미노산 490-516.

청구항 22

제10항 내지 제14항 중 어느 한 항에 있어서,

상기 뉴 월드 알파바이러스는 WEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 WEEV nsP3 초가변 도메인의 아미노산 504-520을 포함하지 않는, RNA 레플리콘.

청구항 23

제22항에 있어서,

상기 뉴 월드 알파바이러스는 WEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 WEEV nsP3 초가변 도메인의 아미노산 504-520을 포함하지 않고, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함하는, RNA 레플리콘:

a. CHIKV nsP3 HVD의 아미노산 479-500;

b. SFV nsP3 HVD의 아미노산 451-471; 또는

c. SINV nsP3 HVD의 아미노산 490-516.

청구항 24

제1항 내지 제23항 중 어느 한 항에 있어서,

이종성 단백질을 인코딩하는 RNA 서열에 작동적으로 연결되고 이종성 단백질을 인코딩하는 RNA 서열의 번역을 제어하는 서브게놈 프로모터를 추가로 포함하는, RNA 레플리콘.

청구항 25

제1항 내지 제24항 중 어느 한 항에 있어서,

5' 캡 및 3' 폴리-A 테일을 추가로 포함하는, RNA 레플리콘.

청구항 26

제1항 내지 제25항 중 어느 한 항에 있어서,

상기 레플리콘은 양성 센스, 단일 가닥 RNA를 포함하는, RNA 레플리콘.

청구항 27

제1항 내지 제26항 중 어느 한 항에 있어서,

상기 레플리콘은 10-12 kb의 RNA를 포함하고 30-50 nm의 직경을 가지는, RNA 레플리콘.

청구항 28

제1항 내지 제27항 중 어느 한 항에 있어서,

상기 이종성 단백질은 생체치료제 단백질 또는 펩티드인, RNA 레플리콘.

청구항 29

제1항 내지 제28항 중 어느 한 항에 있어서,

상기 이종성 단백질은 항체인, RNA 레플리콘.

청구항 30

제1항 또는 제2항에 있어서,

상기 뉴 월드 알파바이러스는 VEEV이고, 상기 HVD는 VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않는 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분, 및 SINV nsP3 HVD의 아미노산 490-493 또는 513-516 또는 490-516 또는 335-538을 가지는 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분을 포함하는, RNA 레플리콘.

청구항 31

제30항에 있어서,

상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 SINV nsP3 HVD의 아미노산 490-516을 포함하는, RNA 레플리콘.

청구항 32

제30항에 있어서,

상기 올드 월드 알파바이러스는 SINV이고 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 SINV nsP3 HVD의 아미노산 335-538을 포함하는, RNA 레플리콘.

청구항 33

제1항 또는 제2항에 있어서,

상기 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열은 nsP1, nsP2 및 nsP4를 인코딩하는 RNA 서열에 작동적으로 연결된, RNA 레플리콘.

청구항 34

이종성 단백질 또는 펩티드를 인코딩하는 제1항 내지 제33항 중 어느 한 항의 RNA 레플리콘을 포유류에 투여하는 단계를 포함하는 포유류에 이종성 단백질 또는 펩티드를 투여하는 방법으로, 상기 이종성 단백질 또는 펩티드는 포유류 내에서 발현되는 것인, 방법.

청구항 35

제34항에 있어서,

상기 뉴 월드 알파바이러스는 VEEV이고, RNA 레플리콘은 제14항의 레플리콘인, 방법.

청구항 36

제34항에 있어서,

상기 RNA 레플리콘은 제19항의 레플리콘인, 방법.

청구항 37

제34항에 있어서,

상기 RNA 레플리콘은 제22항의 레플리콘인, 방법.

청구항 38

제34항에 있어서,

상기 RNA 레플리콘은 제25항의 레플리콘인, 방법.

청구항 39

다음을 포함하는 RNA 레플리콘:

이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열;

뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2 및 nsP4로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및

올드 월드 알파바이러스 nsP3 단백질로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열로, 상기 nsP3 단백질의 N 말단 및/또는 C 말단 측의 첫 1-6 아미노산은 뉴 월드 알파바이러스 서열로부터 유래된, RNA 서열.

발명의 설명

발명의 내용

[0001] **관련된 출원에 관한 상호 참조**

[0002] 본 출원은 2018년 10월 8일자로 출원된 미국 출원 제62/742,868에 대한 우선권을 주장하며, 상기 출원의 기재사항은 전체가 본 명세서에 참조로 통합된다.

[0003] **전자적으로 제출된 서열 목록에 관한 참조**

[0004] 본 출원은 서열 목록을 포함하며, 서열 목록은 형성일자 2019년 10월 7일 및 147 KB의 크기를 가지는 파일명 "689405.107WO_SL"의 ASCII 포맷 서열 목록으로서 EFS-Web을 통하여 전자적으로 제출되었다. 상기 EFS-Web을 통하여 제출된 서열 목록은 명세서의 일부이며 본 명세서에 전체가 참조로서 통합된다.

[0005] **배경기술**

[0006] 알파바이러스 기반 자가 증폭 RNA(레플리콘)은 수십년간 백신을 위한 범용 플랫폼으로 사용되어 왔다. 레플리콘을 이러한 플랫폼에 적합하게 만드는 두가지 특성은 1) 높고 장기적인 단백질 발현 및 2) 강력한 세포성, 체액성 및 점막성 면역을 유도하는 것을 돕는 플랫폼의 자가-애쥬번트 성질이다. 이들 특성은 레플리콘을 백신 플랫폼으로서 탁월하게 만드는 한편, 이들은 다른 영역, 특히 생체치료제(biotherapeutics)의 생체 내(in vivo) 발현을 위한 플랫폼으로서의 성능을 저해할 가능성을 가진다.

[0007] 의학에서 생체치료제의 사용이 증가하고 있다. 그러나 이러한 재조합적으로 생산된 단백질은 종종 이물질로 인식되고 면역 반응 및 항약물 항체(anti-drug antibodies, ADAs)의 유도를 유발하여 단백질의 치료 효과를 감소시킨다. 항약물 항체(ADA) 반응의 전개에 기여하는 요인은 다양하지만 생물학적 또는 생체치료제가 전달되는 염증 환경은 ADA 반응을 촉진 및/또는 향상시킬 수 있다. 따라서, 천연적으로 자가-애쥬번트이고 염증성인 레플리콘으로부터 생물학적 치료제를 전달하면 ADA 반응을 촉진하고 인코딩된 생체치료제의 임상적 효능을 감소시킬 위험이 있다. 레플리콘에서 발현된 이종성 단백질에 대한 면역 반응을 하향 조절하는 능력은 ADA 생성 위험을 줄이고 생체 내(in vivo)에서 생체치료제 발현을 위한 레플리콘의 유용성을 향상시킨다.

[0008] 따라서 바람직하지 않은 면역 반응을 유발할 위험을 줄이거나 제거하면서 인간 또는 동물에게 생체치료제 분자를 투여 할 수 있는 조성물 및 방법을 갖는 것이 유용할 것이다.

[0009] **발명의 요지**

[0010] 본 발명은 인간 또는 동물에 이종성(heterologous) 분자를 투여하고 이종성 분자에 대한 인간 또는 동물로부터의 면역 반응을 감소 또는 제거하는데 유용한 RNA 레플리콘을 제공한다. RNA 레플리콘은 이종성 분자(예컨대, 단백질 또는 펩티드)를 인코딩하는 RNA 서열(예컨대, 관심 유전자(gene of interest, GOI))을 가지며, RNA 서열은 뉴 월드(New World) 알파바이러스 비구조 단백질 nsP1, nsP2 및 nsP4; 및 알파바이러스 nsP3 단백질 매크로 도메인, 중앙 도메인 및 초가변 도메인(hypervariable domain, HVD)을 인코딩한다. 인코딩된 초가변 도메인은 올드 월드(Old World) 알파바이러스 nsP3 초가변 도메인으로부터 유래된 아미노산 서열을 가질 수 있거나, 뉴 월드 알파바이러스 nsP3 초가변 도메인의 부분으로부터 유래된 아미노산 서열, 및 올드 월드 알파바이러스 nsP3

초가변 도메인으로부터 유래된 다른 부분, 즉 키메라 nsP3 초가변 도메인을 가질 수 있다. 뉴 월드 알파바이러스에 기반한 레플리콘이 본 명세서에 기재된 바와 같이 변형될 때, 인코딩된 이종성 단백질 또는 펩티드에 의해 유발된 면역 반응이 감소되거나 제거된다는 것이 밝혀졌다.

[0011] RNA 레플리콘은 단백질 및 펩티드와 같은 생체치료제 분자의 투여에 유용하며, 여기서 본 발명의 레플리콘은 레플리콘에 의해 인코딩되는 생체치료제와 함께 인간 또는 동물에 투여되며, 인코딩된 생체치료제(예컨대, 이종성 단백질 또는 펩티드)는 인간 또는 동물에서 발현된다.

[0012] 첫 번째 양태에서 본 발명은 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열; 5' 및 3' 알파바이러스 비번역 영역; 뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2 및 nsP4로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및 알파바이러스 nsP3 매크로 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 알파바이러스 nsP3 중앙 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 아미노산 서열, 또는 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분, 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분을 포함하는 아미노산 서열을 갖는 초가변 도메인을 인코딩하는 RNA 서열을 갖는 RNA 레플리콘을 제공한다

[0013] 몇몇 구현예에서, 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 뉴 월드 알파바이러스로부터 유래되지만, 다른 구현예에서 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 올드 월드 알파바이러스로부터 유래된다. 다양한 구현예에서 올드 월드 알파바이러스는 CHIKV, SINV 및 SFV로 이루어진 군에서 선택된다. 뉴 월드 알파바이러스는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis virus, VEEV), 웨스턴 이콰인 뇌염 바이러스(western equine encephalitis virus, WEEV), 또는 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV) 일 수 있다. 다양한 구현예에서, 올드 월드 알파바이러스는 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV), 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV), 로스 리버 바이러스(Ross River Virus, RRV), 사기야마 바이러스(Sagiyama virus, SAGV), 게타 바이러스(Getah virus, GETV), 미들버그 바이러스(Middleburg virus, MIDV), 베바루 바이러스(Bebaru virus, BEBV), 오뇽뇽 바이러스(O'nyong nyong virus, ONNV), 엔두무 바이러스(Ndumu, NDUV), 및 바르마 포레스트 바이러스(Barmah Forest virus, BFV) 중 임의의 것일 수 있다.

[0014] 몇몇 구현예에서, 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 FGDF (서열번호 18) 및 FGSF (서열번호 19)로 이루어진 군에서 선택된 모티프를 포함한다. 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 FGDF/FGDF (서열번호 20) 반복, FGSF/FGSF (서열번호 21) 반복, FGDF/FGSF (서열번호 22) 반복, 및 FGSF/FGDF (서열번호 23) 반복으로 이루어진 군에서 선택되는 반복을 포함하고; 및 상기 반복 서열은 적어도 10 및 25 이하의 아미노산에 의해 분리될 수 있다. 일부 구현예에서 반복 서열은 NEGEIESLSSELLT (서열번호 6), SDGEIDELSRRVTTESEPVL (서열번호 7) 및 DEHEVDALASGIT (서열번호 8)로 이루어진 군에서 유래된 아미노산 서열에 의해 분리된다.

[0015] RNA 레플리콘의 임의의 구현예에서, 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 CHIKV nsP3 HVD의 아미노산 479-482 또는 497-500 또는 479-500 또는 335-517 중 임의의 것; 또는 SFV nsP3 HVD의 아미노산 451-454 또는 468-471 또는 451-471 중 임의의 것; 또는 SINV nsP3 HVD의 아미노산 490-493 또는 513-516 또는 490-516 또는 335-538을 가질 수 있다. 이들 구현예 중 임의의 것(또는 본 명세서에 기재된 임의의 구현예에서)에서 뉴 월드 알파바이러스는 VEEV 일 수 있고 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 VEEV nsP3 초가변 도메인의 아미노산 478-518을 포함하지 않으며; VEEV nsP3 초가변 도메인의 아미노산 478-545를 포함하지 않거나; VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않는다. 다른 구현예에서, 뉴 월드 알파바이러스는 EEEV 일 수 있고 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 EEEV 초가변 도메인의 아미노산 531-547을 포함하지 않는다. 또는 뉴 월드 알파바이러스는 WEEV 일 수 있고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 WEEV 초가변 도메인의 아미노산 504-520을 포함하지 않는다.

[0016] 임의의 구현예에서, RNA 레플리콘은 이종성 단백질을 인코딩하는 RNA 서열에 작동적으로 연결되고 이를 제어하는 서브-게놈 프로모터를 가질 수 있다. RNA 레플리콘은 또한 5' 캡 및 3' 폴리-A 테일을 가질 수 있다. RNA 레플리콘은 양성 센스, 단일 가닥 RNA를 가질 수 있다. 다양한 구현예에서 RNA 레플리콘은 10-12 kb의 RNA를 가질 수 있고/있거나 30-50 nm의 직경을 가질 수 있다.

[0017] 다양한 구현예에서 이종성 단백질은 예를 들어 항체 또는 조작된 키메라 항체 또는 항체 단편, 항원성 폴리펩티드 또는 임의의 다른 치료적 또는 면역원성 폴리펩티드 또는 펩티드일 수 있는 생체치료제 단백질 또는 펩티드이다.

[0018] 레플리콘의 몇몇 특정 구현예에서, 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않으며, 올드 월드 알파바이러스 nsP3 초가변 도메인은 SINV nsP3 HVD의 아미노산 490-516을 포함하고; 또는 올드 월드 알파바이러스는 SINV이고 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분은 SINV nsP3 HVD의 아미노산 335-538을 포함한다.

[0019] 임의의 구현예에서, 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열은 nsP1, nsP2 및 nsP4를 인코딩하는 RNA 서열에 작동적으로 연결될 수 있다.

[0020] 또 다른 양태에서 본 발명은 포유 동물에게 이종성 단백질 또는 펩티드를 투여하는 방법을 제공한다. 상기 방법은 이종성 단백질 또는 펩티드를 인코딩하는 본 명세서에 기재된 RNA 레플리콘을 포유류에 투여하는 것을 포함하고, 여기서 이종성 단백질 또는 펩티드는 포유류 내에서 발현된다. RNA 레플리콘은 본 명세서에 기재된 임의의 것일 수 있다.

[0021] 또 다른 양태에서, 본 발명은 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열; 뉴 월드 알파바이러스 비 구조 단백질 nsP1, nsP2 및 nsP4로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열; 및 올드 월드 알파바이러스 nsP3 단백질로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열을 갖는 RNA 레플리콘을 제공하며, 여기서 nsP3 단백질의 N-말단 및/또는 C-말단 측에 있는 처음 1-6 아미노산은 뉴 월드 알파바이러스로부터 유래된 것이다. 따라서 상기 1-6 아미노산이 nsP2와 nsP3 사이의 접합부에 존재할 수 있고; 또는 1-6 아미노산이 nsP3와 nsP4 사이의 접합부에 존재할 수 있다. 다양한 구현예에서 올드 월드 알파바이러스는 본 명세서에 기술 된 임의의 것일 수 있다. 뉴 월드 알파바이러스가 VEEV인 경우 nsP2/nsP3 접합부 서열은 LHEAGC/APSY (서열번호 12)일 수 있고; 접합부가 nsP3/nsP4 접합부인 경우 서열은 RFDAGA/YIFS(서열번호 13) 일 수 있다. 임의의 구현예에서, 끝에서 두번째 글리신(단일 문자 코드 "G"로도 지칭됨)은 보존될 수 있고 나머지 nsP3 아미노산은 본 명세서에 기재된 바와 같이 가변적이다. 접합부 서열은 선택적으로 정지 코돈(TGA)이 선행될 수 있으며, 이는 리드스루(readthrough) 정지 코돈일 수 있다. 뉴 월드 알파바이러스가 EEEV인 다른 구현예에서, nsP2/nsP3 접합부 서열은 QHEAGR/APAY(서열번호 14)일 수 있고, 끝에서 두번째 G는 보존된다. 뉴 월드 알파바이러스가 EEEV 인 경우 nsP3/nsP4 접합부에서의 서열은 RYEAGA/YIFS(서열번호 15)일 수 있고, 끝에서 두번째 글리신은 선택적으로 보존될 수 있는 한편 나머지 nsP3 아미노산은 본 명세서에 기재된 바와 같이 가변적이다. 이러한 서열은 또한 리드스루(read-through) 정지 코돈 (TGA)이 선행될 수 있다. 다른 구현예에서 뉴 월드 알파바이러스는 WEEV이고, nsP2/nsP3 접합부 서열은 RYEAGR/APAY(서열번호 16)일 수 있고, 끝에서 두번째 G는 보존되는 한편 nsP2/nsP3 접합부의 나머지 아미노산은 본 명세서에 기재된 바와 같이 가변적이다. WEEV의 nsP3/nsP4 접합부에 대해, 서열은 RYEAGA/YIFS(서열번호 17)일 수 있으며, 끝에서 두번째 글리신은 보존되고 나머지 nsP3 아미노산은 본 명세서에 기재된 바와 같이 가변적이다; 이러한 서열은 리드스루(read-through) 정지 코돈(TGA)이 선행될 수도 있다. 다양한 구현예에서 서열번호 12-17의 서열은 또한 N-말단 및/또는 C-말단 측에 1개 또는 2개 또는 3개의 치환을 포함할 수 있다.

[0022] 상술한 본 발명의 요지는 제한적이지 않으며 본 발명의 다른 특징 및 이점은 본 발명에 대한 다음의 상세한 설명 및 청구범위로부터 명백해질 것이다. 섹션 제목 또는 하위 제목은 독자의 편의를 위해서만 제공되며 논의에서 벗어나거나 반드시 완전히 새로운 주제 영역을 나타내지는 않는다. 모든 주제는 섹션 제목 또는 하위 제목 아래에서 논의하거나 공개할 수 있다.

도면의 간단한 설명

[0023] 전술한 요지 및 본 발명의 다음의 상세한 설명은 첨부된 도면과 함께 읽을 때 더 잘 이해될 것이다. 본 발명은 도면에 도시된 정확한 실시예로 제한되지 않음을 이해해야 한다.

도 1은 야생형 알파바이러스의 도시적 예를 제공한다. 바이러스 입자는 240개 모노머로 이루어진 T=4 정이십면체 대칭을 갖는 65-70 nm 직경의 캡시드를 갖는 피복된(enveloped), 구형, 정이십면체로 표시된다. 외피(envelope)는 80개의 스파이크를 포함하며 각 스파이크는 E1/E2 단백질의 삼량체이다.

도 2a-도 2c: 도 2a는 적색 반딧불이 루시퍼라제(red firefly luciferase, rFF)를 인코딩하는 VEEV 기반 알파바이러스 레플리콘의 도시적 예시이다. 3개의 구현예가 묘사된다: nsP3의 야생형 VEEV HVD를 갖는 것(WT), SINV의 HVD의 부분을 갖는 VEEV/SINV 하이브리드(VEEV/SINV)(VEEV HVD의 아미노산 잔기 335-538을 SINV HVD의 아미노산 335-538로 치환하여); 및 CHIKV HVD의 부분을 갖는 또 다른 하이브리드(VEEV/CHIKV)(VEEV HVD의 아미노산 잔기 335-518을 CHIKV HVD의 아미노산 335-517로 치환하여). 도 2b 및 2c는 돌연변이 nsP3 단백질을 포함하는

레플리콘이 동일한 수준으로 복제되고(도 2b) 야생형 nsP3를 포함하는 레플리콘과 동일한 수준의 rFF(도 2c)를 발현함을 보여주는 도시적 예시이다.

도 3a-도 3b: 도 3a는 생체 내(in vivo) 루시퍼라제 활성을 모니터링한 결과를 나타내는 그래프이며 총 플럭스(flux)로 보고된다. 식염수 내의 도 2a에 도시된 3개의 VEEV 기반 알파바이러스 레플리콘 RNA 각각 10ug를 BALB/c 마우스의 대퇴사두근에 근육 내로 전달하였으며, 여기서 α.SGI.rFF는 VEEV WT를 인코딩하고, α.SGI.SINV.rFF는 VEEV/SINV를 인코딩하고 α.SGI.CHIKV.rFF는 VEEV/CHIKV를 인코딩한다. 도 3b는 동일하지만 1ug의 레플리콘 RNA로 모니터링한 결과를 나타낸다. nsP3의 돌연변이 형태를 발현하는 레플리콘은 야생형 nsP3를 갖는 레플리콘과 유사한 수준의 루시퍼라제 활성을 나타냈다.

도 4는 H5N1 인플루엔자 바이러스로부터 HA를 발현하는 VEEV 기반 레플리콘의 생체 내 연구 결과를 나타내는 플롯 및 막대 그래프이다. 데이터는 VEEV/CHIKV HVD 키메라를 인코딩하는 레플리콘이 야생형 HVD를 발현하는 레플리콘에 비해 HA 특이적 IgG 역가를 유발하지 않았음을 나타낸다.

도 5a-도 5b: 도 5a는 H5N1 HA를 발현하는 표지된 VEEV 기반 레플리콘으로 면역화된 BALB/c 마우스에서 HA 특이적 단기 이펙터 CD8+ T 세포(SLEC)의 빈도를 나타내는 그래픽 포맷의 플롯을 제공한다. 도 5b는 H5N1 HA를 발현하는 표지된 VEEV 기반 레플리콘으로 면역화된 BALB/c 마우스에서 HA 특이적 메모리 전구체 이펙터 CD8+ T 세포(MPEC)의 빈도를 나타내는 그래픽 포맷의 플롯을 제공한다. WT는 VEEV의 TC-83 균주로부터 유래된 비변형 레플리콘 백본을 지칭한다. SGI는 인터페론 내성을 갖도록 변형된 레플리콘 백본을 지칭한다.

도 6은 뉴 월드 및 올드 월드 알파바이러스의 대표적 구성원의 nsP3 단백질의 도메인 구조 및 서열 정렬의 일부를 제공한다. nsP3 단백질의 도식적 표현은 세 가지 예측된 구조 도메인인 매크로 도메인, 알파 도메인 및 HVD를 나타낸다. 상이한 알파바이러스의 nsP3 단백질의 서열 정렬은 Clustal Omega를 사용하여 수행되었다. 도메인 서열은 도식 프레젠테이션에 사용한 것과 동일한 색상으로 밑줄이 그어져 있다. 예시된 nsP3 단백질 서열은 다음 바이러스로부터 유래했다: SFV(GenBank accession no. NP_740667.1) (표시된 단백질은 서열번호 24), SINV (GenBank accession no. P03317.1) (표시된 단백질은 서열번호 25), CHIKV (GenBank accession no. NP_690588.1) (표시된 단백질은 서열번호 26), VEEV (GenBank accession no. P27282.2) (표시된 단백질은 서열번호 27) 및 EEEV (GenBank accession no. Q4QXJ8.2) (표시된 단백질은 서열번호 28). 이미지는 Foy et al., Journal of Virology, Vol. 87, No. 4, pp. 1997-2010 (2013)로부터 가져왔다.

도 7은 다양한 뉴 월드 및 올드 월드 바이러스의 인코딩된 nsP3 단백질의 HVD에서 상이한 영역을 나타내는 그림을 제공한다(Gotte et al., Viruses, 2018, 10, 105의 Fig. 2로부터 재현됨). 도면에 사용된 nsP3 서열의 Uniprot 항목은 MAYV (Q8QZ73), RRV (P13887), SFV (P08411), CHIKV (Q8JUX6), ONNV (Q8QZ73), BFV (P87515), SINV (P03317), VEEV (P36328), EEEV (Q4QXJ8), WEEV (P13896)이다. G3BP 결합 부위는 다음의 올드 월드 바이러스 nsP3 단백질에 존재한다: MAYV의 경우 아미노산 470-473; RRV의 경우 아미노산 512-515 및 523-526; SFV의 경우 아미노산 451-454 및 468-471; CHIKV의 경우 아미노산 479-482 및 497-500; ONNV의 경우 아미노산 519-522 및 537-540; BFV의 경우 아미노산 429-432 및 447-450; SINV의 경우 아미노산 490-493 및 513-516. 뉴 월드 P1234 바이러스 단백질 G3BP(및 FXR) 결합 부위는 다음과 같이 존재한다: VEEV의 경우 아미노산 478-545는 FXR 결합 부위를 가지고; EEEV의 경우 아미노산 471-483은 G3BP 결합 부위를 갖고 아미노산 531-547은 FXR 결합 부위를 인코딩하고; WEEV의 경우 아미노산 504-520은 FXR 결합 부위를 가진다.

발명을 실시하기 위한 구체적인 내용

[0024] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 일반적으로 이해되는 것과 동일한 의미를 갖는다. 그렇지 않은 경우, 본 명세서에서 사용되는 특정 용어는 명세서에서 설정된 의미를 갖는다. 본 명세서에 인용된 모든 특허, 공개된 특허 출원 및 간행물은 마치 본 명세서에 완전히 설명된 것처럼 참조로 통합된다. 본 명세서 및 첨부된 청구범위에서 사용되는, 단수 형태 "a", "an" 및 "the"는 문맥이 명백하게 달리 지시하지 않는 한 복수 참조를 포함한다는 점에 유의해야 한다.

[0025] 달리 언급되지 않는 한, 본 명세서에 기재된% 서열 동일성 또는% 서열 동일성 범위와 같은 임의의 수치는 용어 "약"에 의해 모든 경우에 변형되는 것으로 이해되어야 한다. 따라서 수치는 일반적으로 인용된 값의 ±10%를 포함한다. 예를 들어, 10 mg의 용량은 9 mg 내지 11 mg을 포함한다. 본 명세서에서 사용된 바와 같이, 수치 범위의 사용은 문맥상 달리 명시적으로 나타내지 않는 한, 이러한 범위 내의 정수 및 값의 분수를 포함하여 모든 가능한 하위 범위, 그 범위 내의 모든 개별 수치 값을 명시적으로 포함한다.

[0026] 본 명세서 및 이하의 청구항 전반에서, 문맥상 달리 요구하지 않는 한, "포함하다(comprise)" 라는 단어 및 "포함하다(comprises)" 및 "포함하는(comprising)"과 같은 변형은 명시된 정수 또는 단계 또는 정수 또는 단계의 그룹을 포함하지만 임의의 다른 정수 또는 단계 또는 정수 또는 단계의 그룹을 배제하지 않는 것을 암시하는 것으로 이해될 것이다. 본 명세서에서 사용된 용어 "포함하는(comprising)"은 용어 "포함하는(containing)" 또는 "포함하는(including)" 또는 때로는 용어 "갖는(having)"과 함께 사용될 때 대체될 수 있다.

[0027] 본 명세서에서 "~로 이루어진다(consisting of)"를 사용할 때 청구항 구성요소에서 특정되지 않은 임의의 요소, 단계 또는 성분을 배제한다. 본 명세서에 사용될 때, "본질적으로 이루어진(consisting essentially of)"는 청구항의 기본 및 신규한 특성에 실질적으로 영향을 미치지 않는 재료 또는 단계를 배제하지 않는다. 본 발명의 양태 또는 구현예의 맥락에서 사용될 때마다, 전술한 용어 "포함하는(comprising)", "포함하는(containing)", "포함하는(including)" 및 "갖는(having)" 중 임의의 것은 개시의 범위를 변경하기 위하여 "구성되는(consisting of)" 또는 "본질적으로 이루어진(consisting essentially of)"의 용어로 대체할 수 있다.

[0028] 하나의 일반적인 측면에서, 본 발명은 이종성 단백질 또는 펩티드를 인코딩하는 RNA 레플리콘 및 RNA 레플리콘을 사용하여 인간 또는 동물에게 투여하는 방법에 관한다. 본 발명의 RNA 레플리콘은 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열, 및 뉴 월드 알파바이러스 nsP1, nsP2 및 nsP4 단백질로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열을 포함한다. 상기 레플리콘은 또한 알파바이러스 nsP3 매크로 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열 및 알파바이러스 nsP3 중앙 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열을 갖는다. 본 발명의 RNA 레플리콘은 추가로 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 전체적으로 유래된 아미노산 서열을 인코딩하는 RNA 서열, 또는 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분, 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래된 부분을 갖는 아미노산 서열을 추가로 가지며, 즉, 상기 HVD는 하이브리드 또는 키메라 뉴 월드/올드 월드 서열일 수 있다. 본 명세서에서 사용 된 "폴리펩티드", "펩티드"또는 "단백질"은 폴리펩티드를 형성하기 위해 펩티드 결합에 의해 연결된 적어도 2개의 아미노산 잔기를 포함하는 분자를 의미한다.

[0029] 본 명세서에서 사용되는, 알파바이러스 nsP3 또는 알파바이러스 nsP3 초가변 도메인(HVD)과 관련하여, 아미노산 잔기는 야생형 알파바이러스 nsP3의 아미노산 서열에 관하여 번호가 매겨진다. 야생형 알파바이러스 nsP3의 아미노산 서열은 본 명세서에 기술되어 있거나 GenBank 데이터베이스와 같은 공개 기록에서 다른 방법으로 입수할 수 있다. 예를 들어, SFV nsP3와 관련하여 아미노산 잔기는 서열번호 24의 야생형 SFV nsP3에 관해 번호가 매겨지고; SINV nsP3와 관련하여, 아미노산 잔기는 서열번호 25의 야생형 SINV nsP3에 관해 번호가 매겨지고; CHIKV nsP3와 관련하여, 아미노산 잔기는 서열번호 26의 야생형 CHIKV nsP3에 관해 번호가 매겨지고; VEEV nsP3와 관련하여, 아미노산 잔기는 서열번호 27의 야생형 VEEV nsP3에 관해 번호가 매겨지고; EEEV nsP3와 관련하여, 아미노산 잔기는 서열번호 28의 야생형 EEEV nsP3에 관해 번호가 매겨지고; WEEV nsP3와 관련하여, 아미노산 잔기는 서열번호 29의 야생형 WEEV nsP3에 관해 번호가 매겨진다.

[0030] 레플리콘에 의해 인코딩된 nsP1, nsP2, nsP3 및 nsP4 단백질은 기능적 또는 생물학적 활성 단백질이다. 본 발명의 RNA 레플리콘은 또한 알파바이러스 3' 및 5' UTR 일 수 있는 3' 비번역 영역(UTR) 및 5' UTR을 인코딩할 수 있다. RNA 레플리콘은 또한 조절 요소(예컨대 하나 이상의 서브게놈 프로모터) 및 폴리-A 테일을 인코딩할 수 있다. 프로모터, 5' 및/또는 3' UTR, 및 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열은 작동적으로 연결되어 레플리콘 RNA가 유기체에 도입될 때 자가 증폭하고 이종성 단백질 또는 펩티드가 유기체에서 발현되도록 할 수 있다.

[0031] 본 출원의 발명자들은 예상치 않게 뉴 월드 알파바이러스(NW) 게놈으로부터 유래된 RNA 레플리콘에서 nsP3 단백질을 인코딩하는 RNA의 적어도 일부가 올드 월드 알파바이러스(OW)로부터 유래된 nsP3의 적어도 일부를 인코딩하는 RNA로 치환되면, 포유류에서 레플리콘에 인코딩된 이종성 단백질 또는 펩타이드에 대한 면역원성이 현저하게 감소하거나 제거된다는 것을 발견하였다. 따라서, 레플리콘의 일부 구현예에서, nsP3 매크로 도메인 및 중앙 도메인은 뉴 월드 알파바이러스 서열로부터 유래할 수 있는 한편, HVD는 a) 올드 월드 알파바이러스 HVD 서열로부터 유래하거나, b) 올드 월드 알파바이러스 HVD 서열로부터 유래된 부분 및 뉴 월드 알파바이러스 HVD 서열로부터 유래된 부분을 갖는다.

[0032] 다른 구현예에서, 매크로 및 중앙 도메인은 올드 월드 알파바이러스 매크로 및 중앙 도메인 서열로부터 유래되고, HVD는 a) 올드 월드 알파바이러스 HVD 서열로부터 유래되거나, 또는 b) 올드 월드 알파바이러스 HVD 서열로부터 유래된 부분 및 뉴 월드 알파바이러스 HVD 서열로부터 유래된 부분을 갖는다.

[0033] 다른 구현예에서, 매크로 도메인은 뉴 월드 알파바이러스 매크로 도메인 서열로부터 유래되고, 중앙 도메인은

올드 월드 알파바이러스 중앙 도메인 서열로부터 유래되고, HVD는 a) 올드 월드 알파바이러스 HVD 서열로부터 유래되거나, 또는 b) 올드 월드 알파바이러스 HVD 서열로부터 유래된 부분 및 뉴 월드 알파바이러스 HVD 서열로부터 유래된 부분을 갖는다.

[0034] 다른 구현예에서, 매크로 도메인은 올드 월드 알파바이러스 매크로 도메인 서열로부터 유래되고, 중앙 도메인은 뉴 월드 알파바이러스 중앙 도메인 서열로부터 유래되고, HVD는 a) 올드 월드 알파바이러스 HVD 서열로부터 유래되거나, 또는 b) 올드 월드 알파바이러스 HVD 서열로부터 유래된 부분 및 뉴 월드 알파바이러스 HVD 서열로부터 유래된 부분을 갖는다.

[0035] 일부 구현예에서, 레플리콘은 뉴 월드 알파바이러스 HVD 서열로부터 유래된 부분 및 올드 월드 HVD 서열로부터 유래된 부분을 갖는 하이브리드 또는 키메라 뉴 월드/올드 월드 서열인 HVD를 인코딩한다. 다양한 구현예에서 올드 월드 부분은 적어도 5 또는 적어도 10 또는 적어도 15 또는 적어도 20 또는 적어도 25 또는 적어도 30 또는 적어도 52 또는 적어도 53 또는 적어도 75 또는 적어도 100 또는 적어도 125 또는 적어도 150 또는 적어도 175 또는 적어도 200개의 아미노산일 수 있다. 상기 부분은 함께 야생형 올드 월드 또는 뉴 월드 알파바이러스 HVD 서열과 동일한 길이를 갖는 HVD를 포함할 수 있거나, 야생형, 올드 월드 또는 뉴 월드 알파바이러스 HVD 서열에 비해 최대 10개 또는 최대 20개 또는 최대 30개 아미노산이 짧거나; 또는 최대 10개 또는 최대 20개 또는 최대 30개 또는 최대 40개 또는 최대 50개 또는 최대 60개 또는 최대 70개 또는 최대 80개 또는 최대 90개 또는 최대 100개 아미노산이 길 수 있다.

[0036] 일부 구현예에서 HVD의 N-말단 부분은 뉴 월드 nsP3 HVD 서열로부터 유래 될 수 있고 HVD의 C-말단 아미노산은 야생형 OW 알파바이러스 HVD 아미노산 서열로부터 유래할 수 있으며, 예를 들어 적어도 5 또는 적어도 10 또는 적어도 15 또는 적어도 20 또는 적어도 25 또는 적어도 30 또는 적어도 또는 적어도 31 또는 적어도 32 또는 적어도 33 또는 적어도 34 또는 적어도 35 또는 35-55 또는 35-65 또는 적어도 40 또는 적어도 45 또는 적어도 50 또는 적어도 52 또는 적어도 53 또는 적어도 60 또는 적어도 70 또는 적어도 80 또는 적어도 100 또는 적어도 125 또는 적어도 150 또는 적어도 175개의 HVD의 C-말단 아미노산은 OW HVD의 아미노산으로부터 유래된(그리고 선택적으로 이에 상응하는) 아미노산 서열일 수 있으며; 이들 구현예 중 임의의 것에서 HVD는 또한 길이가 200개 미만 또는 175개 미만 또는 150개 미만 또는 125개 미만 또는 100개 미만 또는 80개 미만의 아미노산일 수 있다. 추가 구현예에서, C-말단 아미노산은 NW 알파바이러스 C-말단 HVD 서열, 예컨대 말단 1-5 또는 5 또는 5-10 또는 10-12 또는 10-13 또는 10-15 또는 15-20개 아미노산으로부터 보유될 수 있으며, 나머지 C-말단 아미노산은 기술된 바와 같은 OW 알파바이러스 HVD로부터 유래할 수 있다

[0037] 본 명세서에 기술된 임의의 구현예에서, 뉴 월드 알파바이러스는 VEEV 또는 EEEV 또는 WEEV 또는 본 명세서에 기술되거나 당업계에 공지된 임의의 뉴 월드 알파바이러스일 수 있고, 올드 월드 알파바이러스는 CHIKV, SINV 또는 SFV 또는 본 명세서에 기술되거나 당업계에 공지된 임의의 올드 월드 바이러스일 수 있다. 뉴 월드 및 올드 월드 알파바이러스는 본 발명에서 임의의 조합으로 사용될 수 있으며, 모든 가능한 조합 및 하위 조합이 여기에 완전히 설명된 것처럼 개시된다.

[0038] 알파바이러스 레플리콘

[0039] 알파바이러스는 그룹 IV Togaviridae 과 바이러스로 분류된다. 이 바이러스는 일반적으로 11 kb-12 kb 범위의 양성 센스 단일 가닥 RNA 게놈을 가지고 있다. 본 발명의 알파바이러스 레플리콘은 길이가 11 kb-12 kb, 또는 10-13 kb, 또는 7-20 kb 또는 7-25 kb 일 수 있으며, 알파바이러스 5' 캡 및 3' 폴리-A 테일일 수 있는 5' 캡 및 3' 폴리-A 테일을 가질 수 있다. 5' 캡은 당업자에게 알려진 것, 예컨대 7-메틸구아닐레이트 캡, 또는 안티리버스 캡 유사체 3'-O-Me-m7G(5')ppp(5')G 또는 다른 유사체 캡 구조일 수 있다. 그들은 일반적으로 외피(enveloped) 바이러스이며 직경 약 70 nm를 가지는 구형이다. 그들은 또한 등척성 뉴클레오캡시드를 가질 수 있다. 레플리콘은 RNA의 단일 조각 상에서 인코딩될 수 있다. 알파바이러스 게놈 및 레플리콘은 비구조적 및 구조적의 두 개의 오픈 리딩 프레임(ORF)을 가진다. 게놈의 비구조적 부분은 바이러스 RNA의 전사 및 복제 역할을 하고 폴리단백질로서 생성되며 바이러스 복제 기구인 단백질 nsP1-nsP4를 인코딩한다. 그러나 레플리콘은 하나 또는 둘 또는 둘 이상의 오픈 리딩 프레임을 가질 수 있다. 본 발명의 임의의 알파바이러스 레플리콘은 캡시드, 뉴클레오캡시드, 코트 단백질 또는 핵단백질을 결여 또는 포함하지 않거나 내부에 포함되지 않거나 연합되지 않을 수 있다. 알파바이러스 레플리콘은 RNA 분자일 수 있다.

[0040] 게놈의 구조적 부분은 코어 뉴클레오캡시드 단백질 C와 이종성 이량체로서 연합하는 외피 단백질 P62 및 E1을 인코딩한다. 본 발명의 RNA 레플리콘은 알파바이러스의 기술된 특성 중 하나 이상을 가질 수 있다. 일부 구현예에서 본 발명의 RNA 레플리콘은 알파바이러스 구조 단백질을 인코딩하는 서열이 결여되며; 또는 알파바이러스

(또는 선택적으로 다른 임의의) 구조 단백질을 인코딩하지 않는다. 일부 구현예에서, 본 발명의 RNA 레플리콘은 모든 조합 및 하위 조합을 본 명세서에 완전히 기재된 것과 같이 포함하는 단백질 C, P62, 6K 및 E1 중 어느 하나 이상을 인코딩하지 않는다. 일부 구현예에서 본 발명의 RNA 레플리콘은 단백질 C, P62, 6K 및 E1 중 어느 하나를 인코딩하지 않는다.

[0041] 알파바이러스 군의 지리적 분리는 이러한 바이러스가 고유한 환경에 맞게 진화하고 적응하는 요인이 될 수 있다. 순환 알파바이러스 혈청복합체는 올드 월드 또는 뉴 월드 알파바이러스로 추가로 분류될 수 있다. 올드 월드 및 뉴 월드 알파바이러스는 본 명세서에 기술된 바와 같이 본 발명에서 이용될 수 있는 서열을 갖는다. 뉴 월드 알파바이러스에는 임의의 뉴 월드 알파바이러스, 예를 들어 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis virus, VEEV), 웨스턴 이콰인 뇌염 바이러스(western equine encephalitis virus, WEEV), 및 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV), 포트모건(Fort Morgan, FMV), 하일랜드 제이 바이러스(Highland J virus, HJV), 버기 크릭 바이러스(Buggy Creek virus, BCRV), 무캄보 바이러스(Mucambo virus, MUCV), 및 픽수나 바이러스(Pixuna virus, PIXV)를 포함한다. 올드 월드 알파바이러스는 임의의 올드 월드 알파바이러스, 예를 들어 신드비스 바이러스(Sindbis virus, SINV), 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV), 치쿤군야 바이러스(Chickungunya virus, CHIKV), 베바루 바이러스(Bebaru virus, BEBV), 오뇽뇽 바이러스(O'nyong nyong virus, ONNV), 로스 리버 바이러스(Ross River Virus, RRV), 사기야마 바이러스(Sagiyama virus, SAGV), 게타 바이러스(Getah virus, GETV), 미들버그 바이러스(Middleburg virus, MIDV), 엔두무 바이러스(Ndumu, NDUV), 바르마 포레스트 바이러스(Barmah Forest virus, BFV), 마야로 바이러스(Mayaro virus, MAYV), 아우라 바이러스(Aura virus, AURA), 우나 바이러스(Una virus), 와타로아 바이러스(Whataroa virus), 바뱅크 바이러스(Babank virus), 및 키질라가시 바이러스(Kyzylagach virus)를 포함한다. 뉴 월드 및 올드 월드 바이러스 및 이들의 서열은 본 발명의 RNA 레플리콘에서 임의의 조합 또는 하위 조합으로 사용될 수 있으며, 본 명세서에 완전히 기술된 것처럼 모든 가능한 조합 및 하위 조합으로 개시된다.

[0042] 본 발명의 RNA 레플리콘은 알파바이러스 게놈으로부터 유래될 수 있는데, 이는 알파바이러스 게놈의 구조적 특징 중 일부를 가지고 있거나 이와 유사하다는 것을 의미한다. 본 발명의 RNA 레플리콘은 변형된 알파바이러스 게놈일 수 있다. 본 명세서에 개시된 레플리콘의 일부 구현예에서, 레플리콘의 하나 이상의 서열이 "트랜스로(in trans)" 제공 될 수 있으며, 즉, 레플리콘의 서열이 하나 이상의 RNA 분자 상에 제공된다. 다른 구현예에서, 레플리콘의 모든 서열은 단일 RNA 분자 상에 존재하며, 이는 또한 본 명세서에 기재된 바와 같이 치료될 포유류에 투여될 수 있다.

[0043] 유래된(Derived From)

[0044] 본 발명의 RNA 레플리콘은 야생형 뉴 월드 또는 올드 월드 알파바이러스 게놈으로부터의 RNA 서열(또는 이에 의해 인코딩되는 아미노산 서열)을 포함할 수 있다. 본 명세서에 개시된 본 발명의 임의의 RNA 레플리콘은 야생형 알파바이러스 게놈 서열 "유래" 또는 "기반"의 RNA 서열을 포함할 수 있으며, 이는 이들이 뉴 월드 또는 올드 월드 알파바이러스 게놈일 수 있는 야생형 RNA 알파바이러스 게놈으로부터의 RNA 서열(상응하는 RNA 서열일 수 있는)과 적어도 60% 또는 적어도 65% 또는 적어도 68% 또는 적어도 70% 또는 80% 이상 또는 85% 이상 또는 90% 이상 또는 95% 이상 또는 97% 이상 또는 98% 이상 또는 99% 이상 또는 100% 또는 80-99% 또는 90-100% 또는 95-99% 또는 95-100% 또는 97-99% 또는 98-99% 서열 동일성을 가진다는 것을 의미한다. 본 명세서에 개시된 임의의 핵산 또는 아미노산 서열은 기능적 또는 생물학적 활성 일 수 있고 알파바이러스 또는 레플리콘의 자가 복제에 필요한 다른 서열에 작동적으로 연결될 수 있다. 분자는 그의 천연(또는 야생형)의, 상응하는 분자와 적어도 50% 동일한 활성을 수행한다면 기능적 또는 생물학적으로 활성이지만, 기능적 분자는 또한 그의 천연(또는 야생형)의, 상응하는 분자와 적어도 60% 또는 적어도 70% 또는 적어도 90% 또는 적어도 95% 또는 100% 동일한 활성을 수행 할 수 있다. RNA 레플리콘은 또한 야생형 알파바이러스 아미노산 서열로부터 유래되거나 그에 기반한 아미노산 서열을 인코딩할 수 있는데, 이는 이들이 뉴 월드 또는 올드 월드 알파바이러스 게놈일 수 있는 야생형 RNA 알파바이러스 게놈에 의해 인코딩된 아미노산 서열(상응하는 서열일 수 있는)과 적어도 60% 또는 적어도 65% 또는 적어도 68% 또는 적어도 70% 또는 적어도 80% 적어도 70% 또는 적어도 80% 또는 적어도 90% 또는 적어도 95% 또는 적어도 97% 또는 적어도 98% 또는 적어도 99% 또는 100% 또는 80-99% 또는 90-100% 또는 95-99% 또는 95-100% 또는 97-99% 또는 98-99% 서열 동일성을 가진다는 것을 의미한다. 다른 서열로부터 유래된 서열은 본래 서열 보다 최대 5% 또는 최대 10% 또는 최대 20% 또는 최대 30% 더 길거나 짧을 수 있다. 임의의 구현예에서 서열 동일성은 G3BP 또는 FXR 결합 부위를 인코딩하는(또는 G3BP 또는 FXR 결합 부위를 갖는 아미노산 서열) 임의의 뉴클레오티드 서열에 대해 적어도 95% 또는 적어도 97% 또는 적어도 98% 또는 적어도 99% 또는 100%일 수 있다. 이러한 서열은 본래 서열 보다 최대 5% 또는 최대 10% 또는 최대 20% 또는 최대 30% 더 길거나

짧을 수 있다.

[0045] 예를 들어, 일부 구현예에서 nsP1, nsP2, nsP3 매크로 도메인, nsP3 중앙 도메인, nsP3 초가변 도메인, 및/또는 nsP4 단백질 중 임의의 하나 이상을 인코딩하는 RNA 서열은 상응하는 야생형 알파바이러스 서열로부터 유래할 수 있다. 상기 "상응하는" 서열은 다른 형태의 알파바이러스의 유사체 서열일 수 있다. 상응하는 서열은 본 명세서에서 개시되며 또한 당업자에게 공지된 서열 정렬 도구(예컨대, Clustal Omega)를 통해 결정될 수 있다. 도 6은 Clustal Omega를 사용하여 얻은 올드 월드 및 뉴 월드 알파바이러스의 대표적인 구성원으로부터 얻은 nsP3 단백질의 상응하는 서열을 보여주는 서열 정렬을 나타낸다. 그러나 당업자에 의해 허용되는 다른 서열 정렬 도구 역시 사용될 수 있다. 서열 정렬을 수행하는 데 유용한 프로그램은 Molecular Systems Biology (2011) 7, 539에서도 찾아 볼 수 있다. 따라서, 뉴 월드 알파바이러스로부터의 nsP1, nsP2, nsP3, nsP4 서열은 올드 월드 알파바이러스로부터의 nsP1, nsP2, nsP3, nsP4 서열에 각각 "상응"한다. 서열은 상응하는 서열일 수도 있다. 상응하는 아미노산 서열은 적어도 5 또는 적어도 10 또는 적어도 15 또는 적어도 20 또는 적어도 25 또는 적어도 30 또는 적어도 52 또는 적어도 53 또는 적어도 75 또는 적어도 100 또는 적어도 125 또는 적어도 150 또는 적어도 175 또는 적어도 200개의 아미노산일 수 있고, 그리고 본래 서열 보다 최대 5% 또는 최대 10% 또는 최대 20% 또는 최대 30% 더 길거나 더 짧을 수 있고; 상응하는 핵산 서열은 적어도 15 또는 적어도 30 또는 적어도 45 또는 적어도 60 또는 적어도 75 또는 적어도 90 또는 적어도 156 또는 적어도 159 또는 적어도 225 또는 적어도 300 또는 적어도 375 또는 적어도 450 또는 적어도 525 또는 적어도 600개의 뉴클레오티드일 수 있다. 이러한 서열은 본래 서열보다 최대 5% 또는 최대 10% 또는 최대 20% 또는 최대 30% 더 길거나 짧을 수 있다.

[0046] 레플리콘의 일부 구현예에서 각각의 nsP1, nsP2 및 nsP4 서열은 뉴 월드 알파바이러스 게놈으로부터 유래되거나 그에 기반할 수 있다. 일부 구현예에서, 야생형 뉴 월드 알파바이러스 게놈으로부터 유래되거나 그에 기반한 RNA 레플리콘은 야생형 뉴 월드 알파바이러스 게놈으로부터 유래하지 않은 적어도 하나의 RNA 서열(적어도 하나의 이종성 단백질 또는 펩티드 외에)을 포함할 수 있으며, 이는 nsP3, 또는 nsP3의 중앙 및/또는 매크로 도메인(들), 또는 HVD의 적어도 부분의 서열일 수 있다. 일부 구현예에서, 뉴 월드 알파바이러스 게놈으로부터 유래된 RNA 레플리콘은 nsP3, 또는 nsP3의 도메인, 또는 야생형 올드 월드 알파바이러스 게놈으로부터 상응하는 서열로 치환된 nsP3의 도메인의 부분을 인코딩하는 RNA 서열을 가질 수 있다. 전체 레플리콘이 "유래"되거나 또는 "기반"된을 언급할 때, 적어도 하나의 이종성 단백질 또는 펩티드를 인코딩하는 RNA의 서열(들)을 고려하지 않으며, 선택적으로 임의의 조합 또는 하위 조합으로 nsP3 단백질, 임의의 하나 이상의 매크로 도메인, 중앙 도메인, 및/또는 nsP3의 HVD 도메인 또한 고려하지 않을 수 있다.

[0047] 용어 "RNA 레플리콘"은 인간, 포유동물 또는 동물 세포일 수 있는, 허용 세포 내에서 자체 증폭 또는 자가 복제를 지시하기 위해 필요한 모든 유전적 정보를 포함하는 RNA를 지칭한다. RNA 레플리콘은 1) 바이러스 또는 숙주 세포 유래 단백질, 핵산 또는 리보핵단백질과 상호 작용하여 RNA 증폭 과정을 촉매할 수 있는 RNA 의존성 RNA 중합효소를 인코딩한다. 비 구조 단백질은 nsP1, nsP2, nsP3, nsP4를 포함한다; 2) 3' 및 5' UTR(비구조 단백질 매개 증폭을 위한 알파바이러스 뉴클레오티드 서열)과 같은 게놈 및 서브게놈 RNA의 복제 및 전사에 필요한 시스-작용 RNA 서열 및/또는 서브게놈을 포함한다. 이들 서열은 자가-인코딩된 단백질, 또는 비자가-인코딩된 세포 유래 단백질, 핵산 또는 리보핵단백질, 또는 이들 성분들 간의 임의의 복합체에 대한 복제 과정 동안 결합될 수 있다. 일부 구현예에서, 변형된 RNA 레플리콘 분자는 전형적으로 다음의 순서화된 요소를 포함한다: 복제를 위해 시스로(in cis) 요구되는 5' 바이러스 RNA 서열(예컨대, 5' UTR 및 5' CSE), 생물학적 활성 비구조 단백질을 코딩하는 서열(예컨대 nsP1234), 서브게놈 RNA를 전사하기 위한 프로모터, 복제를 위해 시스로(in cis) 요구되는 3' 바이러스 서열(예컨대 3' UTR), 및 폴리아데닐레이트 구역(tract), 및 선택적으로, 서브게놈 프로모터의 제어 후 또는 제어 하에 이종성 단백질 또는 펩티드를 인코딩하는 서열(또는 둘 이상의 서열). 또한, 용어 RNA 레플리콘은 양성 센스(또는 메시지 센스) 분자를 지칭 할 수 있고, RNA 레플리콘은 임의의 공지의, 천연 발생 RNA 바이러스의 길이와 상이할 수 있다. 본 개시내용의 임의의 구현예에서, RNA 레플리콘은 구조적 바이러스 단백질(예컨대, 뉴클레오캡시드 단백질 C, 및 외피 단백질 P62, 6K 및 E1) 중 적어도 하나(또는 모두)의 서열(들)을 결여할(또는 함유하지 않을) 수 있다. 이러한 구현예에서, 하나 이상의 구조적 유전자를 인코딩하는 서열은, 예를 들어, 적어도 하나의 이종성 단백질 또는 펩티드(또는 다른 관심 유전자 (GOI))에 대한 코딩 서열과 같은 하나 이상의 이종성 서열로 치환 될 수 있다.

[0048] 다양한 구현예에서 본 명세서에 개시된 RNA 레플리콘은 조작(engineered), 합성 또는 재조합 RNA 레플리콘일 수 있다. 본 명세서에 사용된 용어 재조합은 폴리뉴클레오티드의 인간 조작(manipulation)으로부터의 또는 간접적으로 이의 결과인, 임의의 분자(예컨대, DNA, RNA 등)를 의미한다. 비제한적 예로서, cDNA는 시험관내(in vitro) 중합효소 반응(들)에 의해 생성된 임의의 핵산 분자 그대로의 것이거나, 여기에 링커가 부착된 것, 또는

클로닝 벡터 또는 발현 벡터와 같은 벡터 내에 통합된 것인, 재조합 DNA 분자이다. 비제한적 예로서, 재조합 RNA 레플리콘은 다음 중 하나 이상일 수 있다: 1) 예컨대 화학적 또는 효소적 기술(예컨대, 화학적 핵산 합성을 사용하여, 또는 복제, 중합, 엑소뉴클레오리틱 분해, 엔도뉴클레오리틱 분해, 라이게이션(ligation), 역전사, 전사, 염기 변형(예컨대, 메틸레이션 포함) 을 위해 효소를 사용하여), 또는 핵산 분자의 재조합(동종 및 부위 특이적 재조합 포함)을 이용하여 시험관내(in vitro) 합성 또는 변형된 것; 2) 천연적으로 결합되지 않은 결합된(conjoined) 뉴클레오티드 서열; 3) 천연 발생 뉴클레오티드 서열과 관련하여 하나 이상의 뉴클레오티드가 결여되도록 분자 클로닝 기술을 사용하여 조작된 것; 및 4) 천연적으로 발생하는 뉴클레오티드 서열에 대해 하나 이상의 서열 변화 또는 재배열을 갖도록 분자 클로닝 기술을 사용하여 조작된 것.

[0049] 본 명세서에 사용되는, 핵산 또는 폴리펩티드 서열에 관한 용어 "% 동일성" 또는 "상동성" 또는 "공유 서열 동일성" 또는 "퍼센트(%) 서열 동일성"은 필요한 경우, 최대 퍼센트 상동성을 달성하기 위하여, 최대 퍼센트 동일성을 위한 정렬 및 갭 도입 후, 뉴클레오티드 또는 아미노산 잔기의 공지의 폴리펩티드와 동일한 후보 서열 내의 백분율로 정의된다. N-말단 또는 C-말단 삽입 또는 결실은 상동성에 영향을 미치는 것으로 해석되지 않아야 하며, 약 30개 미만, 약 20개 미만, 또는 약 10개 미만 또는 5개 미만의 아미노산 잔기의 폴리펩티드 서열 내부 결실 및/또는 서열 내 삽입은 상동성에 영향을 미치는 것으로 해석되지 않아야 한다. 뉴클레오티드 또는 아미노산 서열 수준에서 상동성 또는 동일성은 서열 유사성 검색에 맞게 조정된, blastp, blastn, blastx, tblastn 및 tblastx(Altschul (1997), Nucleic Acids Res. 25, 3389-3402 및 Karlin (1990), Proc. Natl. Acad. Sci. USA 87, 2264-2268) 프로그램에서 사용하는 알고리즘을 사용하여 BLAST(Basic Local Alignment Search Tool) 분석에 의해 결정될 수 있다. BLAST 프로그램에서 사용하는 접근 방식은 먼저 쿼리 서열과 데이터베이스 서열 사이에 갭이 있거나 없는 유사한 세그먼트를 고려한 다음, 식별된 모든 매칭의 통계적 유의성을 평가하고, 마지막으로 유의성의 사전 선택된 역치를 충족하는 매칭 만을 요약하는 것이다. 서열 데이터베이스의 유사성 검색의 기본 이슈에 대한 논의는 Altschul (1994), Nature Genetics 6, 119-129를 참조. 히스토그램, 설명, 정렬, 예상(즉, 데이터베이스 서열에 대한 매칭을 보고하기 위한 통계적 유의성 역치), 컷오프, 매트릭스 및 필터(낮은 복잡성)에 대한 검색 매개 변수는 기본 설정에 있을 수 있다. blastp, blastx, tblastn 및 tblastx에서 사용하는 기본 스코어링 매트릭스는 BLOSUM62 매트릭스(Henikoff (1992), Proc. Natl. Acad. Sci. USA 89, 10915-10919)이며, 길이 85개 이상의 쿼리 서열(뉴클레오티드 염기 또는 아미노산)을 위해 권장된다.

[0050] 뉴클레오티드 서열을 비교하기 위해 설계된 blastn의 경우, 스코어링 매트릭스는 M(즉, 매칭 잔기 쌍에 대한 보상 점수) 대 N(즉, 미스매칭 잔기에 대한 페널티 점수)의 비율로 설정되며, 여기서 M 및 N의 기본값은 각각 +5 및 -4 일 수 있다. 다음과 같이 4 개의 blastn 매개 변수를 조정할 수 있다: Q=10(갭 생성 페널티); R=10(갭 확장 페널티); wink=1(쿼리의 모든 윙크 위치에서 단어 히트를 생성); 및 gapw=16 (갭 정렬이 생성되는 창 너비 설정). 아미노산 서열 비교를 위한 동등한 Blastp 매개 변수 설정은 다음과 같을 수 있다: Q=9; R=2; wink=1; 및 gapw=32. GCG 패키지 버전 10.0에서 사용할 수 있는 서열 간 BESTFIT® 비교는 DNA 매개 변수 GAP=50(갭 생성 페널티) 및 LEN=3(갭 확장 페널티)을 사용할 수 있으며, 단백질 비교에서 동등한 설정은 GAP=8 및 LEN=2이다.

[0051] 본 명세서의 핵산 또는 폴리펩티드 서열, 예를 들어 nsP1, nsP2, nsP3, nsP3 매크로 도메인, nsP3 중앙 도메인, nsP3 초가변 도메인, nsP4, RdRp, P1234의 서열에 대한 개시에서, 본래 서열에 기반하거나 이로부터 유래한 서열 또한 개시된 것으로 간주된다. 따라서 개시된 서열은 본 명세서에 기술된 임의의 폴리펩티드 서열의 전장 폴리펩티드 서열, 예컨대 서열번호 1-20(및 서열번호 1-29 중 임의의 것을 인코딩하는 뉴클레오티드 서열), 및 그 단편과 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 또는 적어도 85%의 서열 동일성을 갖는 폴리펩티드를 포함하고, 예를 들어 적어도 86%, 적어도 87%, 적어도 88%, 적어도 89%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 또는 85-99% 또는 85-95% 또는 90- 99% 또는 95-99% 또는 97-99% 또는 98-99% 의 서열 동일성을 갖는 폴리펩티드 서열을 포함한다. 또한 본 명세서에 개시된 임의의 서열의 단편 또는 부분이 개시된다. 서열의 단편 또는 부분은 전체 서열(또는 이러한 단편을 인코딩하는 핵산)의 적어도 5개 또는 적어도 7개 또는 적어도 10개, 또는 적어도 20개, 또는 적어도 30개, 적어도 50개, 적어도 75개, 적어도 100개, 적어도 125개, 150개 이상 또는 5-10개 또는 10-12개 또는 10-15개 또는 15-20개 또는 20-40개 또는 20-50개 또는 30-50개 또는 30-75개 또는 30-100개 아미노산 잔기, 또는 적어도 100개 또는 적어도 200개 또는 적어도 300개 또는 적어도 400개 또는 적어도 500개 또는 적어도 600개 또는 적어도 700개 또는 적어도 800개 또는 적어도 900개 또는 적어도 1000개 또는 100-200개 또는 100-500개 또는 100-1000개 또는 500-1000개 아미노산 잔기(또는 이러한 단편을 인코딩하는 핵산), 또는 이들 양 중 임의의 것이되 다만 500개 미만 또는 700개 미만 또는 1000개 미만 또는 2000개 미만의 본 명세서에 개시된 서열번호 1-29 또는 임의의 단

편 연속 아미노산, 또는 이러한 단편을 인코딩하는 핵산을 갖는 서열을 포함 할 수 있다. 또한, 예컨대, 적어도 1개 또는 2개 또는 3개 또는 4개 또는 5개의 아미노산 잔기가 삽입 및 치환을 포함하는 개시된 서열(들)에 및/또는 내부에 N- 및/또는 C-말단에 삽입된 이러한 서열의 변이체, 및 이러한 변이체를 인코딩하는 핵산 서열을 개시한다. 고려되는 변이체는 추가적으로 또는 대안적으로, 예컨대 상동성 재조합 또는 부위 지정 또는 PCR 돌연변이 유발에 의한 미리 설정된 돌연변이를 포함하는 것들, 및 본 명세서에 기재된 것들을 포함하지만 이에 제한되지 않는 다른 종의 상응하는 폴리펩티드 또는 핵산, 삽입 및 치환을 포함하는 폴리펩티드 또는 핵산 군의 대립유전자 또는 다른 천연 발생 변이체; 및/또는 폴리펩티드가 삽입 및 치환를 포함하는 천연 발생 아미노산 이외의 모이어티(예를 들어, 효소와 같은 검출 가능한 모이어티)를 가지는 치환, 화학적, 효소적 또는 다른 적절한 수단에 의해 공유결합적으로 변형된 것인 유도체를 포함할 수 있다. 본 명세서에 기재된 핵산 서열은 RNA 서열일 수 있다.

[0052] RNA 레플리콘의 임의의 성분 또는 서열은 임의의 다른 성분 또는 서열에 작동적으로 연결될 수 있다. RNA 레플리콘의 성분 또는 서열은 숙주 세포 또는 치료된 유기체에서 적어도 하나의 이종성 단백질 또는 펩티드(또는 생체치료제)의 발현 및/또는 레플리콘의 자가-복제 능력을 위해 작동적으로 연결될 수 있다. 용어 "작동적으로 연결된"은 일반적인 기능을 수행하도록 구성된 둘 이상의 서열 간의 기능적 연결을 의미한다. 따라서, 코딩 서열에 작동적으로 연결된 프로모터 또는 UTR은 적절한 효소가 존재할 때 코딩 서열의 전사 및 발현에 영향을 미칠 수 있다. 프로모터는 그것의 발현을 지시하는 기능을 하는 한 코딩 서열과 인접 할 필요는 없다. 따라서, 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열과 조절 서열(예컨대, 프로모터 또는 UTR) 사이의 작동적 연결은 관심 폴리뉴클레오티드의 발현을 가능하게 하는 기능적 연결이다. 작동적으로 연결된 서열은 또한 RdRp(예컨대, nsP4), nsP1-4, UTR, 프로모터 및 RNA 레플리콘을 인코딩하는 기타 서열을 인코딩하는 서열과 같은 서열을 지칭할 수 있으며, 이들 서열은 생체치료제 분자의 전사 및 번역 및/또는 레플리콘의 복제를 가능하게 하도록 연결된다. UTR은 다른 인코딩 된 서열의 리보솜에 의한 인식 및 번역에 필수적인 서열 및 간격(spacing)을 제공함으로써 작동적으로 연결될 수 있다.

[0053] G3BP 및 FXR

[0054] G3BP(Ras-GTPase- 활성화 단백질(Src-homology 3(SH3) 도메인)-결합 단백질)와 FXR(fragile X family 단백질)은 둘 다 자가 조립하여 리보핵단백질 복합체(RNP)를 형성하는 RNA 결합 단백질이다. 둘 다 HVD 도메인에 결합한다. G3BP 및 FXR에 의해 형성된 RNP 복합체는 세포 내에서 구별되는 기능을 수행한다. 예를 들어, G3BP는 면역 반응에서 핵형성 및 스트레스 과립 형성에 중요하다. 스트레스 과립은 mRNA 번역을 격리 및 차단하는 것은 물론 I형 인터페론 및 기타 사이토카인의 유도 및 분비를 조절하는 기능을 한다. 이러한 활동은 함께 세포내 항바이러스 상태를 강화하고 적응 면역 반응을 촉진한다. 한편 FXR 계열 단백질은 선천 면역에서 역할을 하는 것으로 여겨지지 않지만 폴리리보솜과 결합하여 뉴런에서 RNA 수송 과립을 형성한다. 도 7은 이러한 RNA 결합 단백질이 알파바이러스 서열에 결합하는 다양한 알파바이러스 nsP3 HVD의 영역을 나타낸다.

[0055] 비구조 단백질

[0056] 알파바이러스 게놈은 비구조 단백질인 nsP1, nsP2, nsP3, 및 nsP4를 인코딩하며, 이는 때때로 P1234(또는 nsP1-4 또는 nsP1234)로 지정되는, 단일 폴리단백질 전구체로서 생성되고, 단백질분해(proteolytic) 가공을 통해 성숙 단백질로 절단된다. nsP1은 크기가 약 60kDa 일 수 있으며 메틸트랜스퍼라제 활성을 가질 수 있으며 바이러스 캡핑 반응에 관여할 수 있다. nsP2는 약 90kDa의 크기를 가지며 헬리카제 및 프로테아제 활성을 가질 수 있는 한편 nsP3은 약 60kDa이며 3개 도메인: 매크로 도메인, 중앙(또는 알파바이러스 고유) 도메인 및 초가변 도메인(HVD)을 포함한다(도 6 참조). nsP4는 크기가 약 70kDa이며 코어 RNA 의존성 RNA 중합효소(RdRp) 촉매 도메인을 포함한다. 감염 후 알파바이러스 게놈 RNA는 번역되어 P1234 폴리단백질을 생성하며, 이는 개별 단백질로 절단된다.

[0057] nsP3

[0058] 알파바이러스 nsP3 단백질은 3 개의 도메인을 포함한다; a) 매크로 도메인, b) 중앙(또는 알파) 도메인, c) 초가변 도메인(HVD). 올드 월드 및 뉴 월드 알파바이러스의 일부 대표적인 구성원의 3개 도메인에 대한 상응하는 아미노산 서열은 도 6에 나타난다. 다양한 구현예에서 본 발명의 레플리콘은 야생형 알파바이러스 nsP3로부터 유래된 nsP3 매크로 도메인, 및 야생형 알파바이러스 nsP3에서 파생 된 nsP3 중앙 도메인을 인코딩하는 RNA 서열을 갖는다. 다양한 구현예에서, 매크로 및 중앙 도메인(들)은 둘 다 뉴 월드 야생형 알파바이러스 nsP3로부터 유래 될 수 있거나, 둘 다 올드 월드 야생형 알파바이러스 nsP3 단백질로부터 유래 될 수 있다. 일부 구현예에서 매크로 도메인은 뉴 월드 야생형 알파바이러스 매크로 도메인으로부터 유래 될 수 있고 중앙 도메인은 올드

월드 야생형 알파바이러스 중앙 도메인으로부터 유래 될 수 있고, 또는 그 반대의 경우도 마찬가지이다. 다양한 도메인은 본 명세서에 기재된 임의의 서열일 수 있다.

[0059] 초가변 도메인(HVD)

[0060] 일부 구현예에서, 레플리콘은 FXR 결합 부위가 시작되는 아미노산의 C-말단 측에 대한 서열이 결실되고 올드 월드 야생형 알파바이러스 HVD 서열, 또는 그 부분의 대체 서열로 대체 될 수 있는 뉴 월드 알파바이러스 HVD를 가질 수 있다. 본 명세서에 올드 월드 알파바이러스 대체 서열이 기술된다. 따라서 뉴 월드 알파바이러스가 VEEV 인 경우 nsP3의 478 아미노산의 C-말단측의 아미노산이 결실될 수 있고; 뉴 월드 알파바이러스가 EEEV 인 경우, nsP3의 531 아미노산의 C-말단측의 아미노산은 결실될 수 있다. 뉴 월드 알파바이러스가 WEEV 인 경우 nsP3의 504 아미노산의 C-말단측의 아미노산이 결실될 수 있다(도 7 참조). 이들 구현예 중 임의의 것에서 대체 서열은 본 명세서에 기술된 바와 같이 치환될 수 있다. 본 명세서에서 달리 기술된 바와 같이, 뉴 월드 알파바이러스 HVD의 C-말단 아미노산의 부분은 그럼에도 불구하고 올드 월드 서열의 C-말단 측에 보유될 수 있다.

[0061] 일부 구현예에서, 뉴 월드 알파바이러스에서 FXR 결합 부위를 인코딩하는 서열의 적어도 부분은 결실되고 본 명세서에 기술된 대체 서열로 대체될 수 있다. 따라서, 뉴 월드 알파바이러스가 VEEV 인 경우, nsP3의 478-517 또는 478-545 아미노산은 결실되고 올드 월드 알파바이러스의 대체 서열로 대체될 수 있다. 또는 뉴 월드 알파바이러스가 VEEV인 경우 nsP3의 아미노산 478-545 사이에 존재하는 반복 중 하나 이상을 결실하고 선택적으로 올드 월드 알파바이러스 대체 서열로 대체할 수 있다. 뉴 월드 알파바이러스가 EEEV인 경우 nsP3의 531-547 아미노산을 결실하고 올드 월드 대체 서열로 대체할 수 있다. 뉴 월드 알파바이러스가 WEEV인 경우 nsP3의 504-520 아미노산을 결실하고 올드 월드 대체 서열로 대체할 수 있다. 다른 구현예에서 FXR 결합 부위를 인코딩하는 전체 서열이 결실될 수 있거나, FXR 결합 부위의 적어도 50% 또는 적어도 70% 또는 적어도 80% 또는 적어도 90%가 결실될 수 있고, 선택적으로 대체 서열로 대체 될 수 있다. 임의의 구현예에서, 언급된 서열은 결실되고 대체 서열이 삽입되지 않을 수 있다.

[0062] 올드 월드 알파바이러스 대체 서열은 하나 이상의 G3BP 결합 부위, 또는 G3BP 결합 부위의 적어도 부분을 갖는 아미노산 단편을 포함할 수 있다. 따라서, 대체 서열은 FGDF(서열번호 18) 또는 FGSF(서열번호 19) 일 수 있다. 대체 서열은 또한 올드 월드 알파바이러스의 야생형 nsP3 초가변 도메인의 적어도 일부로부터 유래될 수 있다. 올드 월드 알파바이러스 대체 서열의 추가적인 예는 아래에 기술된다. 올드 월드 알파바이러스 대체 서열은 본 명세서에 기재된 임의의 뉴 월드 알파바이러스 HVD 서열의 서열을 갖는 레플리콘에 사용될 수 있다. 임의의 구현예에서, 뉴 월드 알파바이러스는 VEEV, EEEV, WEEV 또는 본 명세서에 기재된 임의의 뉴 월드 알파바이러스일 수 있다.

[0063] 올드 월드 알파바이러스가 CHIKV 인 경우, 대체 서열은 CHIKV nsP3의 아미노산 479-582 또는 479-500 또는 479-500 일 수 있다.

[0064] 올드 월드 알파바이러스가 SINV 인 경우, 대체 서열은 SINV nsP3의 아미노산 490-493 또는 513-516 또는 490-516을 포함하는 서열일 수 있다.

[0065] 올드 월드 알파바이러스가 SFV 인 경우 대체 서열은 SFV nsP3의 아미노산 451-471, 또는 451-454, 또는 468-471을 포함하는 서열일 수 있다.

[0066] 올드 월드 알파바이러스가 MAYV 인 경우, 대체 서열은 MAYV nsP3의 아미노산 470-473을 포함하는 서열일 수 있다.

[0067] 올드 월드 알파바이러스가 RRV 인 경우, 대체 서열은 RRV nsP3의 아미노산 412-426, 또는 512-515, 또는 523-526을 포함하는 서열일 수 있다.

[0068] 올드 월드 알파바이러스가 ONNV 인 경우, 대체 서열은 ONNV nsP3의 아미노산 519-540, 또는 519-522, 또는 537-540을 포함하는 서열일 수 있다.

[0069] 올드 월드 알파바이러스가 BFV 인 경우, 대체 서열은 BFV nsP3의 아미노산 429-450, 또는 429-432, 또는 447-450을 포함하는 서열일 수 있다.

[0070] 뉴 월드 및 올드 월드 알파바이러스는 본 명세서에 기술된 임의의 것일 수 있고 임의의 가능한 조합 또는 하위 조합으로 조합될 수 있으며, 이들 모두는 여기에 완전히 설명된 것처럼 개시된다.

[0071] 알파바이러스 게놈은 nsP4에서 코어 RNA 의존성 RNA 중합 효소를 인코딩한다. 폴리단백질의 절단은 nsP2/3 접합

부에서 발생할 수 있으며, 게놈 복제 중에 사용되는 RNA 템플릿에 영향을 준다. 절단 후 nsP3는 nsP2를 둘러싸는 고리 구조를 형성할 수 있으며, 이들 두 단백질은 실질적인 인터페이스를 갖는다. 따라서 nsP2/3 및/또는 nsP3/4의 접합부 주변의 서열을 보존하는 것이 유용할 수 있다.

[0072] 따라서, 일부 구현예에서 nsP3 단백질의 매크로 및/또는 중앙 및/또는 HVD 도메인은 뉴 월드 알파바이러스로부터 유래된 아미노산 서열인 C-말단 및/또는 N-말단 부분(본 명세서에 기술된 바와 같음)을 가질 수 있는 한편 도메인(들)의 나머지 부분은 올드 월드 알파바이러스 서열로부터 유래된다. 예를 들어, 매크로 및/또는 중앙 및/또는 HVD 도메인은 상응하는 올드 월드 알파바이러스 도메인으로부터 유래된 서열을 가질 수 있지만, 뉴 월드 알파바이러스 서열(nsP1, nsP2 및 nsP4가 유래된 뉴 월드 알파바이러스일 수 있음)로부터 유래된 nsP3의 N-말단 및/또는 C-말단의 최초 4 또는 5 또는 6 또는 4-6 또는 6-8 또는 6-10 아미노산을 가질 수 있다. 따라서, 레플리콘은 본 명세서에서 기술된 바와 같이 올드 월드 알파바이러스 nsP3 매크로 및/또는 중앙 및/또는 HVD 도메인으로부터 유래된 아미노산 서열을 인코딩하는 RNA 서열을 가질 수 있고, 도메인(들)의 N-말단 및/또는 C-말단 측의 최초 1-3 또는 1-4 또는 1-5 또는 1-6 또는 1-7 또는 1-8 아미노산은 뉴 월드 알파바이러스로부터 유래되고, 또는 1 또는 2 또는 3 치환을 가질 수 있다. 본 맥락에서 사용된 용어 "C-말단" 및 "N-말단"은 진정한 말단을 가리키는 것이 아니라 폴리단백질(예컨대 P1234)가 개별 폴리펩티드(예컨대 nsP1, nsP2, nsP3 및 nsP4)로 절단될 지점을 가리킨다. nsP를 인코딩하는 서열은 정지 코돈을 특징으로 하며 일반적으로 전사는 그 지점에서 정지된다. 그러나 정지 코돈이 리드스루(readthrough) 정지 코돈으로 취급되는 경우, 말단은 nsP의 N-말단 및/또는 C-말단을 나타낼 수 있는 서열번호 12-17에 지시된 "/" 일 수 있다. 접합부 서열은 말단의 양쪽, 즉, nsP3 측에 있는 1-6 아미노산 일 수 있다. 이들 구현예는 nsP3 서열이 올드 월드 서열로부터 유래되도록 허용하지만 nsP2/nsP3 사이 및 nsP3/nsP4 사이의 접합부를 보존한다. 이러한 접합부의 보존은 뉴 월드 알파바이러스 효소를 사용한 P1234 단백질 접합부의 절단을 허용할 수 있다. 일부 구현예에서 끝에서 두번째 글리신은 접합부에서 보존된다. 올드 월드 알파바이러스는 본 명세서에 기술된 임의의 것일 수 있다. 예를 들어, 뉴 월드 알파바이러스가 VEEV 인 경우 nsP2/nsP3 서열은 LHEAGC/APSY(서열번호 12) 일 수 있으며, 슬래시("/")는 nsP2와 nsP3 사이의 경계를 나타내고, 끝에서 두번째 G가 보존되는 한편 nsP2/nsP3 접합부의 나머지 아미노산은 본 명세서에 기술된 바와 같이 가변적으로 유지된다. VEEV의 nsP3/nsP4 접합의 경우, 서열은 RFDAGA/YIFS(서열번호 13) 일 수 있으며, 끝에서 두번째 글리신은 다시 보존되고 나머지 nsP3 아미노산은 본 명세서에 기술된 바와 같이 가변적이다. 이들 서열은 또한 정지 코돈(TGA)이 선행될 수 있으며, 상기 언급한 바와 같이 때때로 리드스루 정지 코돈으로 취급될 수 있다. 뉴 월드 알파바이러스가 EEEV 인 경우 nsP2/nsP3 서열은 QHEAGR/APAY(서열번호 14) 일 수 있으며, 슬래시("/")는 nsP2와 nsP3 사이의 경계를 나타내고, 끝에서 두번째 G는 보존된다. nsP2/nsP3 접합부의 나머지 아미노산은 본 명세서에 기술된 바와 같이 가변적이다. EEEV의 nsP3/nsP4 접합부의 경우, 서열은 RYEAGA/YIFS(서열번호 15)일 수 있으며, 끝에서 두번째 글리신은 다시 보존되고 나머지 nsP3 아미노산은 본 명세서에 기술된 바와 같이 가변적이다. 이러한 서열은 또한 리드스루 정지 코돈(TGA)이 선행될 수 있다. 뉴 월드 알파바이러스가 WEEV 인 경우 nsP2/nsP3 서열은 RYEAGR/APAY(서열번호 16) 일 수 있으며 슬래시("/")는 nsP2의 끝 또는 말단(및 nsP2와 nsP3 사이의 접합부)을 나타내며, 끝에서 두번째 G는 보존되는 한편 nsP2/nsP3 접합부의 나머지 아미노산이 본 명세서에 기술된 바와 같이 가변적이다. WEEV의 nsP3/nsP4 접합부의 경우, 서열은 RYEAGA/YIFS(서열번호 17) 일 수 있으며, 끝에서 두번째 글리신은 다시 보존되고 나머지 nsP3 아미노산은 본 명세서에 설명된 바와 같이 가변적이다. 이들 서열은 또한 본 명세서에 기술된 바와 같이 리드스루 정지 코돈(TGA)이 선행될 수 있다. 이들 서열(서열번호 12-17) 중 임의의 것은 또한 N-말단 및/또는 C-말단 측에 1 또는 2 또는 3개의 치환을 포함할 수 있다.

[0073] 반복 모티프

[0074] 알파바이러스는 보존 서열 요소(conserved sequence element, CSE)를 포함할 수 있으며, 이는 여러 종에 걸쳐 핵산 서열 또는 폴리펩티드에서 유사하거나 동일한 서열이다. CSE는 뉴 월드 또는 올드 월드 알파바이러스 nsP3의 HVD에서 발생할 수 있으며 당업계에 공지되어 있다.

[0075] 올드 월드 알파바이러스는 또한 FGDF(서열번호 18) 또는 FGSF(서열번호 19) 아미노산 모티프를 포함할 수 있으며, 이는 서열에서 반복되어 반복 서열 또는 반복 모티프를 형성할 수 있다. 본 발명의 RNA 레플리콘의 임의의 구현예에서, 올드 월드 알파바이러스의 HVD는 FGDF/FGDF(서열번호 20) 반복, 또는 FGSF/FGSF(서열번호 21) 반복, 또는 FGDF/FGSF(서열번호 22) 반복, 또는 FGSF/FGDF(서열번호 23) 반복을 포함할 수 있다. 반복이 존재하는 모든 구현예에서 2개의 반복 모티프는 하나 이상의 아미노산 잔기에 의해 분리될 수 있다. 다양한 구현예에서 2개의 반복 모티프는 5 또는 6 또는 7 또는 8 또는 9 또는 10 또는 적어도 10 또는 11 또는 12 또는 13 또는 14 또는 15 또는 16 또는 17 또는 18 또는 19 또는 20 또는 21 또는 22 또는 23 또는 24 또는 25개의 아미노산

잔기 또는 25개 초과의 아미노산 잔기에 의해 분리될 수 있으며, 이는 일 구현예에서 랜덤 아미노산일 수 있다. 일 구현예에서 모티프 또는 반복 모티프는 적어도 10개 내지 25개 이하의 아미노산에 의해 분리되며, 이는 또한 랜덤 아미노산일 수 있다. 다양한 구현예에서 2개의 반복 모티프는 NEGEIESLSSELLT(서열번호 6), SDGEIDELSRRVTTESEPVL(서열번호 7), 또는 DEHEVDALASGIT(서열번호 8), 또는 이들 중 임의의 것으로부터 유래된, 동일한 길이의, 서열에 의해 분리될 수 있고; 따라서, 1) 양 말단에 FGDF(서열번호 18) 모티프; 2) 양 말단에 FGSF(서열번호 19) 모티프; 3) 3' 또는 5' 말단에 FGDF(서열번호 18) 모티프 및 반대쪽 말단에 FGSF(서열번호 19) 모티프를 갖는, 서열번호 6, 7, 또는 8에 의해 분리된 반복 모티프가 개시된다. 다양한 구현예에서 아미노산 서열은 또한 제2 모티프를 추종할 수 있다. 예는 아미노산 서열 DDVLRLGRAGA(서열번호 11) 또는 EPGEVNSIISSRSAVSFPLRKQRRRRRSRRTEY(서열번호 10) 또는 LPGEVDDLTDSDWSTCSDTDDELRLDRAGG(서열번호 9), 또는 이들 중 임의의 것에서 유래딘 서열을 포함하며, 이들 중 임의의 것은 본 명세서에 개시된 모티프 또는 반복 모티프를 추종할 수 있다.

[0076] 비번역 영역

[0077] 본 발명의 임의의 레플리콘은 또한 5' 및 3' 비번역 영역(UTR)을 포함할 수 있다. UTR은 야생형 뉴 월드 또는 올드 월드 알파바이러스 UTR 서열이거나 이들 중 임의의 것으로부터 유래된 서열일 수 있다. 다양한 구현예에서 5' UTR은 약 60 nt 또는 50-70 nt 또는 40-80 nt와 같은 임의의 적절한 길이일 수 있다. 일부 구현예에서 5' UTR은 또한 보존된 1차 또는 2차 구조(예컨대, 하나 이상의 스템-루프(들))를 가질 수 있고 알파바이러스 또는 레플리콘 RNA의 복제에 참여할 수 있다. 일부 구현예에서 3' UTR은 최대 수백 개의 뉴클레오티드일 수 있으며, 예를 들어 50-900 또는 100-900 또는 50-800 또는 100-700 또는 200 nt - 700 nt 일 수 있다. '3 UTR은 또한 2차 구조를 가질 수 있다. 스텝 루프, 폴리아데닐레이트 구역(tract) 또는 폴리-A 테일이 후속될 수 있다. 본 발명의 임의의 구현예에서 5' 및 3' 비번역 영역은 레플리콘에 의해 인코딩된 임의의 다른 서열에 작동적으로 연결될 수 있다. UTR은 다른 인코딩된 서열의 인식 및 전사에 필수적인 서열 및 간격을 제공함으로써 프로모터 및/또는 이종성 단백질 또는 펩티드를 dls코딩하는 서열에 작동적으로 연결될 수 있다.

[0078] 일 구현예에서, 본 발명의 RNA 레플리콘은 이종성 단백질 또는 펩티드 (예컨대, 단일 클론 항체 또는 생체치료제 단백질 또는 펩티드)를 인코딩하는 RNA 서열, 야생형 뉴 월드 알파바이러스 nsP1, nsP2 및 nsP4 단백질 서열로부터 유래된 아미노산 서열을 코딩하는 RNA 서열, 및 5' 및 3' UTR 서열(비구조적 단백질 매개 증폭을 위한)을 가질 수 있다. RNA 레플리콘은 또한 5' 캡 및 폴리아데닐레이트(또는 폴리-A) 테일을 가질 수 있다. RNA 레플리콘은 또한 뉴 월드 알파바이러스 매크로 도메인으로부터 유래된 아미노산 서열, 뉴 월드 알파바이러스 중앙 도메인으로부터 유래된 아미노산 서열 및 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 아미노산 서열을 인코딩할 수 있다. 대안적 구현예에서, RNA 레플리콘은 본 명세서에 기재된 바와 같이 뉴 월드 초가변 도메인으로부터 유래된 아미노산 서열을 갖는 부분 및 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 아미노산 서열을 갖는 다른 부분을 인코딩할 수 있다.

[0079] 이종성 단백질 또는 펩티드의 면역원성은 당업자에게 공지된 수많은 분석, 예를 들어 에피토프-특이적 T-세포 집단에 의한 세포내 사이토카인 또는 분비된 사이토카인의 면역 염색, 또는 에피토프 특이적 T-세포의 수의 빈도 및 총 수의 정량화 및 그들의 분화 및 활성화 상태, 예컨대 단기 이펙터 및 메모리 전구체 이펙터 CD8+ T-세포의 특성화에 의해 결정될 수 있다. 면역원성은 항체 매개 면역 반응, 예컨대 혈청 IgA 또는 IgG 역가를 측정하는 것에 의한 항체 생산을 측정하여 결정할 수도 있다.

[0080] 이종성 단백질 및 펩티드

[0081] 본 발명의 RNA 레플리콘은 알파바이러스에 대해 이종성이고, 또한 체내에서 RNA 서열을 발현하는 인간, 포유 동물 또는 동물에 대해 이종성 일 수 있는(반드시 그런 것은 아님) 적어도 하나의 단백질 또는 펩티드를 인코딩하는 RNA 서열을 포함한다. 임의의 구현예에서, 레플리콘은 2 또는 3 이상의 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열(들)을 가질 수 있다. 다양한 구현예에서 이종성 단백질 또는 펩티드는 본 명세서에 기재된 바와 같은 생체치료제 분자이다. 인간, 포유류 또는 동물에 대한 바이오 치료 분자의 투여는 항-약물 항체 면역 반응을 유발할 위험이 있지만, 본 발명은 인간, 포유류 또는 동물 신체에 레플리콘을 투여할 수 있게 하고, 다음의 발현을 허용한다. 인간, 포유류 또는 동물 신체에서 생체치료제의 발현을 가능하게 하는 한편, 레플리콘을 수용하고 생체치료제 분자를 발현하는 인간, 포유류 또는 동물 세포로부터의 면역 반응을 실질적으로 감소시키거나 전혀 생성하지 않는다. 임의의 구현예에서, 이종성 단백질 또는 펩티드를 인코딩하는 서열은 레플리콘의 하나 이상의 다른 서열(예컨대, 프로모터 또는 5' 또는 3' UTR 서열)에 작동적으로 연결될 수 있고, 이종성 단백질 또는 펩티드가 인간, 포유류 또는 동물에서 발현되도록 서브-게놈 프로모터의 제어하에 있을 수 있다.

[0082] 이종성 단백질 또는 펩티드는 임의의 단백질 또는 펩티드일 수 있으며, 예는 사이토카인, 성장 인자, 면역글로불린, 단일 클론 항체(Fab 항원 결합 단편, Fc 융합 단백질을 포함), 호르몬, 인터페론, 인터루킨, 조절 펩티드 및 단백질을 포함한다. 이종성 단백질이 될 수 있는 단일 클론 항체의 특이적 예로는 락시바쿠맙(raxibacumab), 토실리주맙(tocilizumab), 브렌투지맙베도틴(brentuzimab vedotin), 인자 IX Fc 융합 단백질(Factor IX Fc fusion protein), 릴로나셉트(rilonacept), 오파투무맙(ofatumumab), 베바시주맙(bevacizumab), 벨리무맙(belimumab), 세르톨리주맙 페골(certolizumab pegol), 라무시루맙(ramucirumab), 인자 VIII Fc 융합 단백질(Factor VIII Fc fusion protein), 에타네르셉트(etanercept), 베돌리주맙(vedolizumab), 세툭시맙(cetuximab), 애플리버셉트(aflibercept), 오비누투주맙(obinutuzumab), 트라스투주맙(trastuzumab), 아달리무맙(adalimumab), 카나키누맙(canakinumab), 인플릭시맙(infliximab), 아도-트라스투주맙 엠탄신(ado-trastuzumab emtansine), 펨브롤리주맙(pembrolizumab), 알렘투주맙(alemtuzumab), 라니비주맙(ranibizumab), 로미플로스팀(romiplostim), 벨라타셉트(belatacept), 아바타셉트(abatacept), 페르투주맙(pertuzumab), 데노수맙(denosumab), 인플릭시맙(infliximab), 카투막소맙(catumaxomab), 인플릭시맙(infliximab), 압식시맙(abciximab), 리툭시맙(rituximab), 골리무맙(golimumab), 바실릭시맙(basiliximab), 에쿨리자맙(eculizamab), 우스테키누맙(ustekinumab), 실툭시맙(siltuximab), 팔리비주맙(palivizumab), 나탈리주맙(natalizumab), 파니투무맙(panitumumab), 데노수맙(denosumab), 오말리주맙(omalizumab), 이필리무맙(ipilimumab), 지바플리버셉트(zivaflibercept) 및 이브리투모맙 티우세탄(ibritumomab tiuxetan)을 포함한다. 다른 구현예에서, 이종성 단백질 또는 펩티드는 내피 성장 인자(예컨대, 혈관 EGF, 호르몬(예컨대 인슐린, 릴렉신), 엑손 스키핑 올리고뉴클레오티드, 모르폴리노 올리고머, 모르폴리노 안티센스 올리고머, 또는 종양-특이적 항원을 인코딩하는 RNA일 수 있다. 일부 구현예에서, 이종성 단백질 또는 펩티드는 최대 5 kb 또는 최대 6 kb 또는 최대 7 kb 또는 최대 8 kb, 또는 최대 9 kb 또는 최대 10 kb 또는 최대 11 kb 또는 최대 12 kb 의 RNA 서열에 의해 인코딩될 수 있다. 이종성 단백질은 단일 사슬 항체 분자 일 수도 있다.

[0083] 본 발명의 알파바이러스 레플리콘은 또한 이종성 단백질 또는 펩티드의 발현을 위한 서브게놈 프로모터를 가질 수 있다. 본 명세서에 사용되는 용어 "서브게놈 프로모터"는 바이러스 핵산의 서브게놈 mRNA의 프로모터를 지칭한다. 본 명세서에 사용되는, "알파바이러스 서브게놈 프로모터"는 알파바이러스 복제 과정의 일부로서 서브게놈 메신저 RNA의 전사를 지시하는 야생형 알파바이러스 게놈에서 본래 정의된 바와 같은 프로모터이다.

[0084] 폴리뉴클레오티드, 유전자, 핵산, 폴리펩티드, 단백질 또는 효소와 관련하여 사용될 때 용어 "이종성"은 숙주 종(species)으로부터 유래하지 않은 폴리뉴클레오티드, 유전자, 핵산, 폴리펩티드, 단백질 또는 효소를 지칭한다. 예를 들어, 본 명세서에 사용된 "이종성 유전자" 또는 "이종성 핵산 서열"은 이것이 도입된 숙주 유기체의 종과 다른 종으로부터의 유전자 또는 핵산 서열을 지칭한다. 이종성 서열은 또한 합성될 수 있으며 유기체로부터 유래되지 않거나 자연에서 발견되지 않을 수 있다. 유전자 조절 서열 또는 유전자 서열의 발현을 조작하는데 사용되는 보조 핵산 서열(예컨대 5' 비번역 영역, 3' 비번역 영역, 폴리 A 부가 서열, 인트론 서열, 스플라이스 부위, 리보솜 결합 부위, 내부 리보솜 진입 서열, 게놈 상동성 영역, 재조합 부위 등) 또는 단백질 도메인 또는 단백질 국소화 서열을 인코딩하는 핵산 서열에 대해 지칭할 때, "이종성"은 조절 또는 보조 서열 또는 단백질 도메인 또는 국소화 서열을 인코딩하는 서열이, 조절 또는 보조 핵산 서열 또는 단백질 도메인 또는 국소화 서열을 인코딩하는 핵산 서열이 게놈, 염색체 또는 에피솜에 병치되는 유전자와 다른 출처로부터 유래한 것임을 의미한다. 따라서, 자연 상태에서(예를 들어, 비-유전적으로 조작된 유기체의 게놈에서) 작동적으로 연결되지 않은 유전자에 작동적으로 연결된 프로모터는 본 명세서에서 "이종성 프로모터"로 지칭되지만, 프로모터는 연결된 유전자와 동일한 종(또는 경우에 따라 동일한 유기체)로부터 유래될 수 있다. 유사하게, 단백질 국소화 서열 또는 조작된 단백질의 단백질 도메인을 지칭할 때, "이종성"은 국소화 서열 또는 단백질 도메인이 유전 공학에 의해 통합되는 것과 다른 단백질로부터 유래됨을 의미한다.

[0085] 본 명세서에서 사용되는 용어 "재조합(recombinant)" 또는 "조작된(engineered)" 핵산 분자는 인간 개입을 통해 변경된 핵산 분자를 의미한다. 비제한적 예로서, cDNA는 시험관내(invitro) 중합 효소 반응(들)에 의해 생성되었거나 링커가 부착되었거나 클로닝 벡터 또는 발현 벡터 또는 레플리콘과 같은 벡터에 통합된, 임의의 핵산 분자와 같이 재조합 DNA 분자이다. 비제한적인 예로서, 재조합 핵산 분자는: 1) 예를 들어 핵산 분자의 화학적 또는 효소 기술을 사용하여(예를 들어, 화학적 핵산 합성을 사용하거나, 복제, 중합, 엑소뉴클레오리틱 분해, 엔도뉴클레오리틱 분해, 라이게이션, 역전사, 전사, 염기 변형(예컨대, 메틸화 포함) 또는 재조합(상동성 및 부위 특이적 재조합 포함)를 위한 효소를 사용하여) 시험관 내에서 합성 또는 변형되거나; 2) 자연적으로 결합되지 않은 결합된 뉴클레오티드 서열을 포함하거나, 3) 천연적으로 발생하는 핵산 분자 서열과 관련하여 하나 이상의 뉴클레오티드가 결여되도록 분자 클로닝 기술을 사용하여 조작되거나, 및/또는 4) 천연 발생 핵산 서열과 관련

하여 하나 이상의 서열 변경 또는 재배열을 갖도록 하는 분자 클로닝 기술을 사용하여 조작 된다. 비제한적 예로서, cDNA는 시험관내 중합효소 반응(들)에 의해 생성된 임의의 핵산 분자 그대로의 것이거나, 여기에 링커가 부착된 것, 또는 클로닝 벡터 또는 발현 벡터와 같은 벡터 내에 통합되거나 RNA 레플리콘에 통합된, 재조합 DNA 분자이다.

[0086] 방법

[0087] 본 발명은 또한 인간, 포유류 또는 동물 환자에 핵산을 투여하는 방법을 제공한다. 핵산은 단백질 또는 펩티드(이종성 단백질 또는 펩티드일 수 있음)를 인코딩하는 RNA 서열일 수 있다. 상기 방법은 본 명세서에 기재된 RNA 레플리콘을 환자에게 투여하는 것을 포함하며, 여기서 레플리콘에 의해 인코딩된 단백질 또는 펩티드는 환자에서 발현(또는 전사)되며, 이는 환자 신체의 세포 성분을 이용하여 수행될 수 있다. 본원의(비제한적인) 구현예 중 임의의 것에서, 포유류는 인간, 가축, 식용동물 또는 반려동물일 수 있다. 동물은 또한 임의의 새, 물고기(예컨대, Salmonidae 과), 가금류(poultry) 또는 가금류(fowl), 예컨대, 닭, 오리, 거위, 칠면조, 타조, 에뮤, 백조, 공작, 꿩, 자고새 또는 뿔닭일 수 있다. 레플리콘은 약학적으로 허용되는 담체, 예를 들어 식염수, 물 또는 다른 허용되는 담체 내 투여될 수 있다.

[0088] 본 명세서에 사용 된 "약학적으로 허용되는 담체"는 개인에 대한 투여에 사용하기에 적합한 임의의 물질을 의미한다. 예를 들어, 약학적으로 허용되는 담체는 인산염 완충 식염수(PBS) 또는 주사용수와 같은 멸균 수용액일 수 있다. 일부 구현예에서, 약학적으로 허용되는 담체는 완충제, 보존제, 등장제, 안정제, 계면활성제, 습윤제, 유화제, 항산화제, 증량제 또는 킬레이트제일 수 있다. 물론 이러한 추가 성분은 본 발명의 약학적 제형의 전반적인 안정성에 악영향을 미치지 않아야 한다.

[0089] 레플리콘은 또한 약학적으로 허용되는 염으로 투여될 수 있다. 본 명세서에서 사용되는 "약학적으로 허용되는 염"은 핵산 화합물 또는 폴리뉴클레오티드와 같은 화합물의 생리학적 및 약학적으로 허용되는 염, 즉 모화합물의 원하는 생물학적 활성을 유지하고 이에 원치 않는 독성 효과를 부여하지 않는 염을 의미한다. 본 발명에 사용하기 위한 약학적으로 허용되는 산성/음이온 염은 아세테이트(acetate), 벤젠설포네이트(benzenesulfonate), 벤조에이트(benzoate), 바이카보네이트(bicarbonate), 바이타르트레이트(bitartrate), 브로마이드(bromide), 칼슘 에데테이트(calcium edetate), 캠실레이트(camsylate), 카보네이트(carbonate), 클로라이드(chloride), 시트레이트(citrate), 디하이드로클로라이드(dihydrochloride), 에데테이트(edetate), 에디실레이트(edisylate), 에스톨레이트(estolate,), 에실레이트(esylate), 푸마레이트(fumarate), 글리셉테이트(glyceptate), 글루코네이트(gluconate), 글루타메이트(glutamate), 글리콜릴라사닐레이트(glycollylarsanilate), 헥실레조시네이트(hexylresorcinate), 하이드라바민(hydrabamine), 하이드로브로마이드(hydrobromide), 하이드로클로라이드(hydrochloride), 하이드록시나프토에이트(hydroxynaphthoate), 아이오다이드(iodide), 이세티오네이트(isethionate), 락테이트(lactate), 락토비오네이트(lactobionate), 말레이트(malate), 말리에이트(maleate), 만델레이트(mandelate), 메실레이트(mesylate), 메틸브로마이드(methylbromide), 메틸나이트레이트(methylnitrate), 메틸설페이트(methylsulfate), 뮤케이트(mucate), 나프실레이트(napsylate), 나이트레이트(nitrate), 파모에이트(pamoate), 판토테네이트(pantothenate), 포스페이트/디포스페이트(phosphate/diphosphate), 폴리갈락투로네이트(polygalacturonate), 살리실레이트(salicylate), 스테아레이트(stearate), 서브아세테이트(subacetate), 석시네이트(succinate), 설페이트(sulfate), 탄네이트(tannate), 타르트레이트(tartrate), 테오클레이트(teoclate), 토실레이트(tosylate) 및 트리에티오다이드(triethiodide)를 포함하며 이에 제한되지 않는다. 유기산 또는 무기산은 또한 염소화수소산(hydriodic acid), 과염소산(perchloric acid), 황산(sulfuric acid), 인산(phosphoric acid), 프로피온산(propionic acid), 글리콜산(glycolic acid), 메탄설폰산(methanesulfonic acid), 하이드록시에탄설폰산(hydroxyethanesulfonic acid), 옥살산(oxalic acid), 2-나프탈렌설폰산(2-naphthalenesulfonic acid), p-톨루엔설폰산(p-toluenesulfonic acid), 시클로헥산설팜산(cyclohexanesulfamic acid), 사카린산(saccharinic acid) 또는 트리플루오로아세트산(trifluoroacetic acid)을 포함하며 이에 제한되지 않는다. 약학적으로 허용되는 염기성/양이온성 염은 알루미늄, 2-아미노-2-히드록시메틸-프로판-1,3-디올(2-amino-2-hydroxymethyl-propane-1,3-diol; 트리스(히드록시메틸)아미노메탄(tris(hydroxymethyl)aminomethane), 트로메탄(tromethane) 또는 "TRIS" 라고도 함), 암모니아, 벤자틴(benzathine), t-부틸아민(t-butylamine), 칼슘, 클로로프로카인(chloroprocaine), 콜린(choline), 시클로헥실아민(cyclohexylamine), 디에탄올아민(diethanolamine), 에틸렌디아민(ethylenediamine), 리튬(lithium), L-라이신(L-lysine), 마그네슘, 메글루민(meglumine), N-메틸-D-글루카민(N-methyl-D-glucamine), 피페리딘(piperidine), 칼륨, 프로카인(procaine), 퀴닌(quinine,), 나트륨, 트리에탄올아민(triethanolamine) 또는 아연을 포함하며 이에 제한되지 않는다.

[0090] 본 발명은 또한 이종성 단백질 또는 펩티드를 인코딩하는 본 명세서에 기술된 RNA 레플리콘을 포유류에 투여하는 단계를 포함하는 이종성 단백질 또는 펩티드를 포유류에 투여하는 방법을 제공하고, 상기 이종성 단백질 또는 펩티드가 포유류 내에서 발현한다. 상기 방법은 네이키드(naked) 이종성 단백질 및 펩티드의 투여에 비해 포유류로부터의 면역 반응이 낮거나 제거된 이종성 단백질 또는 펩티드의 포유류에 대한 투여를 가능하게 한다.

[0091] 구현예

[0092] 구현예 1은

[0093] 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열;

[0094] 5' 및 3' 알파바이러스 비번역 영역;

[0095] 뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2, 및 nsP4로부터 유래한 아미노산 서열을 인코딩하는 RNA 서열; 및

[0096] 알파바이러스 nsP3 매크로 도메인으로부터 유래한 아미노산 서열을 인코딩하는 RNA 서열;

[0097] 알파바이러스 nsP3 중앙 도메인으로부터 유래한 아미노산 서열을 인코딩하는 RNA 서열; 및

[0098] 알파바이러스 nsP3 초가변 도메인(HVD; hypervariable domain)으로부터 유래한, 다음을 포함하는 아미노산 서열을 인코딩하는 RNA 서열:

[0099] a. 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 아미노산 서열; 또는

[0100] b. 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분, 및 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분을 포함하는 아미노산 서열

[0101] 을 포함하는 RNA 레플리콘이다.

[0102] 구현예 2는 구현예 1의 RNA 레플리콘으로, 상기 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 뉴 월드 알파바이러스로부터 유래한 것이다.

[0103] 구현예 3은 구현예 1의 RNA 레플리콘으로, 상기 알파바이러스 nsP3 매크로 도메인 및 알파바이러스 nsP3 중앙 도메인은 올드 월드 알파바이러스로부터 유래한 것이다.

[0104] 구현예 4는 구현예 1 내지 3 중 어느 하나의 RNA 레플리콘으로, 상기 알파바이러스 nsP3 초가변 도메인은 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 아미노산 서열을 포함한다.

[0105] 구현예 5는 구현예 1 내지 4 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스는 CHIKV, SINV 및 SFV로 이루어진 군에서 선택된다.

[0106] 구현예 6은 구현예 1 내지 5 중 어느 하나의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis, VEEV)이다.

[0107] 구현예 7은 구현예 1 내지 5 중 어느 하나의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV이다.

[0108] 구현예 8은 구현예 1 내지 5 중 어느 하나의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis virus, VEEV), 웨스턴 이콰인 뇌염 바이러스(western equine encephalitis virus, WEEV), 및 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV)로 이루어진 군에서 선택된다.

[0109] 구현예 9는 구현예 1 내지 8 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스는 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV), 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV), 로스 리버 바이러스(Ross River Virus, RRV), 사기야마 바이러스(Sagiyama virus, SAGV), 게타 바이러스(Getah virus, GETV), 미들버그 바이러스(Middleburg virus, MIDV), 베바루 바이러스(Bebaru virus, BEBV), 오뇽뇽 바이러스(O'nyong nyong virus, ONNV), 엔두무 바이러스(Ndumu, NDUV), 및 바르마 포레스트 바이러스(Barmah Forest virus, BFV)로 이루어진 군에서 선택된다.

[0110] 구현예 10은 구현예 1 내지 9 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분은 FGDF (서열번호 18) 및 FGSF (서열번호 19)로 이루어진 군에서 선택되는 모티프를

포함한다.

[0111] 구현예 11은 구현예 1 내지 10 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분은 FGDF/FGDF (서열번호 20) 반복, FGSF/FGSF (서열번호 21) 반복, FGDF/FGSF (서열번호 22) 반복, 및 FGSF/FGDF (서열번호 23) 반복으로 이루어진 군에서 선택되는 반복을 포함하고; 및 추가로 상기 반복 서열은 적어도 10 및 25 이하의 아미노산에 의해 분리된다.

[0112] 구현예 12는 구현예 11의 RNA 레플리콘으로, 상기 반복 서열은 NEGEIESLSSELLT (서열번호 6), SDGEIDELSRRVTTESEPVL (서열번호 7), 및 DEHEVDALASGIT (서열번호 8)로 이루어진 군에서 유래한 아미노산 서열에 의해 분리된다.

[0113] 구현예 13은 구현예 1 내지 12 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함한다:

[0114] CHIKV nsP3 HVD의 아미노산 479-482 또는 497-500 또는 479-500 또는 335-517; 또는

[0115] SFV nsP3 HVD의 아미노산 451-454 또는 468-471 또는 451-471; 또는

[0116] SINV nsP3 HVD의 아미노산 490-493 또는 513-516 또는 490-516 또는 335-538.

[0117] 구현예 13은 구현예 1 내지 12 중 어느 하나의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래된 부분은 다음을 포함한다:

[0118] CHIKV nsP3 HVD의 아미노산 479-500 또는 335-517; 또는

[0119] SFV nsP3 HVD의 아미노산 451-471; 또는

[0120] SINV nsP3 HVD의 아미노산 490-516.

[0121] 구현예 15는 구현예 13의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 VEEV nsP3 초가변 도메인의 아미노산 478-518을 포함하지 않는다.

[0122] 구현예 16은 구현예 13의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 VEEV nsP3 초가변 도메인의 아미노산 478-545를 포함하지 않는다.

[0123] 구현예 17은 구현예 13의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 VEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않는다.

[0124] 구현예 18은 구현예 14의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스는 CHIKV이고, 올드 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 CHIKV의 아미노산 335-517을 포함한다.

[0125] 구현예 19는 구현예 14의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스는 SINV이고, 올드 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 SINV의 아미노산 335-538을 포함한다.

[0126] 구현예 20은 구현예 13의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 EEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 EEEV nsP3 초가변 도메인의 아미노산 531-547을 포함하지 않는다.

[0127] 구현예 21은 구현예 20의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 EEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 EEEV nsP3 초가변 도메인의 아미노산 531-547을 포함하지 않고, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 다음을 포함한다:

[0128] CHIKV nsP3 HVD의 아미노산 479-500;

[0129] SFV nsP3 HVD의 아미노산 451-471; 또는

[0130] SINV nsP3 HVD의 아미노산 490-516.

[0131] 구현예 22는 구현예 13의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 WEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 WEEV nsP3 초가변 도메인의 아미노산 504-520을 포함하지 않는다.

[0132] 구현예 23은 구현예 22의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 WEEV이고, 뉴 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 WEEV nsP3 초가변 도메인의 아미노산 504-520을 포함하지 않고, 상기 올드 월드 알파바이러스 초가변 도메인으로부터 유래한 부분은 다음을 포함한다:

[0133] CHIKV nsP3 HVD의 아미노산 479-500;

[0134] SFV nsP3 HVD의 아미노산 451-471; 또는

[0135] SINV nsP3 HVD의 아미노산 490-516.

[0136] 구현예 24는 구현예 1 내지 23 중 어느 하나의 RNA 레플리콘으로, 이종성 단백질을 인코딩하는 RNA 서열에 작동적으로 연결되고 이종성 단백질을 인코딩하는 RNA 서열의 번역을 제어하는 서브게놈 프로모터를 추가로 포함한다.

[0137] 구현예 25는 구현예 1 내지 24 중 어느 하나의 RNA 레플리콘으로, 5' 캡 및 3' 폴리-A 테일을 추가로 포함한다.

[0138] 구현예 26은 구현예 1 내지 25 중 어느 하나의 RNA 레플리콘으로, 상기 레플리콘은 양성 센스, 단일 가닥 RNA를 포함한다.

[0139] 구현예 27은 구현예 1 내지 26 중 어느 하나의 RNA 레플리콘으로, 상기 레플리콘은 10-12 kb의 RNA를 포함하고 30-50 nm의 직경을 가진다.

[0140] 구현예 28은 구현예 1 내지 27 중 어느 하나의 RNA 레플리콘으로, 상기 이종성 단백질은 생체치료제 단백질 또는 펩티드이다.

[0141] 구현예 29는 구현예 1 내지 28 중 어느 하나의 RNA 레플리콘으로, 상기 이종성 단백질은 항체이다.

[0142] 구현예 30은 구현예 1의 RNA 레플리콘으로, 상기 뉴 월드 알파바이러스는 VEEV이고, 상기 알파바이러스 nsP3 초가변 도메인은 VEEV nsP3 초가변 도메인의 아미노산 335-518을 포함하지 않는 뉴 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분, 및 SINV nsP3 HVD의 아미노산 490-493 또는 513-516 또는 490-516 또는 335-538을 포함하는 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분을 포함한다.

[0143] 구현예 31은 구현예 30의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분은 SINV nsP3 HVD의 아미노산 490-516을 포함한다.

[0144] 구현예 32는 구현예 30의 RNA 레플리콘으로, 상기 올드 월드 알파바이러스는 SINV이고 상기 올드 월드 알파바이러스 nsP3 초가변 도메인으로부터 유래한 부분은 SINV nsP3 HVD의 아미노산 335-538을 포함한다.

[0145] 구현예 33은 구현예 1 내지 32 중 어느 하나의 RNA 레플리콘이고, 상기 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열은 nsP1, nsP2 및 nsP4를 인코딩하는 RNA 서열에 작동적으로 연결된다.

[0146] 구현예 34는 이종성 단백질 또는 펩티드를 인코딩하는 제1항 내지 제33항 중 어느 한 항의 RNA 레플리콘을 포유류에 투여하는 단계를 포함하는 포유류에 이종성 단백질 또는 펩티드를 투여하는 방법으로, 상기 이종성 단백질 또는 펩티드는 포유류 내에서 발현된다.

[0147] 구현예 35는 구현예 34의 방법으로, 상기 뉴 월드 알파바이러스는 VEEV이고, RNA 레플리콘은 구현예 14의 레플리콘이다.

[0148] 구현예 36은 구현예 34의 방법으로, 상기 RNA 레플리콘은 구현예 19의 레플리콘이다.

[0149] 구현예 37은 구현예 34의 방법으로, 상기 RNA 레플리콘은 구현예 22의 레플리콘이다.

[0150] 구현예 38은 구현예 34의 방법으로, 상기 RNA 레플리콘은 구현예 25의 레플리콘이다.

[0151] 구현예 39는 RNA 레플리콘으로, 다음을 포함한다:

[0152] 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열;

[0153] 뉴 월드 알파바이러스 비구조 단백질 nsP1, nsP2 및 nsP4로부터 유래한 아미노산 서열을 인코딩하는 RNA 서열; 및

[0154] 올드 월드 알파바이러스 nsP3 단백질로부터 유래한 아미노산 서열을 인코딩하는 RNA 서열로, 상기 nsP3 단백질의 N 말단 및/또는 C 말단 측의 첫 1-6 아미노산은 뉴 월드 알파바이러스 서열로부터 유래한, RNA 서열.

[0155] 구현예 40은 알파바이러스 레플리콘 RNA로, 다음을 5' 부터 3' 말단까지의 순서로 포함한다:

[0156] (1) 알파바이러스 레플리콘의 복제를 지시하는 알파바이러스 5' 비번역 서열;

[0157] (2) 알파바이러스 비구조 단백질 nsP1, nsP2, nsP3, 및 nsP4을 인코딩하는 RNA 서열;

[0158] (3) 알파바이러스 서브게놈 프로모터 서열,

[0159] (4) 하나 이상의 이종성 단백질 또는 펩티드를 인코딩하는 RNA 서열, 및

[0160] (5) 알파바이러스 3' 비번역 서열,

[0161] 여기서 nsP1, nsP2 및 nsP4는 하나 이상의 뉴 월드 알파바이러스로부터 유래하고,

[0162] 상기 nsP3는 아미노 말단으로부터 카르복실 말단까지 순서대로, 매크로 도메인, 중앙 도메인 및 초가변 도메인(HVD)를 포함하고, 여기서

[0163] 상기 매크로 도메인 및 중앙 도메인은 하나 이상의 뉴 월드 알파바이러스 및/또는 올드 월드 알파바이러스로부터 유래하고; 및

[0164] 상기 HVD는 올드 월드 알파바이러스의 HVD로부터 유래하거나, 또는 상기 HVD는 뉴 월드 알파바이러스로부터 유래한 HVD의 부분 및 올드 월드 알파바이러스로부터 유래한 HVD의 부분을 포함한다.

[0165] 구현예 41은 구현예 40의 알파바이러스 레플리콘 RNA로, 상기 nsP1, nsP2 및 nsP4는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis virus, VEEV), 웨스턴 이콰인 뇌염 바이러스(western equine encephalitis virus, WEEV), 및 이스턴 이콰인 뇌염 바이러스(eastern equine encephalitis virus, EEEV)로 이루어진 군에서 선택되는 하나 이상의 뉴 월드 알파바이러스로부터 유래한 것이다.

[0166] 구현예 42는 구현예 40의 알파바이러스 레플리콘 RNA로, 상기 nsP1, nsP2 및 nsP4는 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis, VEEV)로부터 유래한 것이다.

[0167] 구현예 43은 구현예 40 내지 42 중 어느 하나의 알파바이러스 레플리콘 RNA로, 상기 올드 월드 알파바이러스는 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV), 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV), 로스 리버 바이러스(Ross River Virus, RRV), 사기야마 바이러스(Sagiyama virus, SAGV), 게타 바이러스(Getah virus, GETV), 미들버그 바이러스(Middleburg virus, MIDV), 베바루 바이러스(Bebaru virus, BEBV), 오뇽뇽 바이러스(O'nyong nyong virus, ONNV), 엔두무 바이러스(Ndumu, NDUV), 및 바르마 포레스트 바이러스(Barmah Forest virus, BFV)로 이루어진 군에서 선택된다.

[0168] 구현예 44는 구현예 43의 알파바이러스 레플리콘 RNA로, 상기 올드 월드 알파바이러스는 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV) 또는 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV)이다.

[0169] 구현예 45는 구현예 40 내지 44 중 어느 하나의 알파바이러스 레플리콘 RNA로, 상기 매크로 도메인 및 중앙 도메인은 하나 이상의 올드 월드 알파바이러스로부터 유래한다.

[0170] 구현예 46은 구현예 40 내지 44 중 어느 하나의 알파바이러스 레플리콘 RNA로, 상기 매크로 도메인 및 중앙 도메인은 하나 이상의 뉴 월드 알파바이러스로부터 유래한다.

[0171] 구현예 47은 구현예 40의 알파바이러스 레플리콘 RNA로, 상기 nsP1, nsP2 및 nsP4, 상기 매크로 도메인 및 중앙 도메인은 베네주엘라 이콰인 뇌염 바이러스(Venezuelan equine encephalitis, VEEV)로부터 유래하고, 상기 HVD는 VEEV로부터 유래한 HVD의 부분과 신드비스 바이러스(Sindbis virus, SINV), 치쿤군야 바이러스(Chickungunya virus, CHIKV) 및 셈리키 포레스트 바이러스(Semliki Forest Virus, SFV)로 이루어진 군에서 선택되는 올드 월드 알파바이러스로부터 유래한 HVD의 부분을 포함한다.

[0172] 구현예 48은 구현예 47의 알파바이러스 레플리콘 RNA로, 상기 HVD는 VEEV로부터 유래한 HVD를 포함하되, VEEV의 nsP3의 아미노산 잔기 335-538이 SINV의 nsP3의 아미노산 잔기 335-538로 치환된다.

[0173] 구현예 49는 구현예 47의 알파바이러스 레플리콘 RNA로, 상기 HVD는 VEEV로부터 유래한 HVD를 포함하되, VEEV의 nsP3의 아미노산 잔기 335-518이 CHIKV의 nsP3의 아미노산 잔기 335-517로 치환된다.

[0174] 실시예 1 - VEEV 기반 레플리콘의 면역원성

[0175] VEEV nsP3의 아미노산 335-518을 인코딩하는 뉴클레오티드 서열을 치쿤쿤야(CHIKV) nsP3의 아미노산 335-517을 인코딩하는 뉴클레오티드 서열로 교체하여 돌연변이 nsP3를 인코딩하는 VEEV 기반 알파바이러스 레플리콘을 구축하여 VEEV/CHIKV nsP3 키메라(서열번호 30)를 발현하는 VEEV 기반 레플리콘을 생성하였다. 이 교체에서 VEEV

로부터 반복 서열의 첫번째 모티프를 제거하고, 이를 CHIKV 게놈(아미노산 479-482 및 497-500에서)의 FGDF/FGDF (서열 번호 20) 반복 서열로 교체하였다. 병행 실험에서, VEEV nsP3(HVD 영역)의 아미노산 335-538을 신드비스 바이러스(SINV) nsP3 아미노산(HVD 영역)의 아미노산 335-538로 교체하여 VEEV/SINV nsP3 키메라(서열번호 31) (도 2 및 7 참조)를 인코딩하는 레플리콘을 생성하였다. 이 교체에서 VEEV로부터 반복 서열을 제거하고 이를 SINV로부터의 FGSF/FGSF(서열번호 21) 반복 서열로 교체했다. WT, VEEV/CHIKV 또는 VEEV/SINV 키메라 nsP3를 포함하고 서브 게놈 RNA(SGIα-rFF)로부터 적색 반딧불이 루시퍼라제(red firefly luciferase, rFF) 리포터를 발현하는 레플리콘을 전기천공에 의해 삼중복(triplicate)으로 BHK-21 세포로 전달하였다. 전기천공 후, 세포의 일부를 6 웰 플레이트 중 하나의 웰과 96 웰 플레이트 중 하나의 웰에 플레이팅하고 20시간 동안 회수되도록 하였다. 전기천공된 세포를 dsRNA의 존재에 대해 염색하고 유세포 분석법으로 분석하여 레플리콘 증폭의 척도로 dsRNA 양성 세포의 빈도를 결정했다. 돌연변이 nsP3를 포함하는 레플리콘은 WT nsP3를 포함하는 레플리콘과 동일한 수준으로 복제되는 것이 밝혀졌다(도 2b). 루시퍼라제 활성에 대해 분석했을 때 WT를 포함하는 레플리콘 또는 표시된 nsP3의 돌연변이 형태간에 차이가 보이지 않았다(도 2c).

[0176] 실시예 2 - 레플리콘에서 이종 단백질의 발현

[0177] 이 실시예는 도 2a에 기술된 바와 같이 돌연변이 nsP3를 인코딩하는 레플리콘(실시예 1)으로부터 재조합 반딧불이 루시퍼라제(rFF)의 생체 내(*in vivo*) 발현을 조사하였다. 식염수 내의 1 또는 10 마이크로그램의 레플리콘 RNA를 BALB/c 마우스의 대퇴사두근 내에 근육 내(IM) 전달했다. 표시된 시점에서 루시퍼라제 활성은 시판 생체 내 이미징 시스템을 사용하여 생체 내에서 모니터링 되었고 총 플럭스로 보고되었다(도 3a 및 3b). 데이터는 nsP3의 돌연변이 형태를 발현하는 레플리콘이 VEEV 유래의 야생형 nsP3를 함유하는 레플리콘과 비교하여 생체 내에서 유사한 수준의 루시퍼라제 활성을 나타냄을 보여준다.

[0178] 실시예 3 - 면역원성

[0179] 이 실시예는 nsP3의 VEEV/CHIKV 키메라 형태를 인코딩하는 VEEV 기반 레플리콘(실시예 1로부터)의 면역원성 대비 야생형(WT) VEEV nsP3를 갖는 레플리콘의 면역원성을 조사했다. 각 레플리콘은 이종성 단백질로서 인플루엔자의 H5N1 균주 유래의 헤마글루티닌(HA)을 인코딩하고 발현했다. 식염수 내의 2.0 ug 또는 0.2 ug의 RNA를 0일에 BALB/c 마우스의 대퇴사 두근에 근육 내로 전달하고 28 일에 동일 레플리콘 RNA 및 용량으로 부스팅했다. 부스팅 후 2주(프라임 후 42일) 비장 및 혈청을 수집하였다. 혈청은 ELISA에 의해 HA 특이적 항체에 대해 분석되었다(도 4). 데이터는 VEEV/CHIKV nsP3 키메라를 인코딩하는 레플리콘이 야생형 nsP3를 갖는 레플리콘과 비교하여 HA 특이적 IgG 역가를 상당히 감소시켰음을 나타낸다.

[0180] 반대로, 단기 이펙터 및 메모리 전구체 이펙터 CD8+ T 세포 분석은 시험된 상이한 레플리콘 사이에 HA 특이적 세포의 빈도에 차이가 없음을 나타낸다(도 5a 및 5b). 도 5a는 야생형, VEEV/SINV nsP3 및 VEEV/CHIKV nsP3 RNA 레플리콘 사이의 HA 특이적 단기 이펙터 CD8+ T 세포의 유사한 빈도를 나타낸다. 도 5b는 메모리 이펙터 CD8+ T 세포에 대한 유사한 결과를 나타낸다.

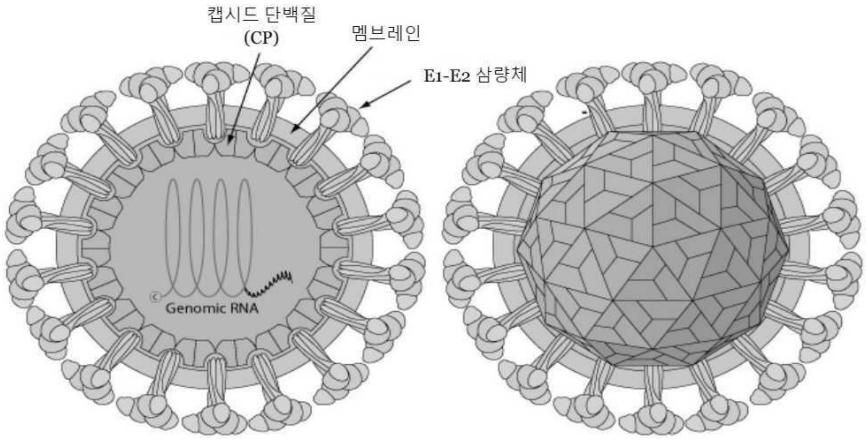
[0181] 본 발명의 범위 및 사상을 벗어나지 않고 본 명세서에 개시된 본 발명에 대해 다양한 치환 및 수정이 이루어질 수 있다는 것은 당업자에게 명백할 것이다.

[0182] 명세서에 언급된 모든 특허 및 간행물은 본 발명이 속하는 기술 분야의 숙련자의 수준의 지표이다.

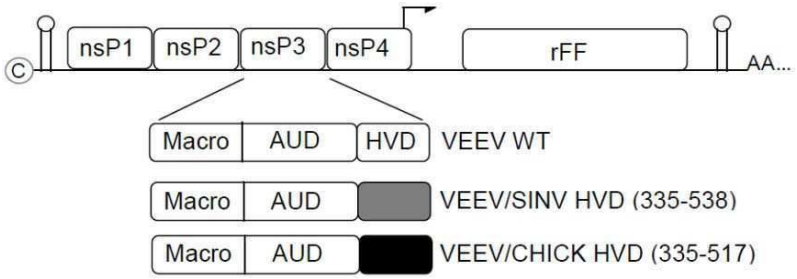
[0183] 본 명세서에 예시적으로 설명된 본 발명은 본 명세서에 구체적으로 개시되지 않은 임의의 요소 또는 요소들, 제한 또는 제한들의 부재 하에 실시될 수 있다. 따라서, 예를 들어, 본 명세서의 각 경우에서 용어 "포함하는", "본질적으로 구성되는" 및 "구성되는"은 다른 두 용어 중 하나로 대체될 수 있다. 사용된 용어 및 표현은 기술을 위한 용어이며 제한이 아니며, 이러한 용어 및 표현에 있어서 제시 및 기술된 특징의 임의의 균등물 또는 그 부분을 배제하려는 의도는 없으나, 청구된 본 발명의 범위 내에서 다양한 변형이 가능함을 인식할 것이다. 또한, 본 발명의 특징 또는 양태가 Markush 군의 관점에서 기술되는 경우, 당업자는 본 발명이 Markush 군의 구성원의 임의의 개별 구성원 또는 서브그룹의 관점에서도 설명된다는 것을 인식할 것이다. 예를 들어, X가 브롬, 염소 및 요오드로 이루어진 군에서 선택된 것으로 기술되는 경우, X는 브롬인 청구항 및 X는 브롬 및 염소인 청구항이 완전히 기술된다. 다른 구현예는 다음의 청구범위 내에 있다.

도면

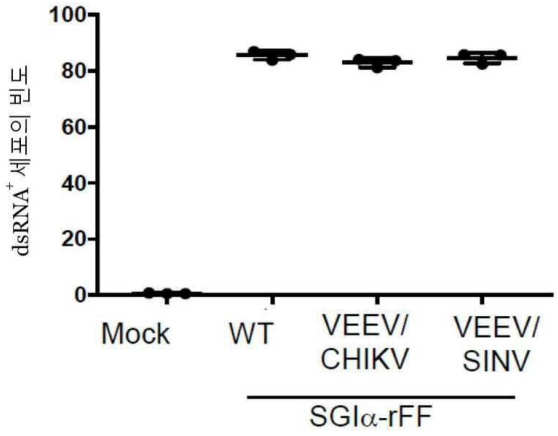
도면1



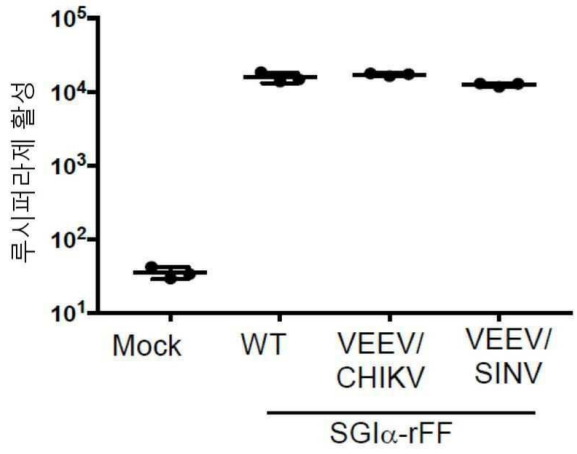
도면2a



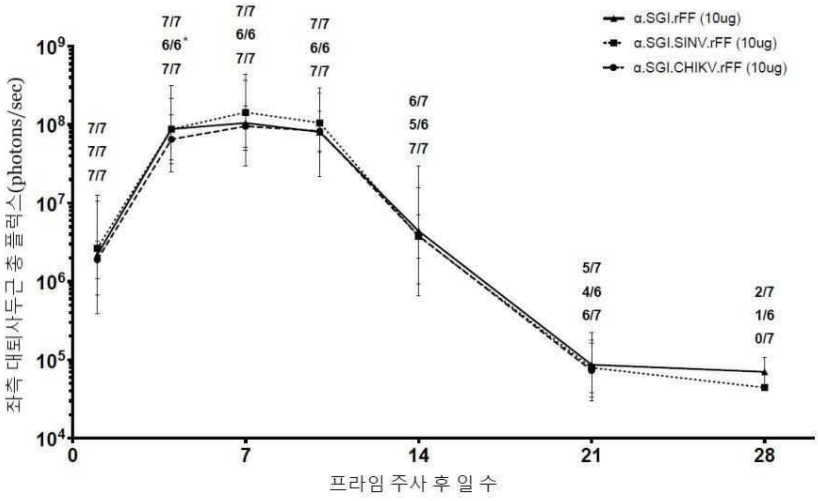
도면2b



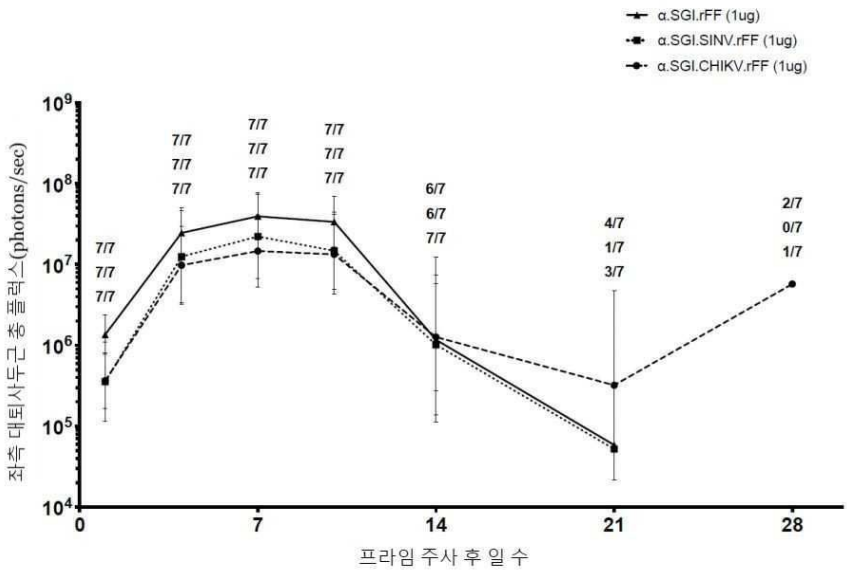
도면2c



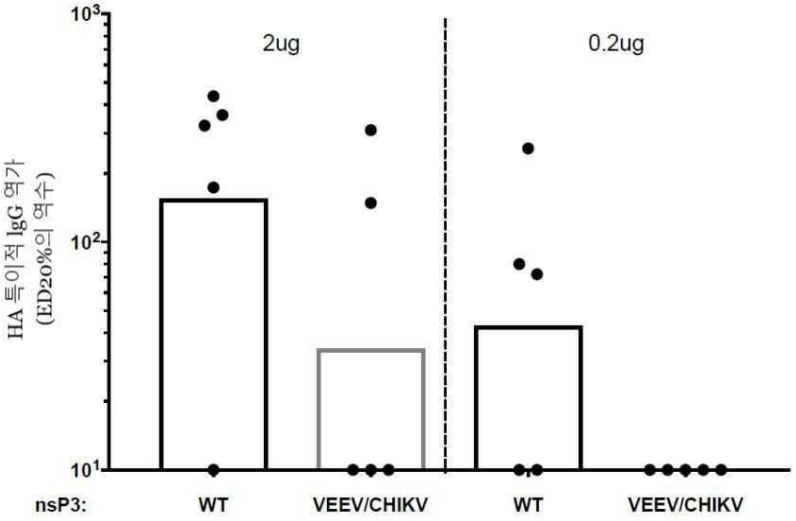
도면3a



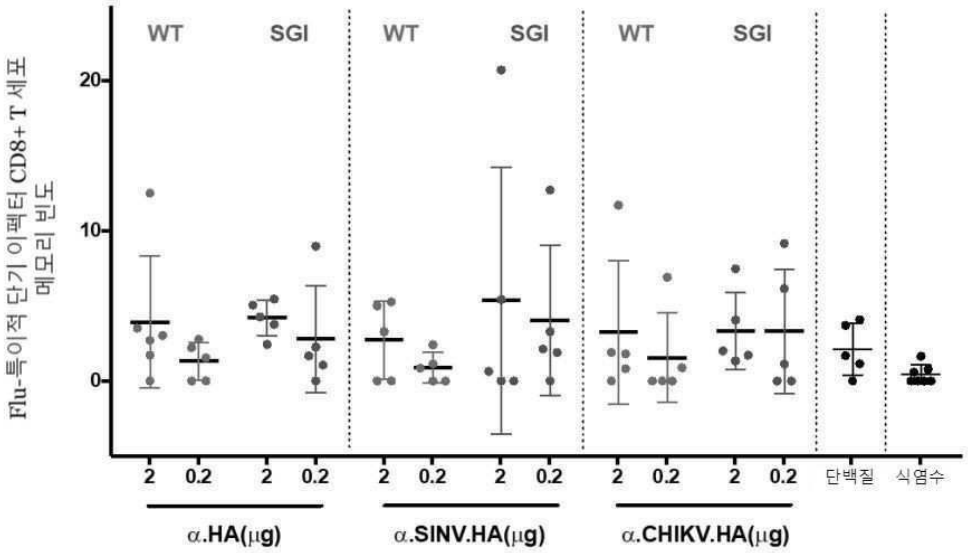
도면3b



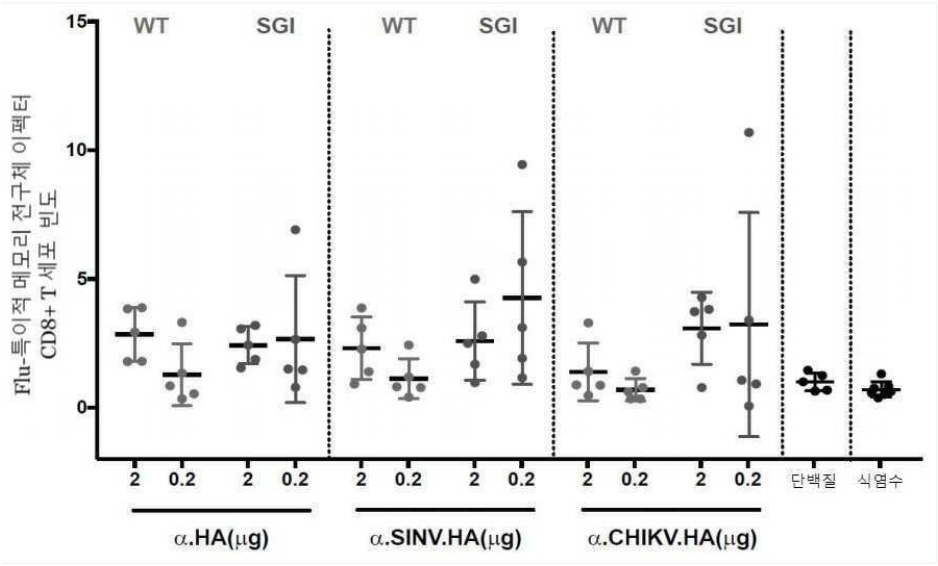
도면4


2ug
0.2ug
HA 특이적 IgG 역가
(ED20%의 역수)
nsP3:
WT
VEEV/CHIKV
WT
VEEV/CHIKV

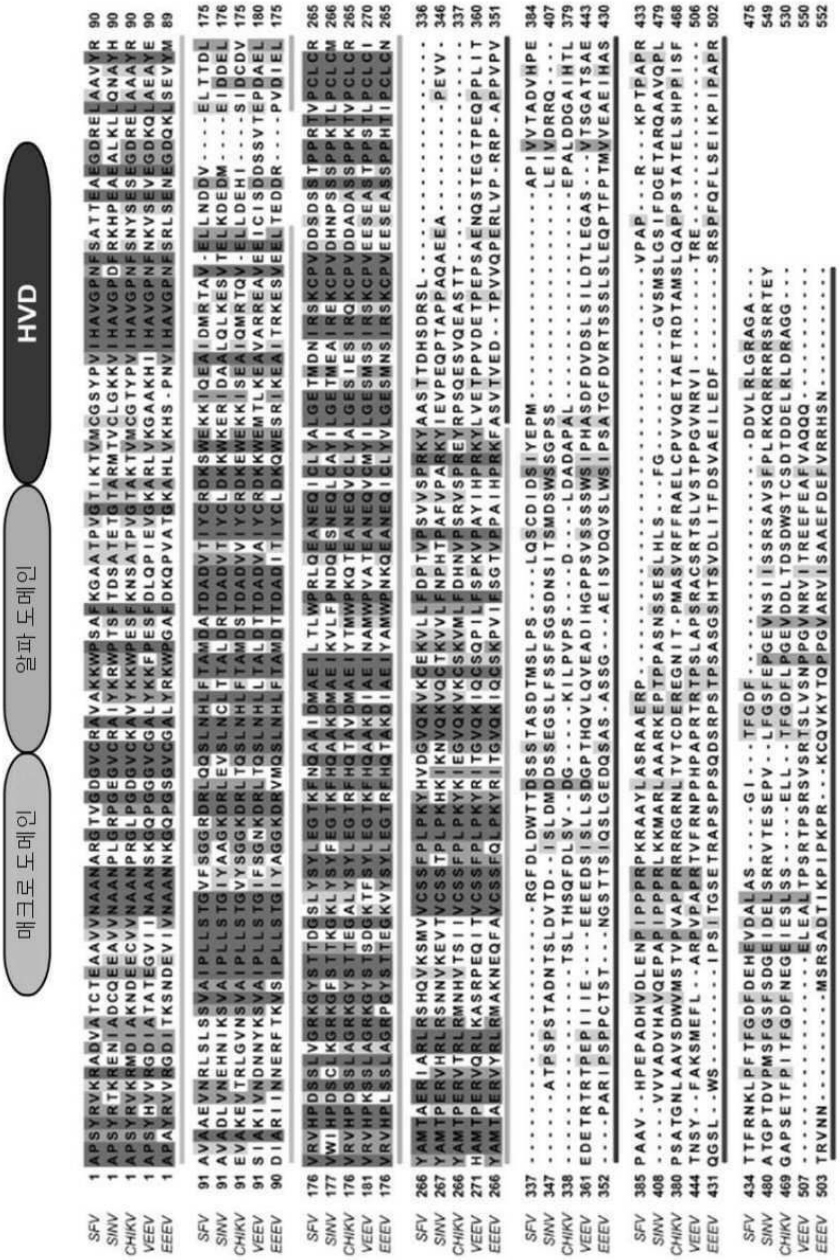
도면5a


Flu-특이적 단기 이펙터 CD8+ T 세포
메모리 빈도
WT
SGI
WT
SGI
WT
SGI
2
0.2
2
0.2
2
0.2
2
0.2
2
0.2
2
0.2
단백질
식염수
α.HA(μg)
α.SINV.HA(μg)
α.CHIKV.HA(μg)

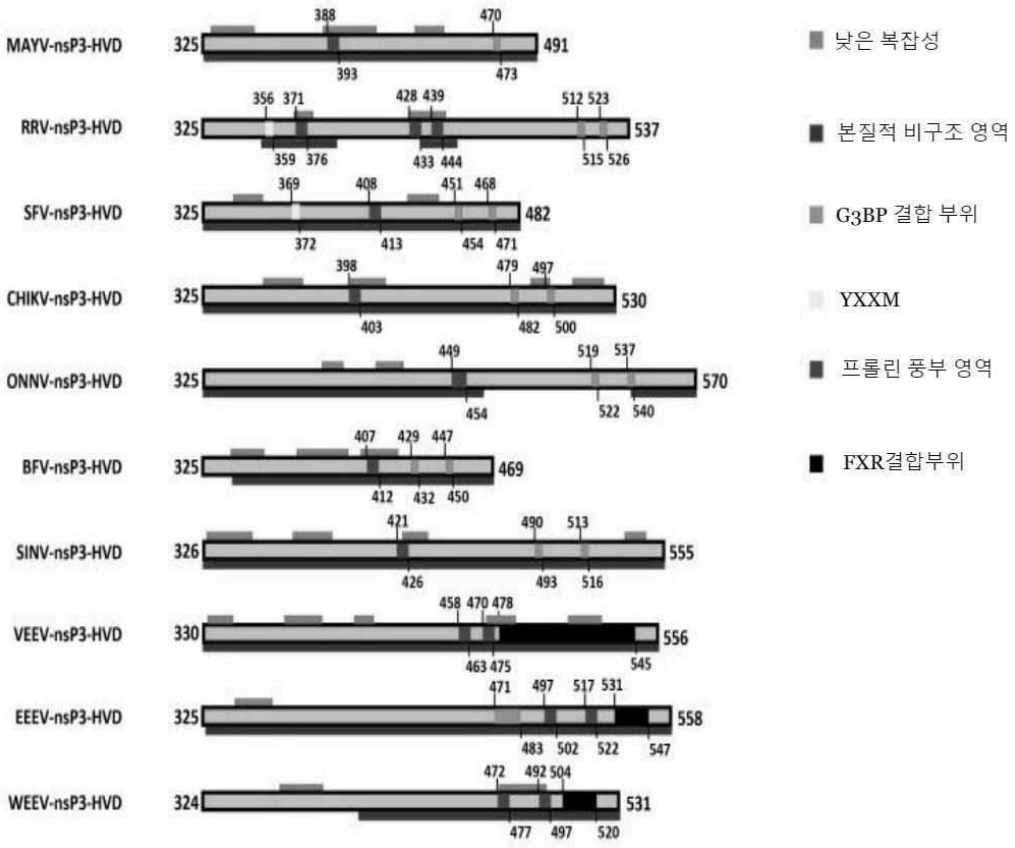
도면5b


Flu-특이적 메모리 전구체 이펙터 CD8+ T 세포 빈도
15
10
5
0
WT
SGI
WT
SGI
WT
SGI
2
0.2
2
0.2
2
0.2
2
0.2
2
0.2
2
0.2
단백질
식염수
α.HA(μg)
α.SINV.HA(μg)
α.CHIKV.HA(μg)

도면6



도면7



서 열 목 록

SEQUENCE LISTING

<110> Janssen Pharmaceuticals, Inc.

<120> Alphavirus-Based Replicons for Administration of Biotherapeutics

<130> 689405.107WO

<150> US 62/742,868

<151> 2018-10-08

<160> 31

<170> PatentIn version 3.5

<210> 1

<211> 2492

<212> PRT

<213> Venezuelan equine encephalitis virus

<400> 1

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 Gly Ser Val Glu Thr Pro Arg Gly Leu

530 535 540
Ile Lys Val Thr Ser Tyr Ala Gly Glu Asp Lys Ile Gly Ser Tyr Ala
545 550 555 560
Val Leu Ser Pro Gln Ala Val Leu Lys Ser Glu Lys Leu Ser Cys Ile
565 570 575
His Pro Leu Ala Glu Gln Val Ile Val Ile Thr His Ser Gly Arg Lys
580 585 590
Gly Arg Tyr Ala Val Glu Pro Tyr His Gly Lys Val Val Val Pro Glu

595 600 605
Gly His Ala Ile Pro Val Gln Asp Phe Gln Ala Leu Ser Glu Ser Ala
610 615 620
Thr Ile Val Tyr Asn Glu Arg Glu Phe Val Asn Arg Tyr Leu His His
625 630 635 640
Ile Ala Thr His Gly Gly Ala Leu Asn Thr Asp Glu Glu Tyr Tyr Lys
645 650 655
Thr Val Lys Pro Ser Glu His Asp Gly Glu Tyr Leu Tyr Asp Ile Asp

660 665 670
Arg Lys Gln Cys Val Lys Lys Glu Leu Val Thr Gly Leu Gly Leu Thr
675 680 685
Gly Glu Leu Val Asp Pro Pro Phe His Glu Phe Ala Tyr Glu Ser Leu
690 695 700
Arg Thr Arg Pro Ala Ala Pro Tyr Gln Val Pro Thr Ile Gly Val Tyr
705 710 715 720
Gly Val Pro Gly Ser Gly Lys Ser Gly Ile Ile Lys Ser Ala Val Thr

725 730 735
Lys Lys Asp Leu Val Val Ser Ala Lys Lys Glu Asn Cys Ala Glu Ile
740 745 750
Ile Arg Asp Val Lys Lys Met Lys Gly Leu Asp Val Asn Ala Arg Thr

```
        755                 760                 765
Val Asp Ser Val Leu Leu Asn Gly Cys Lys His Pro Val Glu Thr Leu
    770                 775                 780
Tyr Ile Asp Glu Ala Phe Ala Cys His Ala Gly Thr Leu Arg Ala Leu

785                 790                 795                 800
Ile Ala Ile Ile Arg Pro Lys Lys Ala Val Leu Cys Gly Asp Pro Lys
                805                 810                 815
Gln Cys Gly Phe Phe Asn Met Met Cys Leu Lys Val His Phe Asn His
            820                 825                 830
Glu Ile Cys Thr Gln Val Phe His Lys Ser Ile Ser Arg Arg Cys Thr
        835                 840                 845
Lys Ser Val Thr Ser Val Val Ser Thr Leu Phe Tyr Asp Lys Lys Met

    850                 855                 860
Arg Thr Thr Asn Pro Lys Glu Thr Lys Ile Val Ile Asp Thr Thr Gly
865                 870                 875                 880
Ser Thr Lys Pro Lys Gln Asp Asp Leu Ile Leu Thr Cys Phe Arg Gly
                885                 890                 895
Trp Val Lys Gln Leu Gln Ile Asp Tyr Lys Gly Asn Glu Ile Met Thr
            900                 905                 910
Ala Ala Ala Ser Gln Gly Leu Thr Arg Lys Gly Val Tyr Ala Val Arg

        915                 920                 925
Tyr Lys Val Asn Glu Asn Pro Leu Tyr Ala Pro Thr Ser Glu His Val
    930                 935                 940
Asn Val Leu Leu Thr Arg Thr Glu Asp Arg Ile Val Trp Lys Thr Leu
945                 950                 955                 960
Ala Gly Asp Pro Trp Ile Lys Thr Leu Thr Ala Lys Tyr Pro Gly Asn
                965                 970                 975
Phe Thr Ala Thr Ile Glu Glu Trp Gln Ala Glu His Asp Ala Ile Met

            980                 985                 990
Arg His Ile Leu Glu Arg Pro Asp  Pro Thr Asp Val Phe  Gln Asn Lys
        995                 1000                1005
```

Ala Asn Val Cys Trp Ala Lys Ala Leu Val Pro Val Leu Lys Thr
1010 1015 1020
Ala Gly Ile Asp Met Thr Thr Glu Gln Trp Asn Thr Val Asp Tyr
1025 1030 1035
Phe Glu Thr Asp Lys Ala His Ser Ala Glu Ile Val Leu Asn Gln

1040 1045 1050
Leu Cys Val Arg Phe Phe Gly Leu Asp Leu Asp Ser Gly Leu Phe
1055 1060 1065
Ser Ala Pro Thr Val Pro Leu Ser Ile Arg Asn Asn His Trp Asp
1070 1075 1080
Asn Ser Pro Ser Pro Asn Met Tyr Gly Leu Asn Lys Glu Val Val
1085 1090 1095
Arg Gln Leu Ser Arg Arg Tyr Pro Gln Leu Pro Arg Ala Val Ala

1100 1105 1110
Thr Gly Arg Val Tyr Asp Met Asn Thr Gly Thr Leu Arg Asn Tyr
1115 1120 1125
Asp Pro Arg Ile Asn Leu Val Pro Val Asn Arg Arg Leu Pro His
1130 1135 1140
Ala Leu Val Leu His His Asn Glu His Pro Gln Ser Asp Phe Ser
1145 1150 1155
Ser Phe Val Ser Lys Leu Lys Gly Arg Thr Val Leu Val Val Gly

1160 1165 1170
Glu Lys Leu Ser Val Pro Gly Lys Met Val Asp Trp Leu Ser Asp
1175 1180 1185
Arg Pro Glu Ala Thr Phe Arg Ala Arg Leu Asp Leu Gly Ile Pro
1190 1195 1200
Gly Asp Val Pro Lys Tyr Asp Ile Ile Phe Val Asn Val Arg Thr
1205 1210 1215
Pro Tyr Lys Tyr His His Tyr Gln Gln Cys Glu Asp His Ala Ile

1220 1225 1230
Lys Leu Ser Met Leu Thr Lys Lys Ala Cys Leu His Leu Asn Pro

1235 1240 1245

Gly Gly Thr Cys Val Ser Ile Gly Tyr Gly Tyr Ala Asp Arg Ala

1250 1255 1260

Ser Glu Ser Ile Ile Gly Ala Ile Ala Arg Gln Phe Lys Phe Ser

1265 1270 1275

Arg Val Cys Lys Pro Lys Ser Ser Leu Glu Glu Thr Glu Val Leu

1280 1285 1290

Phe Val Phe Ile Gly Tyr Asp Arg Lys Ala Arg Thr His Asn Pro

1295 1300 1305

Tyr Lys Leu Ser Ser Thr Leu Thr Asn Ile Tyr Thr Gly Ser Arg

1310 1315 1320

Leu His Glu Ala Gly Cys Ala Pro Ser Tyr His Val Val Arg Gly

1325 1330 1335

Asp Ile Ala Thr Ala Thr Glu Gly Val Ile Ile Asn Ala Ala Asn

1340 1345 1350

Ser Lys Gly Gln Pro Gly Gly Gly Val Cys Gly Ala Leu Tyr Lys

1355 1360 1365

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

1370 1375 1380

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

1385 1390 1395

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

1400 1405 1410

Leu Ala Glu Ala Tyr Glu Ser Ile Ala Lys Ile Val Asn Asp Asn

1415 1420 1425

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

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Lys Trp Glu Met Thr Leu Lys Glu Ala Val Ala Arg Arg
1475 1480 1485
Glu Ala Val Glu Glu Ile Cys Ile Ser Asp Asp Ser Ser Val Thr
1490 1495 1500
Glu Pro Asp Ala Glu Leu Val Arg Val His Pro Lys Ser Ser Leu
1505 1510 1515
Ala Gly Arg Lys Gly Tyr Ser Thr Ser Asp Gly Lys Thr Phe Ser

1520 1525 1530
Tyr Leu Glu Gly Thr Lys Phe His Gln Ala Ala Lys Asp Ile Ala
1535 1540 1545
Glu Ile Asn Ala Met Trp Pro Val Ala Thr Glu Ala Asn Glu Gln
1550 1555 1560
Val Cys Met Tyr Ile Leu Gly Glu Ser Met Ser Ser Ile Arg Ser
1565 1570 1575
Lys Cys Pro Val Glu Glu Ser Glu Ala Ser Thr Pro Pro Ser Thr

1580 1585 1590
Leu Pro Cys Leu Cys Ile His Ala Met Thr Pro Glu Arg Val Gln
1595 1600 1605
Arg Leu Lys Ala Ser Arg Pro Glu Gln Ile Thr Val Cys Ser Ser
1610 1615 1620
Phe Pro Leu Pro Lys Tyr Arg Ile Thr Gly Val Gln Lys Ile Gln
1625 1630 1635
Cys Ser Gln Pro Ile Leu Phe Ser Pro Lys Val Pro Ala Tyr Ile

1640 1645 1650
His Pro Arg Lys Tyr Leu Val Glu Thr Pro Pro Val Asp Glu Thr
1655 1660 1665
Pro Glu Pro Ser Ala Glu Asn Gln Ser Thr Glu Gly Thr Pro Glu
1670 1675 1680
Gln Pro Pro Leu Ile Thr Glu Asp Glu Thr Arg Thr Arg Thr Pro
1685 1690 1695
Glu Pro Ile Ile Ile Glu Glu Glu Glu Glu Asp Ser Ile Ser Leu

1700 1705 1710

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

1715 1720 1725

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

1730 1735 1740

His Ala Ser Asp Phe Asp Val Asp Ser Leu Ser Ile Leu Asp Thr

1745 1750 1755

Leu Glu Gly Ala Ser Val Thr Ser Gly Ala Thr Ser Ala Glu Thr

1760 1765 1770

Asn Ser Tyr Phe Ala Lys Ser Met Glu Phe Leu Ala Arg Pro Val

1775 1780 1785

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

1790 1795 1800

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

1805 1810 1815

Thr Ser Leu Val Ser Thr Pro Pro Gly Val Asn Arg Val Ile Thr

1820 1825 1830

Arg Glu Glu Leu Glu Ala Leu Thr Pro Ser Arg Thr Pro Ser Arg

1835 1840 1845

Ser Val Ser Arg Thr Ser Leu Val Ser Asn Pro Pro Gly Val Asn

1850 1855 1860

Arg Val Ile Thr Arg Glu Glu Phe Glu Ala Phe Val Ala Gln Gln

1865 1870 1875

Gln Arg Phe Asp Ala Gly Ala Tyr Ile Phe Ser Ser Asp Thr Gly

1880 1885 1890

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

1895 1900 1905

Glu Val Val Leu Glu Arg Thr Glu Leu Glu Ile Ser Tyr Ala Pro

1910 1915 1920

Arg Leu Asp Gln Glu Lys Glu Glu Leu Leu Arg Lys Lys Leu Gln

1925 1930 1935

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

1940 1945 1950

Val Glu Asn Met Lys Ala Ile Thr Ala Arg Arg Ile Leu Gln Gly

1955 1960 1965

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

1970 1975 1980

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

1985 1990 1995

Phe Ser Ser Pro Lys Val Ala Val Glu Ala Cys Asn Ala Met Leu

2000 2005 2010

Lys Glu Asn Phe Pro Thr Val Ala Ser Tyr Cys Ile Ile Pro Glu

2015 2020 2025

Tyr Asp Ala Tyr Leu Asp Met Val Asp Gly Ala Ser Cys Cys Leu

2030 2035 2040

Asp Thr Ala Ser Phe Cys Pro Ala Lys Leu Arg Ser Phe Pro Lys

2045 2050 2055

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

2060 2065 2070

Ala Ile Gln Asn Thr Leu Gln Asn Val Leu Ala Ala Ala Thr Lys

2075 2080 2085

Arg Asn Cys Asn Val Thr Gln Met Arg Glu Leu Pro Val Leu Asp

2090 2095 2100

Ser Ala Ala Phe Asn Val Glu Cys Phe Lys Lys Tyr Ala Cys Asn

2105 2110 2115

Asn Glu Tyr Trp Glu Thr Phe Lys Glu Asn Pro Ile Arg Leu Thr

2120 2125 2130

Glu Glu Asn Val Val Asn Tyr Ile Thr Lys Leu Lys Gly Pro Lys

2135 2140 2145

Ala Ala Ala Leu Phe Ala Lys Thr His Asn Leu Asn Met Leu Gln

2150 2155 2160

Asp Ile Pro Met Asp Arg Phe Val Met Asp Leu Lys Arg Asp Val

2165 2170 2175

Lys Val Thr Pro Gly Thr Lys His Thr Glu Glu Arg Pro Lys Val

2180 2185 2190

Gln Val Ile Gln Ala Ala Asp Pro Leu Ala Thr Ala Tyr Leu Cys

2195 2200 2205

Gly Ile His Arg Glu Leu Val Arg Arg Leu Asn Ala Val Leu Leu

2210 2215 2220

Pro Asn Ile His Thr Leu Phe Asp Met Ser Ala Glu Asp Phe Asp

2225 2230 2235

Ala Ile Ile Ala Glu His Phe Gln Pro Gly Asp Cys Val Leu Glu

2240 2245 2250

Thr Asp Ile Ala Ser Phe Asp Lys Ser Glu Asp Asp Ala Met Ala

2255 2260 2265

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

2270 2275 2280

Leu Leu Thr Leu Ile Glu Ala Ala Phe Gly Glu Ile Ser Ser Ile

2285 2290 2295

His Leu Pro Thr Lys Thr Lys Phe Lys Phe Gly Ala Met Met Lys

2300 2305 2310

Ser Gly Met Phe Leu Thr Leu Phe Val Asn Thr Val Ile Asn Ile

2315 2320 2325

Val Ile Ala Ser Arg Val Leu Arg Glu Arg Leu Thr Gly Ser Pro

2330 2335 2340

Cys Ala Ala Phe Ile Gly Asp Asp Asn Ile Val Lys Gly Val Lys

2345 2350 2355

Ser Asp Lys Leu Met Ala Asp Arg Cys Ala Thr Trp Leu Asn Met

2360 2365 2370

Glu Val Lys Ile Ile Asp Ala Val Val Gly Glu Lys Ala Pro Tyr

2375 2380 2385

Phe Cys Gly Gly Phe Ile Leu Cys Asp Ser Val Thr Gly Thr Ala

2390 2395 2400

Cys Arg Val Ala Asp Pro Leu Lys Arg Leu Phe Lys Leu Gly Lys

2405 2410 2415

Pro Leu Ala Ala Asp Asp Glu His Asp Asp Asp Arg Arg Arg Ala

2420 2425 2430

Leu His Glu Glu Ser Thr Arg Trp Asn Arg Val Gly Ile Leu Ser

2435 2440 2445

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

2450 2455 2460

Ser Ile Ile Val Met Ala Met Thr Thr Leu Ala Ser Ser Val Lys

2465 2470 2475

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

2480 2485 2490

<210> 2

<211> 2493

<212> PRT

<213> Eastern equine encephalomyelitis virus

<400> 2

Met Glu Lys Val His Val Asp Leu Asp Ala Asp Ser Pro Phe Val Lys

1 5 10 15

Ser Leu Gln Arg Cys Phe Pro His Phe Glu Ile Glu Ala Thr Gln Val

20 25 30

Thr Asp Asn Asp His Ala Asn Ala Arg Ala Phe Ser His Leu Ala Thr

35 40 45

Lys Leu Ile Glu Gly Glu Val Asp Thr Asp Gln Val Ile Leu Asp Ile

50 55 60

Gly Ser Ala Pro Val Arg His Thr His Ser Lys His Lys Tyr His Cys

65 70 75 80

Ile Cys Pro Met Lys Ser Ala Glu Asp Pro Asp Arg Leu Tyr Arg Tyr

85 90 95

Ala Asp Lys Leu Arg Lys Ser Asp Val Thr Asp Lys Cys Ile Ala Ser

100 105 110

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

Asp Thr Thr Pro Phe Met Tyr Lys Asn Met Ala Gly Ala Tyr Pro Thr
180 185 190
Tyr Asn Thr Asn Trp Ala Asp Glu Ser Val Leu Glu Ala Arg Asn Ile
195 200 205
Gly Leu Gly Ser Ser Asp Leu His Glu Lys Ser Phe Gly Lys Val Ser
210 215 220
Ile Met Arg Lys Lys Lys Leu Gln Pro Thr Asn Lys Val Ile Phe Ser
225 230 235 240

Val Gly Ser Thr Ile Tyr Thr Glu Glu Arg Ile Leu Leu Arg Ser Trp
245 250 255
His Leu Pro Asn Val Phe His Leu Lys Gly Lys Thr Ser Phe Thr Gly
260 265 270
Arg Cys Asn Thr Ile Val Ser Cys Glu Gly Tyr Val Val Lys Lys Ile
275 280 285
Thr Leu Ser Pro Gly Ile Tyr Gly Lys Val Asp Asn Leu Ala Ser Thr
290 295 300

Met His Arg Glu Gly Phe Leu Ser Cys Lys Val Thr Asp Thr Leu Arg
305 310 315 320
Gly Glu Arg Val Ser Phe Pro Val Cys Thr Tyr Val Pro Ala Thr Leu
325 330 335
Cys Asp Gln Met Thr Gly Ile Leu Ala Thr Asp Val Ser Val Asp Asp
340 345 350
Ala Gln Lys Leu Leu Val Gly Leu Asn Gln Arg Ile Val Val Asn Gly

355 360 365

Arg Thr Gln Arg Asn Thr Asn Thr Met Gln Asn Tyr Leu Leu Pro Val
370 375 380
Val Ala Gln Ala Phe Ser Arg Trp Ala Arg Glu His Arg Ala Asp Leu
385 390 395 400
Glu Asp Glu Lys Gly Leu Gly Val Arg Glu Arg Ser Leu Val Met Gly
405 410 415
Cys Cys Trp Ala Phe Lys Thr His Lys Ile Thr Ser Ile Tyr Lys Arg
420 425 430

Pro Gly Thr Gln Thr Ile Lys Lys Val Pro Ala Val Phe Asn Ser Phe
435 440 445
Val Ile Pro Gln Pro Thr Ser Tyr Gly Leu Asp Ile Gly Leu Arg Arg
450 455 460
Arg Ile Lys Met Leu Phe Asp Ala Lys Lys Ala Pro Ala Pro Ile Ile
465 470 475 480
Thr Glu Ala Asp Val Ala His Leu Lys Gly Leu Gln Asp Glu Ala Glu
485 490 495

Ala Val Ala Glu Ala Glu Ala Val Arg Ala Ala Leu Pro Pro Leu Leu
500 505 510
Pro Glu Val Asp Lys Glu Thr Val Glu Ala Asp Ile Asp Leu Ile Met
515 520 525
Gln Glu Ala Gly Ala Gly Ser Val Glu Thr Pro Arg Arg His Ile Lys
530 535 540
Val Thr Thr Tyr Pro Gly Glu Glu Met Ile Gly Ser Tyr Ala Val Leu
545 550 555 560

Ser Pro Gln Ala Val Leu Asn Ser Glu Lys Leu Ala Cys Ile His Pro
565 570 575
Leu Ala Glu Gln Val Leu Val Met Thr His Lys Gly Arg Ala Gly Arg
580 585 590
Tyr Lys Val Glu Pro Tyr His Gly Arg Val Ile Val Pro Ser Gly Thr
595 600 605

Ala Ile Pro Ile Leu Asp Phe Gln Ala Leu Ser Glu Ser Ala Thr Ile
610 615 620

Val Phe Asn Glu Arg Glu Phe Val Asn Arg Tyr Leu His His Ile Ala
625 630 635 640
Val Asn Gly Gly Ala Leu Asn Thr Asp Glu Glu Tyr Tyr Lys Val Val
645 650 655
Lys Ser Thr Glu Thr Asp Ser Glu Tyr Val Phe Asp Ile Asp Ala Lys
660 665 670
Lys Cys Val Lys Lys Gly Asp Ala Gly Pro Met Cys Leu Val Gly Glu
675 680 685

Leu Val Asp Pro Pro Phe His Glu Phe Ala Tyr Glu Ser Leu Lys Thr
690 695 700
Arg Pro Ala Ala Pro His Lys Val Pro Thr Ile Gly Val Tyr Gly Val
705 710 715 720
Pro Gly Ser Gly Lys Ser Gly Ile Ile Lys Ser Ala Val Thr Lys Arg
725 730 735
Asp Leu Val Val Ser Ala Lys Lys Glu Asn Cys Met Glu Ile Ile Lys
740 745 750

Asp Val Lys Arg Met Arg Gly Met Asp Ile Ala Ala Arg Thr Val Asp
755 760 765
Ser Val Leu Leu Asn Gly Val Lys His Ser Val Asp Thr Leu Tyr Ile
770 775 780
Asp Glu Ala Phe Ala Cys His Ala Gly Thr Leu Leu Ala Leu Ile Ala
785 790 795 800
Ile Val Lys Pro Lys Lys Val Val Leu Cys Gly Asp Pro Lys Gln Cys
805 810 815

Gly Phe Phe Asn Met Met Cys Leu Lys Val His Phe Asn His Glu Ile
820 825 830
Cys Thr Glu Val Tyr His Lys Ser Ile Ser Arg Arg Cys Thr Lys Thr
835 840 845
Val Thr Ser Ile Val Ser Thr Leu Phe Tyr Asp Lys Arg Met Arg Thr

```
    850                 855                 860
Val Asn Pro Cys Asn Asp Lys Ile Ile Ile Asp Thr Thr Ser Thr Thr
865                 870                 875                 880

Lys Pro Leu Lys Asp Asp Ile Ile Leu Thr Cys Phe Arg Gly Trp Val
                885                 890                 895
Lys Gln Leu Gln Ile Asp Tyr Lys Asn His Glu Ile Met Thr Ala Ala
            900                 905                 910
Ala Ser Gln Gly Leu Thr Arg Lys Gly Val Tyr Ala Val Arg Tyr Lys
        915                 920                 925
Val Asn Glu Asn Pro Leu Tyr Ala Gln Thr Ser Glu His Val Asn Val
    930                 935                 940

Leu Leu Thr Arg Thr Glu Lys Arg Ile Val Trp Lys Thr Leu Ala Gly
945                 950                 955                 960
Asp Pro Trp Ile Lys Thr Leu Thr Ala Ser Tyr Pro Gly Asn Phe Thr
                965                 970                 975
Ala Thr Leu Glu Glu Trp Gln Ala Glu His Asp Ala Ile Met Ala Lys
            980                 985                 990
Ile Leu Glu Thr Pro Ala Ser Ser  Asp Val Phe Gln Asn  Lys Val Asn
        995                 1000                1005

Val Cys  Trp Ala Lys Ala Leu  Glu Pro Val Leu Ala  Thr Ala Asn
    1010                1015                1020
Ile Thr  Leu Thr Arg Ser Gln  Trp Glu Thr Ile Pro  Ala Phe Lys
    1025                1030                1035
Asp Asp  Lys Ala Tyr Ser Pro  Glu Met Ala Leu Asn  Phe Phe Cys
    1040                1045                1050
Thr Arg  Phe Phe Gly Val Asp  Ile Asp Ser Gly Leu  Phe Ser Ala
    1055                1060                1065

Pro Thr  Val Pro Leu Thr Tyr  Thr Asn Glu His Trp  Asp Asn Ser
    1070                1075                1080
Pro Gly  Pro Asn Met Tyr Gly  Leu Cys Met Arg Thr  Ala Lys Glu
    1085                1090                1095
```

Leu Ala Arg Arg Tyr Pro Cys Ile Leu Lys Ala Val Asp Thr Gly
1100 1105 1110
Arg Val Ala Asp Val Arg Thr Asp Thr Ile Lys Asp Tyr Asn Pro
1115 1120 1125

Leu Ile Asn Val Val Pro Leu Asn Arg Arg Leu Pro His Ser Leu
1130 1135 1140
Val Val Thr His Arg Tyr Thr Gly Asn Gly Asp Tyr Ser Gln Leu
1145 1150 1155
Val Thr Lys Met Thr Gly Lys Thr Val Leu Val Val Gly Thr Pro
1160 1165 1170
Met Asn Ile Pro Gly Lys Arg Val Glu Thr Leu Gly Pro Ser Pro
1175 1180 1185

Gln Cys Thr Tyr Lys Ala Glu Leu Asp Leu Gly Ile Pro Ala Ala
1190 1195 1200
Leu Gly Lys Tyr Asp Ile Ile Phe Ile Asn Val Arg Thr Pro Tyr
1205 1210 1215
Arg His His His Tyr Gln Gln Cys Glu Asp His Ala Ile His His
1220 1225 1230
Ser Met Leu Thr Arg Lys Ala Val Asp His Leu Asn Lys Gly Gly
1235 1240 1245

Thr Cys Ile Ala Leu Gly Tyr Gly Thr Ala Asp Arg Ala Thr Glu
1250 1255 1260
Asn Ile Ile Ser Ala Val Ala Arg Ser Phe Arg Phe Ser Arg Val
1265 1270 1275
Cys Gln Pro Lys Cys Ala Trp Glu Asn Thr Glu Val Ala Phe Val
1280 1285 1290
Phe Phe Gly Lys Asp Asn Gly Asn His Leu Gln Asp Gln Asp Arg
1295 1300 1305

Leu Ser Val Val Leu Asn Asn Ile Tyr Gln Gly Ser Thr Gln His
1310 1315 1320
Glu Ala Gly Arg Ala Pro Ala Tyr Arg Val Val Arg Gly Asp Ile

1325 1330 1335

Thr Lys Ser Asn Asp Glu Val Ile Val Asn Ala Ala Asn Asn Lys

1340 1345 1350

Gly Gln Pro Gly Ser Gly Val Cys Gly Ala Leu Tyr Arg Lys Trp

1355 1360 1365

Pro Gly Ala Phe Asp Lys Gln Pro Val Ala Thr Gly Lys Ala His

1370 1375 1380

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

1385 1390 1395

Phe Ser Arg Leu Ser Glu Asn Glu Gly Asp Gln Lys Leu Ser Glu

1400 1405 1410

Val Tyr Met Asp Ile Ala Arg Ile Ile Asn Asn Glu Arg Phe Thr

1415 1420 1425

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

1430 1435 1440

Lys Asp Arg Val Met Gln Ser Leu Asn His Leu Phe Thr Ala Met

1445 1450 1455

Asp Thr Thr Asp Ala Asp Ile Thr Ile Tyr Cys Leu Asp Lys Gln

1460 1465 1470

Trp Glu Ser Arg Ile Lys Glu Ala Ile Thr Arg Lys Glu Ser Val

1475 1480 1485

Glu Glu Leu Thr Glu Asp Asp Arg Pro Val Asp Ile Glu Leu Val

1490 1495 1500

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

1505 1510 1515

Thr Thr Glu Gly Lys Val Tyr Ser Tyr Leu Glu Gly Thr Arg Phe

1520 1525 1530

His Gln Thr Ala Lys Asp Ile Ala Glu Ile Tyr Ala Met Trp Pro

1535 1540 1545

Asn Lys Gln Glu Ala Asn Glu Gln Ile Cys Leu Tyr Val Leu Gly

1550 1555 1560

Glu Ser Met Asn Ser Ile Arg Ser Lys Cys Pro Val Glu Glu Ser
1565 1570 1575
Glu Ala Ser Ser Pro Pro His Thr Ile Pro Cys Leu Cys Asn Tyr
1580 1585 1590
Ala Met Thr Ala Glu Arg Val Tyr Arg Leu Arg Met Ala Lys Asn
1595 1600 1605

Glu Gln Phe Ala Val Cys Ser Ser Phe Gln Leu Pro Lys Tyr Arg
1610 1615 1620
Ile Thr Gly Val Gln Lys Ile Gln Cys Ser Lys Pro Val Ile Phe
1625 1630 1635
Ser Gly Thr Val Pro Pro Ala Ile His Pro Arg Lys Phe Ala Ser
1640 1645 1650
Val Thr Val Glu Asp Thr Pro Val Val Gln Pro Glu Arg Leu Val
1655 1660 1665

Pro Arg Arg Pro Ala Pro Pro Val Pro Val Pro Ala Arg Ile Pro
1670 1675 1680
Ser Pro Pro Cys Thr Ser Thr Asn Gly Ser Thr Thr Ser Ile Gln
1685 1690 1695
Ser Leu Gly Glu Asp Gln Ser Ala Ser Ala Ser Ser Gly Ala Glu
1700 1705 1710
Ile Ser Val Asp Gln Val Ser Leu Trp Ser Ile Pro Ser Ala Thr
1715 1720 1725

Gly Phe Asp Val Arg Thr Ser Ser Ser Leu Ser Leu Glu Gln Pro
1730 1735 1740
Thr Phe Pro Thr Met Val Val Glu Ala Glu Ile His Ala Ser Gln
1745 1750 1755
Gly Ser Leu Trp Ser Ile Pro Ser Ile Thr Gly Ser Glu Thr Arg
1760 1765 1770
Ala Pro Ser Pro Pro Ser Gln Asp Ser Arg Pro Ser Thr Pro Ser
1775 1780 1785

Ala Ser Gly Ser His Thr Ser Val Asp Leu Ile Thr Phe Asp Ser

1790 1795 1800

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

1805 1810 1815

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

1820 1825 1830

Asn Met Ser Arg Ser Ala Asp Thr Ile Lys Pro Ile Pro Lys Pro

1835 1840 1845

Arg Lys Cys Gln Val Lys Tyr Thr Gln Pro Pro Gly Val Ala Arg

1850 1855 1860

Val Ile Ser Ala Ala Glu Phe Asp Glu Phe Val Arg Arg His Ser

1865 1870 1875

Asn Arg Tyr Glu Ala Gly Ala Tyr Ile Phe Ser Ser Glu Thr Gly

1880 1885 1890

Gln Gly His Leu Gln Gln Lys Ser Thr Arg Gln Cys Lys Leu Gln

1895 1900 1905

Tyr Pro Ile Leu Glu Arg Ser Val His Glu Lys Phe Tyr Ala Pro

1910 1915 1920

Arg Leu Asp Leu Glu Arg Glu Lys Leu Leu Gln Lys Lys Leu Gln

1925 1930 1935

Leu Cys Ala Ser Glu Gly Asn Arg Ser Arg Tyr Gln Ser Arg Lys

1940 1945 1950

Val Glu Asn Met Lys Ala Ile Thr Val Glu Arg Leu Leu Gln Gly

1955 1960 1965

Ile Gly Ser Tyr Leu Ser Ala Glu Pro Gln Pro Val Glu Cys Tyr

1970 1975 1980

Lys Val Thr Tyr Pro Ala Pro Met Tyr Ser Ser Thr Ala Ser Asn

1985 1990 1995

Ser Phe Ser Ser Ala Glu Val Ala Val Lys Val Cys Asn Leu Val

2000 2005 2010

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

2015 2020 2025

Glu Tyr Asp Ala Tyr Leu Asp Met Val Asp Gly Ala Ser Cys Cys
2030 2035 2040
Leu Asp Thr Ala Thr Phe Cys Pro Ala Lys Leu Arg Ser Phe Pro
2045 2050 2055
Lys Lys His Ser Tyr Leu Arg Pro Glu Ile Arg Ser Ala Val Pro
2060 2065 2070
Ser Pro Ile Gln Asn Thr Leu Gln Asn Val Leu Ala Ala Ala Thr
2075 2080 2085

Lys Arg Asn Cys Asn Val Thr Gln Met Arg Glu Leu Pro Val Leu
2090 2095 2100
Asp Ser Ala Ala Phe Asn Val Glu Cys Phe Lys Lys Tyr Ala Cys
2105 2110 2115
Asn Asp Glu Tyr Trp Asp Phe Tyr Lys Thr Asn Pro Ile Arg Leu
2120 2125 2130
Thr Ala Glu Asn Val Thr Gln Tyr Val Thr Lys Leu Lys Gly Pro
2135 2140 2145

Lys Ala Ala Ala Leu Phe Ala Lys Thr His Asn Leu Gln Pro Leu
2150 2155 2160
His Glu Ile Pro Met Asp Arg Phe Val Met Asp Leu Lys Arg Asp
2165 2170 2175
Val Lys Val Thr Pro Gly Thr Lys His Thr Glu Glu Arg Pro Lys
2180 2185 2190
Val Gln Val Ile Gln Ala Ala Asp Pro Leu Ala Thr Ala Tyr Leu
2195 2200 2205

Cys Gly Ile His Arg Glu Leu Val Arg Arg Leu Asn Ala Val Leu
2210 2215 2220
Leu Pro Asn Ile His Thr Leu Phe Asp Met Ser Ala Glu Asp Phe
2225 2230 2235
Asp Ala Ile Ile Ala Glu His Phe Gln Phe Gly Asp Ala Val Leu
2240 2245 2250
Glu Thr Asp Ile Ala Ser Phe Asp Lys Ser Glu Asp Asp Ala Ile

2255 2260 2265

Ala Met Ser Ala Leu Met Ile Leu Glu Asp Leu Gly Val Asp Gln
2270 2275 2280
Ala Leu Leu Asn Leu Ile Glu Ala Ala Phe Gly Asn Ile Thr Ser
2285 2290 2295
Val His Leu Pro Thr Gly Thr Arg Phe Lys Phe Gly Ala Met Met
2300 2305 2310
Lys Ser Gly Met Phe Leu Thr Leu Phe Ile Asn Thr Val Val Asn
2315 2320 2325

Ile Met Ile Ala Ser Arg Val Leu Arg Glu Arg Leu Thr Thr Ser
2330 2335 2340
Pro Cys Ala Ala Phe Ile Gly Asp Asp Asn Ile Val Lys Gly Val
2345 2350 2355
Thr Ser Asp Ala Leu Met Ala Glu Arg Cys Ala Thr Trp Leu Asn
2360 2365 2370
Met Glu Val Lys Ile Ile Asp Ala Val Val Gly Val Lys Ala Pro
2375 2380 2385

Tyr Phe Cys Gly Gly Phe Ile Val Val Asp Gln Ile Thr Gly Thr
2390 2395 2400
Ala Cys Arg Val Ala Asp Pro Leu Lys Arg Leu Phe Lys Leu Gly
2405 2410 2415
Lys Pro Leu Pro Leu Asp Asp Asp Gln Asp Val Asp Arg Arg Arg
2420 2425 2430
Ala Leu His Asp Glu Ala Ala Arg Trp Asn Arg Ile Gly Ile Thr
2435 2440 2445

Glu Glu Leu Val Lys Ala Val Glu Ser Arg Tyr Glu Val Asn Tyr
2450 2455 2460
Val Ser Leu Ile Ile Thr Ala Leu Thr Thr Leu Ala Ser Ser Val
2465 2470 2475
Ser Asn Phe Lys His Ile Arg Gly His Pro Ile Thr Leu Tyr Gly
2480 2485 2490

<210> 3

<211> 2431

<212> PRT

<213> Semliki Forest virus

<400> 3

Met Ala Ala Lys Val His Val Asp Ile Glu Ala Asp Ser Pro Phe Ile

1 5 10 15

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

20 25 30

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

35 40 45

Thr Lys Leu Ile Glu Gln Glu Thr Asp Lys Asp Thr Leu Ile Leu Asp

50 55 60

Ile Gly Ser Ala Pro Ser Arg Arg Met Met Ser Thr His Lys Tyr His

65 70 75 80

Cys Val Cys Pro Met Arg Ser Ala Glu Asp Pro Glu Arg Leu Asp Ser

85 90 95

Tyr Ala Lys Lys Leu Ala Ala Ala Ser Gly Lys Val Leu Asp Arg Glu

100 105 110

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

115 120 125

Ala Glu Ser Pro Thr Phe Cys Leu His Thr Asp Val Thr Cys Arg Thr

130 135 140

Ala Ala Glu Val Ala Val Tyr Gln Asp Val Tyr Ala Val His Ala Pro

145 150 155 160

Thr Ser Leu Tyr His Gln Ala Met Lys Gly Val Arg Thr Ala Tyr Trp

165 170 175

Ile Gly Phe Asp Thr Thr Pro Phe Met Phe Asp Ala Leu Ala Gly Ala

180 185 190

Tyr Pro Thr Tyr Ala Thr Asn Trp Ala Asp Glu Gln Val Leu Gln Ala

195 200 205

Arg Asn Ile Gly Leu Cys Ala Ala Ser Leu Thr Glu Gly Arg Leu Gly
210 215 220
Lys Leu Ser Ile Leu Arg Lys Lys Gln Leu Lys Pro Cys Asp Thr Val
225 230 235 240
Met Phe Ser Val Gly Ser Thr Leu Tyr Thr Glu Ser Arg Lys Leu Leu
245 250 255
Arg Ser Trp His Leu Pro Ser Val Phe His Leu Lys Gly Lys Gln Ser

260 265 270
Phe Thr Cys Arg Cys Asp Thr Ile Val Ser Cys Glu Gly Tyr Val Val
275 280 285
Lys Lys Ile Thr Met Cys Pro Gly Leu Tyr Gly Lys Thr Val Gly Tyr
290 295 300
Ala Val Thr Tyr His Ala Glu Gly Phe Leu Val Cys Lys Thr Thr Asp
305 310 315 320
Thr Val Lys Gly Glu Arg Val Ser Phe Pro Val Cys Thr Tyr Val Pro

325 330 335
Ser Thr Ile Cys Asp Gln Met Thr Gly Ile Leu Ala Thr Asp Val Thr
340 345 350
Pro Glu Asp Ala Gln Lys Leu Leu Val Gly Leu Asn Gln Arg Ile Val
355 360 365
Val Asn Gly Arg Thr Gln Arg Asn Thr Asn Thr Met Lys Asn Tyr Leu
370 375 380
Leu Pro Ile Val Ala Val Ala Phe Ser Lys Trp Ala Arg Glu Tyr Lys

385 390 395 400
Ala Asp Leu Asp Asp Glu Lys Pro Leu Gly Val Arg Glu Arg Ser Leu
405 410 415
Thr Cys Cys Cys Leu Trp Ala Phe Lys Thr Arg Lys Met His Thr Met
420 425 430
Tyr Lys Lys Pro Asp Thr Gln Thr Ile Val Lys Val Pro Ser Glu Phe
435 440 445
Asn Ser Phe Val Ile Pro Ser Leu Trp Ser Thr Gly Leu Ala Ile Pro

450 455 460

Val Arg Ser Arg Ile Lys Met Leu Leu Ala Lys Lys Thr Lys Arg Glu

465 470 475 480

Leu Ile Pro Val Leu Asp Ala Ser Ser Ala Arg Asp Ala Glu Gln Glu

485 490 495

Glu Lys Glu Arg Leu Glu Ala Glu Leu Thr Arg Glu Ala Leu Pro Pro

500 505 510

Leu Val Pro Ile Ala Pro Ala Glu Thr Gly Val Val Asp Val Asp Val

515 520 525

Glu Glu Leu Glu Tyr His Ala Gly Ala Gly Val Val Glu Thr Pro Arg

530 535 540

Ser Ala Leu Lys Val Thr Ala Gln Pro Asn Asp Val Leu Leu Gly Asn

545 550 555 560

Tyr Val Val Leu Ser Pro Gln Thr Val Leu Lys Ser Ser Lys Leu Ala

565 570 575

Pro Val His Pro Leu Ala Glu Gln Val Lys Ile Ile Thr His Asn Gly

580 585 590

Arg Ala Gly Gly Tyr Gln Val Asp Gly Tyr Asp Gly Arg Val Leu Leu

595 600 605

Pro Cys Gly Ser Ala Ile Pro Val Pro Glu Phe Gln Ala Leu Ser Glu

610 615 620

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

625 630 635 640

Tyr His Ile Ala Val His Gly Pro Ser Leu Asn Thr Asp Glu Glu Asn

645 650 655

Tyr Glu Lys Val Arg Ala Glu Arg Thr Asp Ala Glu Tyr Val Phe Asp

660 665 670

Val Asp Lys Lys Cys Cys Val Lys Arg Glu Glu Ala Ser Gly Leu Val

675 680 685

Leu Val Gly Glu Leu Thr Asn Pro Pro Phe His Glu Phe Ala Tyr Glu

690 695 700

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

705 710 715 720

Val Phe Gly Val Pro Gly Ser Gly Lys Ser Ala Ile Ile Lys Ser Leu

725 730 735

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

740 745 750

Glu Ile Val Asn Asp Val Lys Lys His Arg Gly Lys Gly Thr Ser Arg

755 760 765

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

770 775 780

Ile Leu Tyr Val Asp Glu Ala Phe Ala Cys His Ser Gly Thr Leu Leu

785 790 795 800

Ala Leu Ile Ala Leu Val Lys Pro Arg Ser Lys Val Val Leu Cys Gly

805 810 815

Asp Pro Lys Gln Cys Gly Phe Phe Asn Met Met Gln Leu Lys Val Asn

820 825 830

Phe Asn His Asn Ile Cys Thr Glu Val Cys His Lys Ser Ile Ser Arg

835 840 845

Arg Cys Thr Arg Pro Val Thr Ala Ile Val Ser Thr Leu His Tyr Gly

850 855 860

Gly Lys Met Arg Thr Thr Asn Pro Cys Asn Lys Pro Ile Ile Ile Asp

865 870 875 880

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

885 890 895

Phe Arg Gly Trp Ala Lys Gln Leu Gln Leu Asp Tyr Arg Gly His Glu

900 905 910

Val Met Thr Ala Ala Ala Ser Gln Gly Leu Thr Arg Lys Gly Val Tyr

915 920 925

Ala Val Arg Gln Lys Val Asn Glu Asn Pro Leu Tyr Ala Pro Ala Ser

930 935 940

Glu His Val Asn Val Leu Leu Thr Arg Thr Glu Asp Arg Leu Val Trp

945 950 955 960

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

965 970 975

Gln Gly Asn Phe Thr Ala Thr Leu Glu Glu Trp Gln Glu Glu His Asp

980 985 990

Lys Ile Met Lys Val Ile Glu Gly Pro Ala Ala Pro Val Asp Ala Phe

995 1000 1005

Gln Asn Lys Ala Asn Val Cys Trp Ala Lys Ser Leu Val Pro Val

1010 1015 1020

Leu Asp Thr Ala Gly Ile Arg Leu Thr Ala Glu Glu Trp Ser Thr

1025 1030 1035

Ile Ile Thr Ala Phe Lys Glu Asp Arg Ala Tyr Ser Pro Val Val

1040 1045 1050

Ala Leu Asn Glu Ile Cys Thr Lys Tyr Tyr Gly Val Asp Leu Asp

1055 1060 1065

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

1070 1075 1080

Asn His Trp Asp Asn Arg Pro Gly Gly Arg Met Tyr Gly Phe Asn

1085 1090 1095

Ala Ala Thr Ala Ala Arg Leu Glu Ala Arg His Thr Phe Leu Lys

1100 1105 1110

Gly Gln Trp His Thr Gly Lys Gln Ala Val Ile Ala Glu Arg Lys

1115 1120 1125

Ile Gln Pro Leu Ser Val Leu Asp Asn Val Ile Pro Ile Asn Arg

1130 1135 1140

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

1145 1150 1155

Ser Arg Val Glu Trp Leu Val Asn Lys Val Arg Gly Tyr His Val

1160 1165 1170

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

1175 1180 1185

```
Thr Trp  Leu Ser Pro Leu Asn  Val Thr Gly Ala Asp  Arg Cys Tyr
    1190                 1195                 1200
Asp Leu  Ser Leu Gly Leu Pro  Ala Asp Ala Gly Arg  Phe Asp Leu

    1205                 1210                 1215
Val Phe  Val Asn Ile His Thr  Glu Phe Arg Ile His  His Tyr Gln
    1220                 1225                 1230
Gln Cys  Val Asp His Ala Met  Lys Leu Gln Met Leu  Gly Gly Asp
    1235                 1240                 1245
Ala Leu  Arg Leu Leu Lys Pro  Gly Gly Ile Leu Met  Arg Ala Tyr
    1250                 1255                 1260
Gly Tyr  Ala Asp Lys Ile Ser  Glu Ala Val Val Ser  Ser Leu Ser

    1265                 1270                 1275
Arg Lys  Phe Ser Ser Ala Arg  Val Leu Arg Pro Asp  Cys Val Thr
    1280                 1285                 1290
Ser Asn  Thr Glu Val Phe Leu  Leu Phe Ser Asn Phe  Asp Asn Gly
    1295                 1300                 1305
Lys Arg  Pro Ser Thr Leu His  Gln Met Asn Thr Lys  Leu Ser Ala
    1310                 1315                 1320
Val Tyr  Ala Gly Glu Ala Met  His Thr Ala Gly Cys  Ala Pro Ser

    1325                 1330                 1335
Tyr Arg  Val Lys Arg Ala Asp  Ile Ala Thr Cys Thr  Glu Ala Ala
    1340                 1345                 1350
Val Val  Asn Ala Ala Asn Ala  Arg Gly Thr Val Gly  Asp Gly Val
    1355                 1360                 1365
Cys Arg  Ala Val Ala Lys Lys  Trp Pro Ser Ala Phe  Lys Gly Ala
    1370                 1375                 1380
Ala Thr  Pro Val Gly Thr Ile  Lys Thr Val Met Cys  Gly Ser Tyr

    1385                 1390                 1395
Pro Val  Ile His Ala Val Ala  Pro Asn Phe Ser Ala  Thr Thr Glu
    1400                 1405                 1410
Ala Glu  Gly Asp Arg Glu Leu  Ala Ala Val Tyr Arg  Ala Val Ala
```

1415 1420 1425

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

1430 1435 1440

Leu Ser Thr Gly Val Phe Ser Gly Gly Arg Asp Arg Leu Gln Gln

1445 1450 1455

Ser Leu Asn His Leu Phe Thr Ala Met Asp Ala Thr Asp Ala Asp

1460 1465 1470

Val Thr Ile Tyr Cys Arg Asp Lys Ser Trp Glu Lys Lys Ile Gln

1475 1480 1485

Glu Ala Ile Asp Met Arg Thr Ala Val Glu Leu Leu Asn Asp Asp

1490 1495 1500

Val Glu Leu Thr Thr Asp Leu Val Arg Val His Pro Asp Ser Ser

1505 1510 1515

Leu Val Gly Arg Lys Gly Tyr Ser Thr Thr Asp Gly Ser Leu Tyr

1520 1525 1530

Ser Tyr Phe Glu Gly Thr Lys Phe Asn Gln Ala Ala Ile Asp Met

1535 1540 1545

Ala Glu Ile Leu Thr Leu Trp Pro Arg Leu Gln Glu Ala Asn Glu

1550 1555 1560

Arg Ile Cys Leu Tyr Ala Leu Gly Glu Thr Met Asp Asn Ile Gly

1565 1570 1575

Ser Lys Cys Pro Val Asn Asp Ser Asp Ser Ser Thr Pro Pro Arg

1580 1585 1590

Thr Val Pro Cys Leu Cys Arg Tyr Ala Met Thr Ala Glu Arg Ile

1595 1600 1605

Ala Arg Leu Arg Ser His Gln Val Lys Ser Met Val Val Cys Ser

1610 1615 1620

Ser Phe Pro Leu Pro Lys Tyr His Val Asp Gly Val Gln Lys Val

1625 1630 1635

Lys Cys Glu Lys Val Leu Leu Phe Asp Pro Thr Val Pro Ser Val

1640 1645 1650

Val Ser Pro Arg Lys Tyr Ala Ala Ser Thr Thr Asp His Ser Asp

1655 1660 1665

Arg Ser Leu Arg Gly Phe Asp Leu Asp Trp Thr Thr Asp Ser Ser

1670 1675 1680

Ser Thr Ala Ser Asp Thr Met Ser Leu Pro Ser Leu Gln Ser Cys

1685 1690 1695

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

1700 1705 1710

Ala Asp Val His Pro Glu Pro Ala Gly Ile Ala Asp Leu Ala Ala

1715 1720 1725

Asp Val His Pro Glu Pro Ala Asp His Val Asp Leu Glu Asn Pro

1730 1735 1740

Ile Pro Pro Pro Arg Pro Lys Arg Ala Ala Tyr Leu Ala Ser Arg

1745 1750 1755

Ala Ala Glu Arg Pro Val Pro Ala Pro Arg Lys Pro Thr Pro Ala

1760 1765 1770

Pro Arg Thr Ala Phe Arg Asn Lys Leu Pro Leu Thr Phe Gly Asp

1775 1780 1785

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

1790 1795 1800

Gly Asp Phe Asp Asp Val Leu Arg Leu Gly Arg Ala Gly Ala Tyr

1805 1810 1815

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

1820 1825 1830

Val Arg Gln His Asn Leu Gln Cys Ala Gln Leu Asp Ala Val Gln

1835 1840 1845

Glu Glu Lys Met Tyr Pro Pro Lys Leu Asp Thr Glu Arg Glu Lys

1850 1855 1860

Leu Leu Leu Leu Lys Met Gln Met His Pro Ser Glu Ala Asn Lys

1865 1870 1875

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

1880 1885 1890

Val Asp Arg Leu Thr Ser Gly Ala Arg Leu Tyr Thr Gly Ala Asp

1895 1900 1905

Val Gly Arg Ile Pro Thr Tyr Ala Val Arg Tyr Pro Arg Pro Val

1910 1915 1920

Tyr Ser Pro Thr Val Ile Glu Arg Phe Ser Ser Pro Asp Val Ala

1925 1930 1935

Ile Ala Ala Cys Asn Glu Tyr Leu Ser Arg Asn Tyr Pro Thr Val

1940 1945 1950

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

1955 1960 1965

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

1970 1975 1980

Ala Lys Leu Arg Cys Tyr Pro Lys His His Ala Tyr His Gln Pro

1985 1990 1995

Thr Val Arg Ser Ala Val Pro Ser Pro Phe Gln Asn Thr Leu Gln

2000 2005 2010

Asn Val Leu Ala Ala Ala Thr Lys Arg Asn Cys Asn Val Thr Gln

2015 2020 2025

Met Arg Glu Leu Pro Thr Met Asp Ser Ala Val Phe Asn Val Glu

2030 2035 2040

Cys Phe Lys Arg Tyr Ala Cys Ser Gly Glu Tyr Trp Glu Glu Tyr

2045 2050 2055

Ala Lys Gln Pro Ile Arg Ile Thr Thr Glu Asn Ile Thr Thr Tyr

2060 2065 2070

Val Thr Lys Leu Lys Gly Pro Lys Ala Ala Ala Leu Phe Ala Lys

2075 2080 2085

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

2090 2095 2100

Thr Val Asp Met Lys Arg Asp Val Lys Val Thr Pro Gly Thr Lys

2105 2110 2115

His Thr Glu Glu Arg Pro Lys Val Gln Val Ile Gln Ala Ala Glu

2120 2125 2130

Pro Leu Ala Thr Ala Tyr Leu Cys Gly Ile His Arg Glu Leu Val

2135 2140 2145

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

2150 2155 2160

Asp Met Ser Ala Glu Asp Phe Asp Ala Ile Ile Ala Ser His Phe

2165 2170 2175

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

2180 2185 2190

Lys Ser Gln Asp Asp Ser Leu Ala Leu Thr Gly Leu Met Ile Leu

2195 2200 2205

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

2210 2215 2220

Ala Phe Gly Glu Ile Ser Ser Cys His Leu Pro Thr Gly Thr Arg

2225 2230 2235

Phe Lys Phe Gly Ala Met Met Lys Ser Gly Met Phe Leu Thr Leu

2240 2245 2250

Phe Ile Asn Thr Val Leu Asn Ile Thr Ile Ala Ser Arg Val Leu

2255 2260 2265

Glu Gln Arg Leu Thr Asp Ser Ala Cys Ala Ala Phe Ile Gly Asp

2270 2275 2280

Asp Asn Ile Val His Gly Val Ile Ser Asp Lys Leu Met Ala Glu

2285 2290 2295

Arg Cys Ala Ser Trp Val Asn Met Glu Val Lys Ile Ile Asp Ala

2300 2305 2310

Val Met Gly Glu Lys Pro Pro Tyr Phe Cys Gly Gly Phe Ile Val

2315 2320 2325

Phe Asp Ser Val Thr Gln Thr Ala Cys Arg Val Ser Asp Pro Leu

2330 2335 2340

Lys Arg Leu Phe Lys Leu Gly Lys Pro Leu Thr Ala Glu Asp Lys

2345 2350 2355

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

2360 2365 2370

Trp Phe Arg Thr Gly Leu Gly Ala Glu Leu Glu Val Ala Leu Thr

2375 2380 2385

Ser Arg Tyr Glu Val Glu Gly Cys Lys Ser Ile Leu Ile Ala Met

2390 2395 2400

Thr Thr Leu Ala Arg Asp Ile Lys Ala Phe Lys Lys Leu Arg Gly

2405 2410 2415

Pro Val Ile His Leu Tyr Gly Gly Pro Arg Leu Val Arg

2420 2425 2430

<210> 4

<211> 2512

<212> PRT

<213> Sindbis virus

<400> 4

Met Glu Lys Pro Val Val Asn Val Asp Val Asp Pro Gln Ser Pro Phe

1 5 10 15

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

20 25 30

Gln Val Thr Pro Asn Asp His Ala Asn Ala Arg Ala Phe Ser His Leu

35 40 45

Ala Ser Lys Leu Ile Glu Leu Glu Val Pro Thr Thr Ala Thr Ile Leu

50 55 60

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

65 70 75 80

His Cys Val Cys Pro Met Arg Ser Pro Glu Asp Pro Asp Arg Met Met

85 90 95

Lys Tyr Ala Ser Lys Leu Ala Glu Lys Ala Cys Lys Ile Thr Asn Lys

100 105 110

Asn Leu His Glu Lys Ile Lys Asp Leu Arg Thr Val Leu Asp Thr Pro

115 120 125

```
Asp Ala Glu Thr Pro Ser Leu Cys Phe His Asn Asp Val Thr Cys Asn
    130                 135                 140
Met Arg Ala Glu Tyr Ser Val Met Gln Asp Val Tyr Ile Asn Ala Pro
145                 150                 155                 160

Gly Thr Ile Tyr His Gln Ala Met Lys Gly Val Arg Thr Leu Tyr Trp
                165                 170                 175
Ile Gly Phe Asp Thr Thr Gln Phe Met Phe Ser Ala Met Ala Gly Ser
            180                 185                 190
Tyr Pro Ala Tyr Asn Thr Asn Trp Ala Asp Glu Lys Val Leu Glu Ala
        195                 200                 205
Arg Asn Ile Gly Leu Cys Ser Thr Lys Leu Ser Glu Gly Arg Thr Gly
    210                 215                 220

Lys Leu Ser Ile Met Arg Lys Lys Glu Leu Lys Pro Gly Ser Arg Val
225                 230                 235                 240
Tyr Phe Ser Val Gly Ser Thr Leu Tyr Pro Glu His Arg Ala Ser Leu
                245                 250                 255
Gln Ser Trp His Leu Pro Ser Val Phe His Leu Asn Gly Lys Gln Ser
            260                 265                 270
Tyr Thr Cys Arg Cys Asp Thr Val Val Ser Cys Glu Gly Tyr Val Val
        275                 280                 285

Lys Lys Ile Thr Ile Ser Pro Gly Ile Thr Gly Glu Thr Val Gly Tyr
    290                 295                 300
Ala Val Thr His Asn Ser Glu Gly Phe Leu Leu Cys Lys Val Thr Asp
305                 310                 315                 320
Thr Val Lys Gly Glu Arg Val Ser Phe Pro Val Cys Thr Tyr Ile Pro
                325                 330                 335
Ala Thr Ile Cys Asp Gln Met Thr Gly Ile Met Ala Thr Asp Ile Ser
            340                 345                 350

Pro Asp Asp Ala Gln Lys Leu Leu Val Gly Leu Asn Gln Arg Ile Val
        355                 360                 365
Ile Asn Gly Arg Thr Asn Arg Asn Thr Asn Thr Met Gln Asn Tyr Leu
```

370 375 380

Leu Pro Ile Ile Ala Gln Gly Phe Ser Lys Trp Ala Lys Glu Arg Lys
385 390 395 400

Asp Asp Leu Asp Asn Glu Lys Met Leu Gly Thr Arg Glu Arg Lys Leu
405 410 415

Thr Tyr Gly Cys Leu Trp Ala Phe Arg Thr Lys Lys Val His Ser Phe
420 425 430

Tyr Arg Pro Pro Gly Thr Gln Thr Cys Val Lys Val Pro Ala Ser Phe
435 440 445

Ser Ala Phe Pro Met Ser Ser Val Trp Thr Thr Ser Leu Pro Met Ser
450 455 460

Leu Arg Gln Lys Leu Lys Leu Ala Leu Gln Pro Lys Lys Glu Glu Lys
465 470 475 480

Leu Leu Gln Val Ser Glu Glu Leu Val Met Glu Ala Lys Ala Ala Phe
485 490 495

Glu Asp Ala Gln Glu Glu Ala Arg Ala Glu Lys Leu Arg Glu Ala Leu
500 505 510

Pro Pro Leu Val Ala Asp Lys Gly Ile Glu Ala Ala Ala Glu Val Val
515 520 525

Cys Glu Val Glu Gly Leu Gln Ala Asp Ile Gly Ala Ala Leu Val Glu
530 535 540

Thr Pro Arg Gly His Val Arg Ile Ile Pro Gln Ala Asn Asp Arg Met
545 550 555 560

Ile Gly Gln Tyr Ile Val Val Ser Pro Asn Ser Val Leu Lys Asn Ala
565 570 575

Lys Leu Ala Pro Ala His Pro Leu Ala Asp Gln Val Lys Ile Ile Thr
580 585 590

His Ser Gly Arg Ser Gly Arg Tyr Ala Val Glu Pro Tyr Asp Ala Lys
595 600 605

Val Leu Met Pro Ala Gly Gly Ala Val Pro Trp Pro Glu Phe Leu Ala
610 615 620

Leu Ser Glu Ser Ala Thr Leu Val Tyr Asn Glu Arg Glu Phe Val Asn
625 630 635 640
Arg Lys Leu Tyr His Ile Ala Met His Gly Pro Ala Lys Asn Thr Glu
645 650 655
Glu Glu Gln Tyr Lys Val Thr Lys Ala Glu Leu Ala Glu Thr Glu Tyr
660 665 670

Val Phe Asp Val Asp Lys Lys Arg Cys Val Lys Lys Glu Glu Ala Ser
675 680 685
Gly Leu Val Leu Ser Gly Glu Leu Thr Asn Pro Pro Tyr His Glu Leu
690 695 700
Ala Leu Glu Gly Leu Lys Thr Arg Pro Ala Val Pro Tyr Lys Val Glu
705 710 715 720
Thr Ile Gly Val Ile Gly Thr Pro Gly Ser Gly Lys Ser Ala Ile Ile
725 730 735

Lys Ser Thr Val Thr Ala Arg Asp Leu Val Thr Ser Gly Lys Lys Glu
740 745 750
Asn Cys Arg Glu Ile Glu Ala Asp Val Leu Arg Leu Arg Gly Met Gln
755 760 765
Ile Thr Ser Lys Thr Val Asp Ser Val Met Leu Asn Gly Cys His Lys
770 775 780
Ala Val Glu Val Leu Tyr Val Asp Glu Ala Phe Ala Cys His Ala Gly
785 790 795 800

Ala Leu Leu Ala Leu Ile Ala Ile Val Arg Pro Arg Lys Lys Val Val
805 810 815
Leu Cys Gly Asp Pro Met Gln Cys Gly Phe Phe Asn Met Met Gln Leu
820 825 830
Lys Val His Phe Asn His Pro Glu Lys Asp Ile Cys Thr Lys Thr Phe
835 840 845
Tyr Lys Tyr Ile Ser Arg Arg Cys Thr Gln Pro Val Thr Ala Ile Val
850 855 860

Ser Thr Leu His Tyr Asp Gly Lys Met Lys Thr Thr Asn Pro Cys Lys

865 870 875 880

Lys Asn Ile Glu Ile Asp Ile Thr Gly Ala Thr Lys Pro Lys Pro Gly

885 890 895

Asp Ile Ile Leu Thr Cys Phe Arg Gly Trp Val Lys Gln Leu Gln Ile

900 905 910

Asp Tyr Pro Gly His Glu Val Met Thr Ala Ala Ala Ser Gln Gly Leu

915 920 925

Thr Arg Lys Gly Val Tyr Ala Val Arg Gln Lys Val Asn Glu Asn Pro

930 935 940

Leu Tyr Ala Ile Thr Ser Glu His Val Asn Val Leu Leu Thr Arg Thr

945 950 955 960

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

965 970 975

Gln Pro Thr Asn Ile Pro Lys Gly Asn Phe Gln Ala Thr Ile Glu Asp

980 985 990

Trp Glu Ala Glu His Lys Gly Ile Ile Ala Ala Ile Asn Ser Pro Thr

995 1000 1005

Pro Arg Ala Asn Pro Phe Ser Cys Lys Thr Asn Val Cys Trp Ala

1010 1015 1020

Lys Ala Leu Glu Pro Ile Leu Ala Thr Ala Gly Ile Val Leu Thr

1025 1030 1035

Gly Cys Gln Trp Ser Glu Leu Phe Pro Gln Phe Ala Asp Asp Lys

1040 1045 1050

Pro His Ser Ala Ile Tyr Ala Leu Asp Val Ile Cys Ile Lys Phe

1055 1060 1065

Phe Gly Met Asp Leu Thr Ser Gly Leu Phe Ser Lys Gln Ser Ile

1070 1075 1080

Pro Leu Thr Tyr His Pro Ala Asp Ser Ala Arg Pro Val Ala His

1085 1090 1095

Trp Asp Asn Ser Pro Gly Thr Arg Lys Tyr Gly Tyr Asp His Ala

1100 1105 1110

Ile Ala Ala Glu Leu Ser Arg Arg Phe Pro Val Phe Gln Leu Ala
1115 1120 1125
Gly Lys Gly Thr Gln Leu Asp Leu Gln Thr Gly Arg Thr Arg Val
1130 1135 1140
Ile Ser Ala Gln His Asn Leu Val Pro Val Asn Arg Asn Leu Pro
1145 1150 1155
His Ala Leu Val Pro Glu Tyr Lys Glu Lys Gln Pro Gly Pro Val
1160 1165 1170

Lys Lys Phe Leu Asn Gln Phe Lys His His Ser Val Leu Val Val
1175 1180 1185
Ser Glu Glu Lys Ile Glu Ala Pro Arg Lys Arg Ile Glu Trp Ile
1190 1195 1200
Ala Pro Ile Gly Ile Ala Gly Ala Asp Lys Asn Tyr Asn Leu Ala
1205 1210 1215
Phe Gly Phe Pro Pro Gln Ala Arg Tyr Asp Leu Val Phe Ile Asn
1220 1225 1230

Ile Gly Thr Lys Tyr Arg Asn His His Phe Gln Gln Cys Glu Asp
1235 1240 1245
His Ala Ala Thr Leu Lys Thr Leu Ser Arg Ser Ala Leu Asn Cys
1250 1255 1260
Leu Asn Pro Gly Gly Thr Leu Val Val Lys Ser Tyr Gly Tyr Ala
1265 1270 1275
Asp Arg Asn Ser Glu Asp Val Val Thr Ala Leu Ala Arg Lys Phe
1280 1285 1290

Val Arg Val Ser Ala Ala Arg Pro Asp Cys Val Ser Ser Asn Thr
1295 1300 1305
Glu Met Tyr Leu Ile Phe Arg Gln Leu Asp Asn Ser Arg Thr Arg
1310 1315 1320
Gln Phe Thr Pro His His Leu Asn Cys Val Ile Ser Ser Val Tyr
1325 1330 1335
Glu Gly Thr Arg Asp Gly Val Gly Ala Ala Pro Ser Tyr Arg Thr

1340 1345 1350

Lys Arg Glu Asn Ile Ala Asp Cys Gln Glu Glu Ala Val Val Asn
1355 1360 1365
Ala Ala Asn Pro Leu Gly Arg Pro Gly Glu Gly Val Cys Arg Ala
1370 1375 1380
Ile Tyr Lys Arg Trp Pro Thr Ser Phe Thr Asp Ser Ala Thr Glu
1385 1390 1395
Thr Gly Thr Ala Arg Met Thr Val Cys Leu Gly Lys Lys Val Ile
1400 1405 1410

His Ala Val Gly Pro Asp Phe Arg Lys His Pro Glu Ala Glu Ala
1415 1420 1425
Leu Lys Leu Leu Gln Asn Ala Tyr His Ala Val Ala Asp Leu Val
1430 1435 1440
Asn Glu His Asn Ile Lys Ser Val Ala Ile Pro Leu Leu Ser Thr
1445 1450 1455
Gly Ile Tyr Ala Ala Gly Lys Asp Arg Leu Glu Val Ser Leu Asn
1460 1465 1470

Cys Leu Thr Thr Ala Leu Asp Arg Thr Asp Ala Asp Val Thr Ile
1475 1480 1485
Tyr Cys Leu Asp Lys Lys Trp Lys Glu Arg Ile Asp Ala Ala Leu
1490 1495 1500
Gln Leu Lys Glu Ser Val Thr Glu Leu Lys Asp Glu Asp Met Glu
1505 1510 1515
Ile Asp Asp Glu Leu Val Trp Ile His Pro Asp Ser Cys Leu Lys
1520 1525 1530

Gly Arg Lys Gly Phe Ser Thr Thr Lys Gly Lys Leu Tyr Ser Tyr
1535 1540 1545
Phe Glu Gly Thr Lys Phe His Gln Ala Ala Lys Asp Met Ala Glu
1550 1555 1560
Ile Lys Val Leu Phe Pro Asn Asp Gln Glu Ser Asn Glu Gln Leu
1565 1570 1575

Cys Ala Tyr Ile Leu Gly Glu Thr Met Glu Ala Ile Arg Glu Lys
1580 1585 1590

Cys Pro Val Asp His Asn Pro Ser Ser Ser Pro Pro Lys Thr Leu
1595 1600 1605
Pro Cys Leu Cys Met Tyr Ala Met Thr Pro Glu Arg Val His Arg
1610 1615 1620
Leu Arg Ser Asn Asn Val Lys Glu Val Thr Val Cys Ser Ser Thr
1625 1630 1635
Pro Leu Pro Lys His Lys Ile Lys Asn Val Gln Lys Val Gln Cys
1640 1645 1650

Thr Lys Val Val Leu Phe Asn Pro His Thr Pro Ala Phe Val Pro
1655 1660 1665
Ala Arg Lys Tyr Ile Glu Val Pro Glu Gln Pro Thr Ala Pro Pro
1670 1675 1680
Ala Gln Ala Glu Glu Ala Pro Glu Val Val Ala Thr Pro Ser Pro
1685 1690 1695
Ser Thr Ala Asp Asn Thr Ser Leu Asp Val Thr Asp Ile Ser Leu
1700 1705 1710

Asp Met Asp Asp Ser Ser Glu Gly Ser Leu Phe Ser Ser Phe Ser
1715 1720 1725
Gly Ser Asp Asn Ser Ile Thr Ser Met Asp Ser Trp Ser Ser Gly
1730 1735 1740
Pro Ser Ser Leu Glu Ile Val Asp Arg Arg Gln Val Val Val Ala
1745 1750 1755
Asp Val His Ala Val Gln Glu Pro Ala Pro Ile Pro Pro Pro Arg
1760 1765 1770

Leu Lys Lys Met Ala Arg Leu Ala Ala Ala Arg Lys Glu Pro Thr
1775 1780 1785
Pro Pro Ala Ser Asn Ser Ser Glu Ser Leu His Leu Ser Phe Gly
1790 1795 1800
Gly Val Ser Met Ser Leu Gly Ser Ile Phe Asp Gly Glu Thr Ala

1805 1810 1815

Arg Gln Ala Ala Val Gln Pro Leu Ala Thr Gly Pro Thr Asp Val

1820 1825 1830

Pro Met Ser Phe Gly Ser Phe Ser Asp Gly Glu Ile Asp Glu Leu

1835 1840 1845

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

1850 1855 1860

Glu Pro Gly Glu Val Asn Ser Ile Ile Ser Ser Arg Ser Ala Val

1865 1870 1875

Ser Phe Pro Leu Arg Lys Gln Arg Arg Arg Arg Arg Ser Arg Arg

1880 1885 1890

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

1895 1900 1905

Thr Gly Pro Gly His Leu Gln Lys Lys Ser Val Leu Gln Asn Gln

1910 1915 1920

Leu Thr Glu Pro Thr Leu Glu Arg Asn Val Leu Glu Arg Ile His

1925 1930 1935

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

1940 1945 1950

Tyr Gln Met Met Pro Thr Glu Ala Asn Lys Ser Arg Tyr Gln Ser

1955 1960 1965

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

1970 1975 1980

Ser Gly Leu Arg Leu Tyr Asn Ser Ala Thr Asp Gln Pro Glu Cys

1985 1990 1995

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

2000 2005 2010

Ala Asn Tyr Ser Asp Pro Gln Phe Ala Val Ala Val Cys Asn Asn

2015 2020 2025

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

2030 2035 2040

```
Asp Glu  Tyr Asp Ala Tyr Leu  Asp Met Val Asp Gly  Thr Val Ala
    2045                 2050                 2055
Cys Leu  Asp Thr Ala Thr Phe  Cys Pro Ala Lys Leu  Arg Ser Tyr
    2060                 2065                 2070

Pro Lys  Lys His Glu Tyr Arg  Ala Pro Asn Ile Arg  Ser Ala Val
    2075                 2080                 2085
Pro Ser  Ala Met Gln Asn Thr  Leu Gln Asn Val Leu  Ile Ala Ala
    2090                 2095                 2100
Thr Lys  Arg Asn Cys Asn Val  Thr Gln Met Arg Glu  Leu Pro Thr
    2105                 2110                 2115
Leu Asp  Ser Ala Thr Phe Asn  Val Glu Cys Phe Arg  Lys Tyr Ala
    2120                 2125                 2130

Cys Asn  Asp Glu Tyr Trp Glu  Glu Phe Ala Arg Lys  Pro Ile Arg
    2135                 2140                 2145
Ile Thr  Thr Glu Phe Val Thr  Ala Tyr Val Ala Arg  Leu Lys Gly
    2150                 2155                 2160
Pro Lys  Ala Ala Ala Leu Phe  Ala Lys Thr Tyr Asn  Leu Val Pro
    2165                 2170                 2175
Leu Gln  Glu Val Pro Met Asp  Arg Phe Val Met Asp  Met Lys Arg
    2180                 2185                 2190

Asp Val  Lys Val Thr Pro Gly  Thr Lys His Thr Glu  Glu Arg Pro
    2195                 2200                 2205
Lys Val  Gln Val Ile Gln Ala  Ala Glu Pro Leu Ala  Thr Ala Tyr
    2210                 2215                 2220
Leu Cys  Gly Ile His Arg Glu  Leu Val Arg Arg Leu  Thr Ala Val
    2225                 2230                 2235
Leu Leu  Pro Asn Ile His Thr  Leu Phe Asp Met Ser  Ala Glu Asp
    2240                 2245                 2250

Phe Asp  Ala Ile Ile Ala Glu  His Phe Lys Gln Gly  Asp Pro Val
    2255                 2260                 2265
Leu Glu  Thr Asp Ile Ala Ser  Phe Asp Lys Ser Gln  Asp Asp Ala
```

2270 2275 2280

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

2285 2290 2295

Gln Pro Leu Leu Asp Leu Ile Glu Cys Ala Phe Gly Glu Ile Ser

2300 2305 2310

Ser Thr His Leu Pro Thr Gly Thr Arg Phe Lys Phe Gly Ala Met

2315 2320 2325

Met Lys Ser Gly Met Phe Leu Thr Leu Phe Val Asn Thr Val Leu

2330 2335 2340

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

2345 2350 2355

Ser Arg Cys Ala Ala Phe Ile Gly Asp Asp Asn Ile Ile His Gly

2360 2365 2370

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

2375 2380 2385

Asn Met Glu Val Lys Ile Ile Asp Ala Val Ile Gly Glu Arg Pro

2390 2395 2400

Pro Tyr Phe Cys Gly Gly Phe Ile Leu Gln Asp Ser Val Thr Ser

2405 2410 2415

Thr Ala Cys Arg Val Ala Asp Pro Leu Lys Arg Leu Phe Lys Leu

2420 2425 2430

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

2435 2440 2445

Arg Ala Leu Leu Asp Glu Thr Lys Ala Trp Phe Arg Val Gly Ile

2450 2455 2460

Thr Gly Thr Leu Ala Val Ala Val Thr Thr Arg Tyr Glu Val Asp

2465 2470 2475

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

2480 2485 2490

Lys Arg Ala Phe Gln Ala Ile Arg Gly Glu Ile Lys His Leu Tyr

2495 2500 2505

Gly Gly Pro Lys

2510

<210> 5

<211> 2474

<212> PRT

<213> Chikungunya virus

<400> 5

Met Asp Pro Val Tyr Val Asp Ile Asp Ala Asp Ser Ala Phe Leu Lys
1 5 10 15

Ala Leu Gln Arg Ala Tyr Pro Met Phe Glu Val Glu Pro Arg Gln Val
20 25 30

Thr Pro Asn Asp His Ala Asn Ala Arg Ala Phe Ser His Leu Ala Ile

35 40 45

Lys Leu Ile Glu Gln Glu Ile Asp Pro Asp Ser Thr Ile Leu Asp Ile
50 55 60

Gly Ser Ala Pro Ala Arg Arg Met Met Ser Asp Arg Lys Tyr His Cys
65 70 75 80

Val Cys Pro Met Arg Ser Ala Glu Asp Pro Glu Arg Leu Ala Asn Tyr
85 90 95

Ala Arg Lys Leu Ala Ser Ala Ala Gly Lys Val Leu Asp Arg Asn Ile

100 105 110

Ser Gly Lys Ile Gly Asp Leu Gln Ala Val Met Ala Val Pro Asp Thr
115 120 125

Glu Thr Pro Thr Phe Cys Leu His Thr Asp Val Ser Cys Arg Gln Arg
130 135 140

Ala Asp Val Ala Ile Tyr Gln Asp Val Tyr Ala Val His Ala Pro Thr
145 150 155 160

Ser Leu Tyr His Gln Ala Ile Lys Gly Val Arg Leu Ala Tyr Trp Val

165 170 175

Gly Phe Asp Thr Thr Pro Phe Met Tyr Asn Ala Met Ala Gly Ala Tyr
180 185 190

Pro Ser Tyr Ser Thr Asn Trp Ala Asp Glu Gln Val Leu Lys Ala Lys

195 200 205

Asn Ile Gly Leu Cys Ser Thr Asp Leu Thr Glu Gly Arg Arg Gly Lys

210 215 220

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

225 230 235 240

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

245 250 255

Ser Trp His Leu Pro Ser Val Phe His Leu Lys Gly Lys Leu Ser Phe

260 265 270

Thr Cys Arg Cys Asp Thr Val Val Ser Cys Glu Gly Tyr Val Val Lys

275 280 285

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

290 295 300

Val Thr His His Ala Asp Gly Phe Leu Met Cys Lys Thr Thr Asp Thr

305 310 315 320

Val Asp Gly Glu Arg Val Ser Phe Ser Val Cys Thr Tyr Val Pro Ala

325 330 335

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

340 345 350

Glu 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 Met Ile

370 375 380

Pro Val Val Ala Gln Ala Phe Ser Lys Trp Ala Lys Glu Cys Arg Lys

385 390 395 400

Asp Met Glu Asp Glu Lys Leu Leu Gly Val Arg Glu Arg Thr Leu Thr

405 410 415

Cys Cys Cys Leu Trp Ala Phe Lys Lys Gln Lys Thr His Thr Val Tyr

420 425 430

Lys Arg Pro Asp Thr Gln Ser Ile Gln Lys Val Gln Ala Glu Phe Asp

435 440 445

Ser Phe Val Val Pro Ser Leu Trp Ser Ser Gly Leu Ser Ile Pro Leu
450 455 460
Arg Thr Arg Ile Lys Trp Leu Leu Ser Lys Val Pro Lys Thr Asp Leu
465 470 475 480
Thr Pro Tyr Ser Gly Asp Ala Gln Glu Ala Arg Asp Ala Glu Lys Glu

485 490 495
Ala Glu Glu Glu Arg Glu Ala Glu Leu Thr Leu Glu Ala Leu Pro Pro
500 505 510
Leu Gln Ala Ala Gln Glu Asp Val Gln Val Glu Ile Asp Val Glu Gln
515 520 525
Leu Glu Asp Arg Ala Gly Ala Gly Ile Ile Glu Thr Pro Arg Gly Ala
530 535 540
Ile Lys Val Thr Ala Gln Pro Thr Asp His Val Val Gly Glu Tyr Leu

545 550 555 560
Val Leu Ser Pro Gln Thr Val Leu Arg Ser Gln Lys Leu Ser Leu Ile
565 570 575
His Ala Leu Ala Glu Gln Val Lys Thr Cys Thr His Ser Gly Arg Ala
580 585 590
Gly Arg Tyr Ala Val Glu Ala Tyr Asp Gly Arg Val Leu Val Pro Ser
595 600 605
Gly Tyr Ala Ile Ser Pro Glu Asp Phe Gln Ser Leu Ser Glu Ser Ala

610 615 620
Thr Met Val Tyr Asn Glu Arg Glu Phe Val Asn Arg Lys Leu His His
625 630 635 640
Ile Ala Met His Gly Pro Ala Leu Asn Thr Asp Glu Glu Ser Tyr Glu
645 650 655
Leu Val Arg Ala Glu Arg Thr Glu His Glu Tyr Val Tyr Asp Val Asp
660 665 670
Gln Arg Arg Cys Cys Lys Lys Glu Glu Ala Ala Gly Leu Val Leu Val

675 680 685
Gly Asp Leu Thr Asn Pro Pro Tyr His Glu Phe Ala Tyr Glu Gly Leu

690 695 700

Lys Ile Arg Pro Ala Cys Pro Tyr Lys Ile Ala Val Ile Gly Val Phe

705 710 715 720

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

725 730 735

Arg Gln Asp Leu Val Thr Ser Gly Lys Lys Glu Asn Cys Gln Glu Ile

740 745 750

Thr Thr Asp Val Met Arg Gln Arg Gly Leu Glu Ile Ser Ala Arg Thr

755 760 765

Val Asp Ser Leu Leu Leu Asn Gly Cys Asn Arg Pro Val Asp Val Leu

770 775 780

Tyr Val Asp Glu Ala Phe Ala Cys His Ser Gly Thr Leu Leu Ala Leu

785 790 795 800

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

805 810 815

Lys Gln Cys Gly Phe Phe Asn Met Met Gln Met Lys Val Asn Tyr Asn

820 825 830

His Asn Ile Cys Thr Gln Val Tyr His Lys Ser Ile Ser Arg Arg Cys

835 840 845

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

850 855 860

Met Arg Thr Thr Asn Glu Tyr Asn Lys Pro Ile Val Val Asp Thr Thr

865 870 875 880

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

885 890 895

Gly Trp Val Lys Gln Leu Gln Ile Asp Tyr Arg Gly His Glu Val Met

900 905 910

Thr Ala Ala Ala Ser Gln Gly Leu Thr Arg Lys Gly Val Tyr Ala Val

915 920 925

Arg Gln Lys Val Asn Glu Asn Pro Leu Tyr Ala Ser Thr Ser Glu His

930 935 940

```
Val Asn Val Leu Leu Thr Arg Thr Glu Gly Lys Leu Val Trp Lys Thr
945                 950                 955                 960
Leu Ser Gly Asp Pro Trp Ile Lys Thr Leu Gln Asn Pro Pro Lys Gly
                965                 970                 975
Asn Phe Lys Ala Thr Ile Lys Glu Trp Glu Val Glu His Ala Ser Ile
            980                 985                 990
Met Ala Gly Ile Cys Ser His Gln  Met Thr Phe Asp Thr  Phe Gln Asn

        995                 1000                1005
Lys Ala  Asn Val Cys Trp Ala  Lys Ser Leu Val Pro  Ile Leu Glu
    1010                1015                1020
Thr Ala  Gly Ile Lys Leu Asn  Asp Arg Gln Trp Ser  Gln Ile Ile
    1025                1030                1035
Gln Ala  Phe Lys Glu Asp Lys  Ala Tyr Ser Pro Glu  Val Ala Leu
    1040                1045                1050
Asn Glu  Ile Cys Thr Arg Met  Tyr Gly Val Asp Leu  Asp Ser Gly

    1055                1060                1065
Leu Phe  Ser Lys Pro Leu Val  Ser Val Tyr Tyr Ala  Asp Asn His
    1070                1075                1080
Trp Asp  Asn Arg Pro Gly Gly  Lys Met Phe Gly Phe  Asn Pro Glu
    1085                1090                1095
Ala Ala  Ser Ile Leu Glu Arg  Lys Tyr Pro Phe Thr  Lys Gly Lys
    1100                1105                1110
Trp Asn  Ile Asn Lys Gln Ile  Cys Val Thr Thr Arg  Arg Ile Glu

    1115                1120                1125
Asp Phe  Asn Pro Thr Thr Asn  Ile Ile Pro Ala Asn  Arg Arg Leu
    1130                1135                1140
Pro His  Ser Leu Val Ala Glu  His Arg Pro Val Lys  Gly Glu Arg
    1145                1150                1155
Met Glu  Trp Leu Val Asn Lys  Ile Asn Gly His His  Val Leu Leu
    1160                1165                1170
Val Ser  Gly Cys Ser Leu Ala  Leu Pro Thr Lys Arg  Val Thr Trp
```

1175 1180 1185

Val Ala Pro Leu Gly Val Arg Gly Ala Asp Tyr Thr Tyr Asn Leu

1190 1195 1200

Glu Leu Gly Leu Pro Ala Thr Leu Gly Arg Tyr Asp Leu Val Val

1205 1210 1215

Ile Asn Ile His Thr Pro Phe Arg Ile His His Tyr Gln Gln Cys

1220 1225 1230

Val Asp His Ala Met Lys Leu Gln Met Leu Gly Gly Asp Ser Leu

1235 1240 1245

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

1250 1255 1260

Tyr Ala Asp Arg Thr Ser Glu Arg Val Ile Cys Val Leu Gly Arg

1265 1270 1275

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

1280 1285 1290

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

1295 1300 1305

Arg Asn Phe Thr Thr His Val Met Asn Asn Gln Leu Asn Ala Ala

1310 1315 1320

Phe Val Gly Gln Ala Thr Arg Ala Gly Cys Ala Pro Ser Tyr Arg

1325 1330 1335

Val Lys Arg Met Asp Ile Ala Lys Asn Asp Glu Glu Cys Val Val

1340 1345 1350

Asn Ala Ala Asn Pro Arg Gly Leu Pro Gly Asp Gly Val Cys Lys

1355 1360 1365

Ala Val Tyr Lys Lys Trp Pro Glu Ser Phe Lys Asn Ser Ala Thr

1370 1375 1380

Pro Val Gly Thr Ala Lys Thr Val Met Cys Gly Thr Tyr Pro Val

1385 1390 1395

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

1400 1405 1410

Gly Asp Arg Glu Leu Ala Ala Ala Tyr Arg Glu Val Ala Lys Glu

1415 1420 1425

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

1430 1435 1440

Thr Gly Val Tyr Ser Gly Gly Lys Asp Arg Leu Thr Gln Ser Leu

1445 1450 1455

Asn His Leu Phe Thr Ala Met Asp Ser Thr Asp Ala Asp Val Val

1460 1465 1470

Ile Tyr Cys Arg Asp Lys Glu Trp Glu Lys Lys Ile Ser Glu Ala

1475 1480 1485

Ile Gln Met Arg Thr Gln Val Glu Leu Leu Asp Glu His Ile Ser

1490 1495 1500

Ile Asp Cys Asp Val Val Arg Val His Pro Asp Ser Ser Leu Ala

1505 1510 1515

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

1520 1525 1530

Leu Glu Gly Thr Arg Phe His Gln Thr Ala Val Asp Met Ala Glu

1535 1540 1545

Ile Tyr Thr Met Trp Pro Lys Gln Thr Glu Ala Asn Glu Gln Val

1550 1555 1560

Cys Leu Tyr Ala Leu Gly Glu Ser Ile Glu Ser Ile Arg Gln Lys

1565 1570 1575

Cys Pro Val Asp Asp Ala Asp Ala Ser Ser Pro Pro Lys Thr Val

1580 1585 1590

Pro Cys Leu Cys Arg Tyr Ala Met Thr Pro Glu Arg Val Thr Arg

1595 1600 1605

Leu Arg Met Asn His Val Thr Ser Ile Ile Val Cys Ser Ser Phe

1610 1615 1620

Pro Leu Pro Lys Tyr Lys Ile Glu Gly Val Gln Lys Val Lys Cys

1625 1630 1635

Ser Lys Val Met Leu Phe Asp His Asn Val Pro Ser Arg Val Ser

1640 1645 1650

Pro Arg Glu Tyr Arg Pro Ser Gln Glu Ser Val Gln Glu Ala Ser

1655 1660 1665

Thr Thr Thr Ser Leu Thr His Ser Gln Phe Asp Leu Ser Val Asp

1670 1675 1680

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

1685 1690 1695

Ala Leu Glu Pro Ala Leu Asp Asp Gly Ala Ile His Thr Leu Pro

1700 1705 1710

Ser Ala Thr Gly Asn Leu Ala Ala Val Ser Asp Trp Val Met Ser

1715 1720 1725

Thr Val Pro Val Ala Pro Pro Arg Arg Arg Arg Gly Arg Asn Leu

1730 1735 1740

Thr Val Thr Cys Asp Glu Arg Glu Gly Asn Ile Thr Pro Met Ala

1745 1750 1755

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

1760 1765 1770

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

1775 1780 1785

Ser Thr Ala Thr Glu Leu Ser His Pro Pro Ile Ser Phe Gly Ala

1790 1795 1800

Pro Ser Glu Thr Phe Pro Ile Thr Phe Gly Asp Phe Asn Glu Gly

1805 1810 1815

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

1820 1825 1830

Leu Pro Gly Glu Val Asp Asp Leu Thr Asp Ser Asp Trp Ser Thr

1835 1840 1845

Cys Ser Asp Thr Asp Asp Glu Leu Arg Leu Asp Arg Ala Gly Gly

1850 1855 1860

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

1865 1870 1875

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

1880 1885 1890

His Glu Glu Lys Cys Tyr Pro Pro Lys Leu Asp Glu Ala Lys Glu

1895 1900 1905

Gln Leu Leu Leu Lys Lys Leu Gln Glu Ser Ala Ser Met Ala Asn

1910 1915 1920

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

1925 1930 1935

Ile Ile Gln Arg Leu Lys Arg Gly Cys Arg Leu Tyr Leu Met Ser

1940 1945 1950

Glu Thr Pro Lys Val Pro Thr Tyr Arg Thr Thr Tyr Pro Ala Pro

1955 1960 1965

Val Tyr Ser Pro Pro Ile Asn Val Arg Leu Ser Asn Pro Glu Ser

1970 1975 1980

Ala Val Ala Ala Cys Asn Glu Phe Leu Ala Arg Asn Tyr Pro Thr

1985 1990 1995

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

2000 2005 2010

Met Val Asp Gly Ser Glu Ser Cys Leu Asp Arg Ala Thr Phe Asn

2015 2020 2025

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

2030 2035 2040

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

2045 2050 2055

Gln Asn Val Leu Ala Ala Ala Thr Lys Arg Asn Cys Asn Val Thr

2060 2065 2070

Gln Met Arg Glu Leu Pro Thr Leu Asp Ser Ala Val Phe Asn Val

2075 2080 2085

Glu Cys Phe Lys Lys Phe Ala Cys Asn Gln Glu Tyr Trp Glu Glu

2090 2095 2100

Phe Ala Ala Ser Pro Ile Arg Ile Thr Thr Glu Asn Leu Thr Thr

2105 2110 2115

Tyr Val Thr Lys Leu Lys Gly Pro Lys Ala Ala Ala Leu Phe Ala

2120 2125 2130

Lys Thr His Asn Leu Leu Pro Leu Gln Glu Val Pro Met Asp Arg

2135 2140 2145

Phe Thr Val Asp Met Lys Arg Asp Val Lys Val Thr Pro Gly Thr

2150 2155 2160

Lys His Thr Glu Glu Arg Pro Lys Val Gln Val Ile Gln Ala Ala

2165 2170 2175

Glu Pro Leu Ala Thr Ala Tyr Leu Cys Gly Ile His Arg Glu Leu

2180 2185 2190

Val Arg Arg Leu Asn Ala Val Leu Leu Pro Asn Val His Thr Leu

2195 2200 2205

Phe Asp Met Ser Ala Glu Asp Phe Asp Ala Ile Ile Ala Ala His

2210 2215 2220

Phe Lys Pro Gly Asp Thr Val Leu Glu Thr Asp Ile Ala Ser Phe

2225 2230 2235

Asp Lys Ser Gln Asp Asp Ser Leu Ala Leu Thr Ala Leu Met Leu

2240 2245 2250

Leu Glu Asp Leu Gly Val Asp His Ser Leu Leu Asp Leu Ile Glu

2255 2260 2265

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

2270 2275 2280

Arg Phe Lys Phe Gly Ala Met Met Lys Ser Gly Met Phe Leu Thr

2285 2290 2295

Leu Phe Val Asn Thr Leu Leu Asn Ile Thr Ile Ala Ser Arg Val

2300 2305 2310

Leu Glu Asp Arg Leu Thr Lys Ser Ala Cys Ala Ala Phe Ile Gly

2315 2320 2325

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

2330 2335 2340

Ala Arg Cys Ala Thr Trp Met Asn Met Glu Val Lys Ile Ile Asp

2345 2350 2355

Ala Val Val Ser Gln Lys Ala Pro Tyr Phe Cys Gly Gly Phe Ile

2360 2365 2370

Leu His Asp Ile Val Thr Gly Thr Ala Cys Arg Val Ala Asp Pro

2375 2380 2385

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

2390 2395 2400

Glu Gln Asp Glu Asp Arg Arg Arg Ala Leu Ala Asp Glu Val Val

2405 2410 2415

Arg Trp Gln Arg Thr Gly Leu Ile Asp Glu Leu Glu Lys Ala Val

2420 2425 2430

Tyr Ser Arg Tyr Glu Val Gln Gly Ile Ser Val Val Val Met Ser

2435 2440 2445

Met Ala Thr Phe Ala Ser Ser Arg Ser Asn Phe Glu Lys Leu Arg

2450 2455 2460

Gly Pro Val Val Thr Leu Tyr Gly Gly Pro Lys

2465 2470

<210> 6

<211> 14

<212> PRT

<213> Chikungunya virus

<400> 6

Asn Glu Gly Glu Ile Glu Ser Leu Ser Ser Glu Leu Leu Thr

1 5 10

<210> 7

<211> 20

<212>

PRT

<213> Sindbis virus

<400> 7

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

1 5 10 15

Glu Pro Val Leu

20

<210> 8

<211> 13

<212> PRT

<213> Semliki Forest virus

<400> 8

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

1 5 10

<210> 9

<211> 30

<212> PRT

<213> Chikungunya virus

<400> 9

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

1 5 10 15

Ser Asp Thr Asp Asp Glu Leu Arg Leu Asp Arg Ala Gly Gly

20 25 30

<210> 10

<211> 33

<212> PRT

<213> Sindbis virus

<400> 10

Glu Pro Gly Glu Val Asn Ser Ile Ile Ser Ser Arg Ser Ala Val Ser

1 5 10 15

Phe Pro Leu Arg Lys Gln Arg Arg Arg Arg Arg Ser Arg Arg Thr Glu

20 25 30

Tyr

<210> 11

<211> 11

<212> PRT

<213> Semliki Forest virus

<400> 11

Asp Asp Val Leu Arg Leu Gly Arg Ala Gly Ala

1 5 10

<210> 12

<211> 10

<212> PRT

<213> Venezuelan equine encephalitis virus

<400> 12

Leu His Glu Ala Gly Cys Ala Pro Ser Tyr

1 5 10

<210> 13

<211> 10

<212> PRT

<213> Venezuelan equine encephalitis virus

<400> 13

Arg Phe Asp Ala Gly Ala Tyr Ile Phe Ser

1 5 10

<210> 14

<211> 10

<212> PRT

<213> Eastern equine encephalomyelitis virus

<400> 14

Gln His Glu Ala Gly Arg Ala Pro Ala Tyr

1 5 10

<210> 15

<211> 10

<212> PRT

<213> Eastern equine encephalomyelitis virus

<400> 15

Arg Tyr Glu Ala Gly Ala Tyr Ile Phe Ser

1 5 10

<210> 16

<211> 10

<212> PRT

<213> Western equine encephalomyelitis virus

<400> 16

Arg Tyr Glu Ala Gly Arg Ala Pro Ala Tyr

1 5 10

<210> 17

<211> 10

<212> PRT

<213> Western equine encephalomyelitis virus

<400> 17

Arg Tyr Glu Ala Gly Ala Tyr Ile Phe Ser

1 5 10

<210> 18

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 18

Phe Gly Asp Phe

1

<210> 19

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 19

Phe Gly Ser Phe

1

<210> 20

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 20

Phe Gly Asp Phe Phe Gly Asp Phe

1 5

<210> 21

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 21

Phe Gly Ser Phe Phe Gly Ser Phe

1 5

<210> 22

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 22

Phe Gly Asp Phe Phe Gly Ser Phe

1 5

<210> 23

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> OW repeat motif

<400> 23

Phe Gly Ser Phe Phe Gly Asp Phe

1 5

<210> 24

<211> 482

<212> PRT

<213> Semliki Forest virus

<400> 24

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

1 5 10 15

Ala Ala Val Val Asn Ala Ala Asn Ala Arg Gly Thr Val Gly Asp Gly

20 25 30

Val Cys Arg Ala Val Ala Lys Lys Trp Pro Ser Ala Phe Lys Gly Ala

35 40 45

Ala Thr Pro Val Gly Thr Ile Lys Thr Val Met Cys Gly Ser Tyr Pro

50 55 60

Val Ile His Ala Val Ala Pro Asn Phe Ser Ala Thr Thr Glu Ala Glu

65 70 75 80

Gly Asp Arg Glu Leu Ala Ala Val Tyr Arg Ala Val Ala Ala Glu Val

85 90 95

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

100 105 110

Val Phe Ser Gly Gly Arg Asp Arg Leu Gln Gln Ser Leu Asn His Leu

115 120 125

Phe Thr Ala Met Asp Ala Thr Asp Ala Asp Val Thr Ile Tyr Cys Arg

130 135 140

Asp Lys Ser Trp Glu Lys Lys Ile Gln Glu Ala Ile Asp Met Arg Thr

145 150 155 160

Ala Val Glu Leu Leu Asn Asp Asp Val Glu Leu Thr Thr Asp Leu Val

165 170 175

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

180 185 190

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

195 200 205

Ala Ala Ile Asp Met Ala Glu Ile Leu Thr Leu Trp Pro Arg Leu Gln

210 215 220

Glu Ala Asn Glu Arg Ile Cys Leu Tyr Ala Leu Gly Glu Thr Met Asp

225 230 235 240

Asn Ile Gly Ser Lys Cys Pro Val Asn Asp Ser Asp Ser Ser Thr Pro

245 250 255

Pro Arg Thr Val Pro Cys Leu Cys Arg Tyr Ala Met Thr Ala Glu Arg

260 265 270

Ile Ala Arg Leu Arg Ser His Gln Val Lys Ser Met Val Val Cys Ser
275 280 285
Ser Phe Pro Leu Pro Lys Tyr His Val Asp Gly Val Gln Lys Val Lys

290 295 300
Cys Glu Lys Val Leu Leu Phe Asp Pro Thr Val Pro Ser Val Val Ser
305 310 315 320
Pro Arg Lys Tyr Ala Ala Ser Thr Thr Asp His Ser Asp Arg Ser Leu
325 330 335
Arg Gly Phe Asp Leu Asp Trp Thr Thr Asp Ser Ser Ser Thr Ala Ser
340 345 350
Asp Thr Met Ser Leu Pro Ser Leu Gln Ser Cys Asp Ile Asp Ser Ile

355 360 365
Tyr Glu Pro Met Ala Pro Ile Val Val Thr Ala Asp Val His Pro Glu
370 375 380
Pro Ala Gly Ile Ala Asp Leu Ala Ala Asp Val His Pro Glu Pro Ala
385 390 395 400
Asp His Val Asp Leu Glu Asn Pro Ile Pro Pro Pro Arg Pro Lys Arg
405 410 415
Ala Ala Tyr Leu Ala Ser Arg Ala Ala Glu Arg Pro Val Pro Ala Pro

420 425 430
Arg Lys Pro Thr Pro Ala Pro Arg Thr Ala Phe Arg Asn Lys Leu Pro
435 440 445
Leu Thr Phe Gly Asp Phe Asp Glu His Glu Val Asp Ala Leu Ala Ser
450 455 460
Gly Ile Thr Phe Gly Asp Phe Asp Asp Val Leu Arg Leu Gly Arg Ala
465 470 475 480
Gly Ala

<210> 25
<211> 549
<212
> PRT

<213> Sindbis virus

<400> 25

Ala Pro Ser Tyr Arg Thr Lys Arg Glu Asn Ile Ala Asp Cys Gln Glu
1 5 10 15

Glu Ala Val Val Asn Ala Ala Asn Pro Leu Gly Arg Pro Gly Glu Gly
20 25 30

Val Cys Arg Ala Ile Tyr Lys Arg Trp Pro Thr Ser Phe Thr Asp Ser
35 40 45

Ala Thr Glu Thr Gly Thr Ala Arg Met Thr Val Cys Leu Gly Lys Lys
50 55 60

Val Ile His Ala Val Gly Pro Asp Phe Arg Lys His Pro Glu Ala Glu
65 70 75 80

Ala Leu Lys Leu Leu Gln Asn Ala Tyr His Ala Val Ala Asp Leu Val
85 90 95

Asn Glu His Asn Ile Lys Ser Val Ala Ile Pro Leu Leu Ser Thr Gly
100 105 110

Ile Tyr Ala Ala Gly Lys Asp Arg Leu Glu Val Ser Leu Asn Cys Leu
115 120 125

Thr Thr Ala Leu Asp Arg Thr Asp Ala Asp Val Thr Ile Tyr Cys Leu
130 135 140

Asp Lys Lys Trp Lys Glu Arg Ile Asp Ala Ala Leu Gln Leu Lys Glu
145 150 155 160

Ser Val Thr Glu Leu Lys Asp Glu Asp Met Glu Ile Asp Asp Glu Leu
165 170 175

Val Trp Ile His Pro Asp Ser Cys Leu Lys Gly Arg Lys Gly Phe Ser
180 185 190

Thr Thr Lys Gly Lys Leu Tyr Ser Tyr Phe Glu Gly Thr Lys Phe His
195 200 205

Gln Ala Ala Lys Asp Met Ala Glu Ile Lys Val Leu Phe Pro Asn Asp
210 215 220

Gln Glu Ser Asn Glu Gln Leu Cys Ala Tyr Ile Leu Gly Glu Thr Met
225 230 235 240

Glu Ala Ile Arg Glu Lys Cys Pro Val Asp His Asn Pro Ser Ser Ser
245 250 255

Pro Pro Lys Thr Leu Pro Cys Leu Cys Met Tyr Ala Met Thr Pro Glu
260 265 270
Arg Val His Arg Leu Arg Ser Asn Asn Val Lys Glu Val Thr Val Cys
275 280 285
Ser Ser Thr Pro Leu Pro Lys His Lys Ile Lys Asn Val Gln Lys Val
290 295 300
Gln Cys Thr Lys Val Val Leu Phe Asn Pro His Thr Pro Ala Phe Val
305 310 315 320

Pro Ala Arg Lys Tyr Ile Glu Val Pro Glu Gln Pro Thr Ala Pro Pro
325 330 335
Ala Gln Ala Glu Glu Ala Pro Glu Val Val Ala Thr Pro Ser Pro Ser
340 345 350
Thr Ala Asp Asn Thr Ser Leu Asp Val Thr Asp Ile Ser Leu Asp Met
355 360 365
Asp Asp Ser Ser Glu Gly Ser Leu Phe Ser Ser Phe Ser Gly Ser Asp
370 375 380

Asn Ser Ile Thr Ser Met Asp Ser Trp Ser Ser Gly Pro Ser Ser Leu
385 390 395 400
Glu Ile Val Asp Arg Arg Gln Val Val Val Ala Asp Val His Ala Val
405 410 415
Gln Glu Pro Ala Pro Ile Pro Pro Pro Arg Leu Lys Lys Met Ala Arg
420 425 430
Leu Ala Ala Ala Arg Lys Glu Pro Thr Pro Pro Ala Ser Asn Ser Ser
435 440 445

Glu Ser Leu His Leu Ser Phe Gly Gly Val Ser Met Ser Leu Gly Ser
450 455 460
Ile Phe Asp Gly Glu Thr Ala Arg Gln Ala Ala Val Gln Pro Leu Ala
465 470 475 480
Thr Gly Pro Thr Asp Val Pro Met Ser Phe Gly Ser Phe Ser Asp Gly

485 490 495

Glu Ile Asp Glu Leu Ser Arg Arg Val Thr Glu Ser Glu Pro Val Leu

500 505 510

Phe Gly Ser Phe Glu Pro Gly Glu Val Asn Ser Ile Ile Ser Ser Arg

515 520 525

Ser Ala Val Ser Phe Pro Leu Arg Lys Gln Arg Arg Arg Arg Arg Ser

530 535 540

Arg Arg Thr Glu Tyr

545

<210> 26

<211> 530

<212> PRT

<213> Chikungunya virus

<400> 26

Ala Pro Ser Tyr Arg Val Lys Arg Met Asp Ile Ala Lys Asn Asp Glu

1 5 10 15

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

20 25 30

Val Cys Lys Ala Val Tyr Lys Lys Trp Pro Glu Ser Phe Lys Asn Ser

35 40 45

Ala Thr Pro Val Gly Thr Ala Lys Thr Val Met Cys Gly Thr Tyr Pro

50 55 60

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

65 70 75 80

Gly Asp Arg Glu Leu Ala Ala Ala Tyr Arg Glu Val Ala Lys Glu Val

85 90 95

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

100 105 110

Val Tyr Ser Gly Gly Lys Asp Arg Leu Thr Gln Ser Leu Asn His Leu

115 120 125

Phe Thr Ala Met Asp Ser Thr Asp Ala Asp Val Val Ile Tyr Cys Arg

130 135 140

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

Thr Ala Val Asp Met Ala Glu Ile Tyr Thr Met Trp Pro Lys Gln Thr
210 215 220
Glu Ala Asn Glu Gln Val Cys Leu Tyr Ala Leu Gly Glu Ser Ile Glu
225 230 235 240
Ser Ile Arg Gln Lys Cys Pro Val Asp Asp Ala Asp Ala Ser Ser Pro
245 250 255
Pro Lys Thr Val Pro Cys Leu Cys Arg Tyr Ala Met Thr Pro Glu Arg
260 265 270

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

Thr Thr Ser Leu Thr His Ser Gln Phe Asp Leu Ser Val Asp Gly Lys
340 345 350
Ile Leu Pro Val Pro Ser Asp Leu Asp Ala Asp Ala Pro Ala Leu Glu
355 360 365
Pro Ala Leu Asp Asp Gly Ala Ile His Thr Leu Pro Ser Ala Thr Gly
370 375 380

Asn Leu Ala Ala Val Ser Asp Trp Val Met Ser Thr Val Pro Val Ala
385 390 395 400

Pro Pro Arg Arg Arg Arg Gly Arg Asn Leu Thr Val Thr Cys Asp Glu
405 410 415
Arg Glu Gly Asn Ile Thr Pro Met Ala Ser Val Arg Phe Phe Arg Ala
420 425 430
Glu Leu Cys Pro Val Val Gln Glu Thr Ala Glu Thr Arg Asp Thr Ala
435 440 445
Met Ser Leu Gln Ala Pro Pro Ser Thr Ala Thr Glu Leu Ser His Pro
450 455 460

Pro Ile Ser Phe Gly Ala Pro Ser Glu Thr Phe Pro Ile Thr Phe Gly
465 470 475 480
Asp Phe Asn Glu Gly Glu Ile Glu Ser Leu Ser Ser Glu Leu Leu Thr
485 490 495
Phe Gly Asp Phe Leu Pro Gly Glu Val Asp Asp Leu Thr Asp Ser Asp
500 505 510
Trp Ser Thr Cys Ser Asp Thr Asp Asp Glu Leu Arg Leu Asp Arg Ala
515 520 525

Gly Gly
530

<210> 27
<211> 556
<212> PRT
<213> Venezuelan equine encephalitis virus
<400> 27
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> 28

<

211> 552

<212> PRT

<213> Eastern equine encephalomyelitis virus

<400> 28

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

1 5 10 15

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

20 25 30

Val Cys Gly Ala Leu Tyr Arg Lys Trp Pro Gly Ala Phe Asp Lys Gln

35 40 45

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

50 55 60

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

65 70 75 80

Asp Gln Lys Leu Ser Glu Val Tyr Met Asp Ile Ala Arg Ile Ile Asn

85 90 95

Asn Glu Arg Phe Thr Lys Val Ser Ile Pro Leu Leu Ser Thr Gly Ile

100 105 110

Tyr Ala Gly Gly Lys Asp Arg Val Met Gln Ser Leu Asn His Leu Phe

115 120 125

Thr Ala Met Asp Thr Thr Asp Ala Asp Ile Thr Ile Tyr Cys Leu Asp

130 135 140

Lys Gln Trp Glu Ser Arg Ile Lys Glu Ala Ile Thr Arg Lys Glu Ser

145 150 155 160

Val Glu Glu Leu Thr Glu Asp Asp Arg Pro Val Asp Ile Glu Leu Val

165 170 175

Arg Val His Pro Leu Ser Ser Leu Ala Gly Arg Pro Gly Tyr Ser Thr

180 185 190

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

245 250 255
Pro His Thr Ile Pro Cys Leu Cys Asn Tyr Ala Met Thr Ala Glu Arg
260 265 270
Val Tyr Arg Leu Arg Met Ala Lys Asn Glu Gln Phe Ala Val Cys Ser
275 280 285
Ser Phe Gln Leu Pro Lys Tyr Arg Ile Thr Gly Val Gln Lys Ile Gln
290 295 300
Cys Ser Lys Pro Val Ile Phe Ser Gly Thr Val Pro Pro Ala Ile His

305 310 315 320
Pro Arg Lys Phe Ala Ser Val Thr Val Glu Asp Thr Pro Val Val Gln
325 330 335
Pro Glu Arg Leu Val Pro Arg Arg Pro Ala Pro Pro Val Pro Val Pro
340 345 350
Ala Arg Ile Pro Ser Pro Pro Cys Thr Ser Thr Asn Gly Ser Thr Thr
355 360 365
Ser Ile Gln Ser Leu Gly Glu Asp Gln Ser Ala Ser Ala Ser Ser Gly

370 375 380
Ala Glu Ile Ser Val Asp Gln Val Ser Leu Trp Ser Ile Pro Ser Ala
385 390 395 400
Thr Gly Phe Asp Val Arg Thr Ser Ser Ser Leu Ser Leu Glu Gln Pro
405 410 415
Thr Phe Pro Thr Met Val Val Glu Ala Glu Ile His Ala Ser Gln Gly
420 425 430
Ser Leu Trp Ser Ile Pro Ser Ile Thr Gly Ser Glu Thr Arg Ala Pro

435 440 445

Ser Pro Pro Ser Gln Asp Ser Arg Pro Ser Thr Pro Ser Ala Ser Gly

450 455 460

Ser His Thr Ser Val Asp Leu Ile Thr Phe Asp Ser Val Ala Glu Ile

465 470 475 480

Leu Glu Asp Phe Ser Arg Ser Pro Phe Gln Phe Leu Ser Glu Ile Lys

485 490 495

Pro Ile Pro Ala Pro Arg Thr Arg Val Asn Asn Met Ser Arg Ser Ala

500 505 510

Asp Thr Ile Lys Pro Ile Pro Lys Pro Arg Lys Cys Gln Val Lys Tyr

515 520 525

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

530 535 540

Glu Phe Val Arg Arg His Ser Asn

545 550

<210> 29

<211> 532

<212> PRT

<213> Western equine encephalomyelitis virus

<220><221> misc_feature

<222> (526)..(526)

<223> Xaa can be any naturally occurring amino acid

<400> 29

Ala Pro Ala Tyr Arg Val Ile Arg Gly Asp Ile Ser Lys Ser Ala Asp

1 5 10 15

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

20 25 30

Val Cys Gly Ala Leu Tyr Arg Lys Trp Pro Ala Ala Phe Asp Arg Gln

35 40 45

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

50 55 60

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

65 70 75 80

Asp Leu Lys Leu Ala Ala Ala Tyr Met Ser Ile Ala Ser Ile Val Asn

85 90 95

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

100 105 110

Tyr Ser Gly Gly Lys Asp Arg Val Met Gln Ser Leu His His Leu Phe

115 120 125

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

130 135 140

Lys Gln Trp Glu Thr Arg Ile Ile Glu Ala Ile His Arg Lys Glu Ser

145 150 155 160

Val Glu Ile Leu Asp Asp Asp Lys Pro Val Asp Ile Asp Leu Val Arg

165 170 175

Val His Pro Asn Ser Ser Leu Ala Gly Arg Pro Gly Tyr Ser Val Asn

180 185 190

Glu Gly Lys Leu Tyr Ser Tyr Leu Glu Gly Thr Arg Phe His Gln Thr

195 200 205

Ala Lys Asp Ile Ala Glu Ile His Ala Met Trp Pro Asn Lys Ser Glu

210 215 220

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

225 230 235 240

Ile Arg Ser Lys Cys Pro Val Glu Glu Ser Glu Ala Ser Ala Pro Pro

245 250 255

His Thr Leu Pro Cys Leu Cys Asn Tyr Ala Met Thr Ala Glu Arg Val

260 265 270

Tyr Arg Leu Arg Ser Ala Lys Lys Glu Gln Phe Ala Val Cys Ser Ser

275 280 285

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

290 295 300

Ser Lys Pro Val Leu Phe Ser Gly Val Val Pro Pro Ala Val His Pro

305 310 315 320

Arg Lys Tyr Ala Glu Ile Ile Leu Glu Thr Pro Pro Pro Pro Ala Thr
325 330 335
Thr Thr Val Ile Cys Glu Pro Thr Val Pro Glu Arg Ile Pro Ser Pro
340 345 350
Val Ile Ser Arg Ala Pro Ser Ala Glu Ser Leu Leu Ser Leu Gly Gly
355 360 365
Val Ser Phe Ser Ser Ser Ala Thr Arg Ser Ser Thr Ala Trp Ser Asp

370 375 380
Tyr Asp Arg Arg Phe Val Val Thr Ala Asp Val His Gln Ala Asn Thr
385 390 395 400
Ser Thr Trp Ser Ile Pro Ser Ala Pro Gly Leu Asp Val Gln Leu Pro
405 410 415
Ser Asp Val Thr Asp Ser His Trp Ser Ile Pro Ser Ala Ser Gly Phe
420 425 430
Glu Val Arg Thr Pro Ser Val Gln Asp Leu Thr Ala Glu Cys Ala Lys

435 440 445
Pro Arg Gly Leu Ala Glu Ile Met Gln Asp Phe Asn Thr Ala Pro Phe
450 455 460
Gln Phe Leu Ser Asp Tyr Arg Pro Val Pro Ala Pro Arg Arg Arg Pro
465 470 475 480
Ile Pro Ser Pro Arg Ser Thr Ala Ser Ala Pro Pro Val Pro Lys Pro
485 490 495
Arg Arg Thr Lys Tyr Gln Gln Pro Pro Gly Val Ala Arg Ala Ile Ser

500 505 510
Glu Ala Glu Leu Asp Glu Tyr Ile Arg Gln His Ser Asn Xaa Arg Tyr
515 520 525
Glu Ala Gly Ala
530

<210> 30
<211> 555
<212> PRT
<213> Artificial Sequence

<220><223> VEEV/CHIKV nsP3 chimera

<400> 30

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 Ser Thr

325 330 335

Thr Thr Ser Leu Thr His Ser Gln Phe Asp Leu Ser Val Asp Gly Lys

340 345 350

Ile Leu Pro Val Pro Ser Asp Leu Asp Ala Asp Ala Pro Ala Leu Glu

355 360 365

Pro Ala Leu Asp Asp Gly Ala Ile His Thr Leu Pro Ser Ala Thr Gly

370 375 380

Asn Leu Ala Ala Val Ser Asp Trp Val Met Ser Thr Val Pro Val Ala

385 390 395 400

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

405 410 415

Arg Glu Gly Asn Ile Thr Pro Met Ala Ser Val Arg Phe Phe Arg Ala

420 425 430

Glu Leu Cys Pro Val Val Gln Glu Thr Ala Glu Thr Arg Asp Thr Ala

435 440 445

Met Ser Leu Gln Ala Pro Pro Ser Thr Ala Thr Glu Leu Ser His Pro

450 455 460

Pro Ile Ser Phe Gly Ala Pro Ser Glu Thr Phe Pro Ile Thr Phe Gly

465 470 475 480

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

530 535 540
Val Ala Gln Gln Gln Arg Phe Asp Ala Gly Ala
545 550 555

<210> 31
<211> 556
<212> PRT
<213> Artificial Sequence
<220><223> VEEV/SINV nsP3 chimera
<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 Pro
                325                 330                 335
Ala Gln Ala Glu Glu Ala Pro Glu Val Val Ala Thr Pro Ser Pro Ser
            340                 345                 350

Thr Ala Asp Asn Thr Ser Leu Asp Val Thr Asp Ile Ser Leu Asp Met
        355                 360                 365
Asp Asp Ser Ser Glu Gly Ser Leu Phe Ser Ser Phe Ser Gly Ser Asp
```

370 375 380

Asn Ser Ile Thr Ser Met Asp Ser Trp Ser Ser Gly Pro Ser Ser Leu

385 390 395 400

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

405 410 415

Gln Glu Pro Ala Pro Ile Pro Pro Pro Arg Leu Lys Lys Met Ala Arg

420 425 430

Leu Ala Ala Ala Arg Lys Glu Pro Thr Pro Pro Ala Ser Asn Ser Ser

435 440 445

Glu Ser Leu His Leu Ser Phe Gly Gly Val Ser Met Ser Leu Gly Ser

450 455 460

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

465 470 475 480

Thr Gly Pro Thr Asp Val Pro Met Ser Phe Gly Ser Phe Ser Asp Gly

485 490 495

Glu Ile Asp Glu Leu Ser Arg Arg Val Thr Glu Ser Glu Pro Val Leu

500 505 510

Phe Gly Ser Phe Glu Pro Gly Glu Val Asn Ser Ile Ile Ser Ser Arg

515 520 525

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

530 535 540

Phe Val Ala Gln Gln Gln Arg Phe Asp Ala Gly Ala

545 550 555