(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0008801
(43) 공개일자 2023년01월16일

(51) 국제특허분류(Int. Cl.)
*C07K 14/005* (2006.01) *A61K 39/00* (2006.01)
*A61K 39/215* (2006.01) *A61P 31/14* (2006.01)
(52) CPC특허분류
*C07K 14/005* (2013.01)
*A61K 39/215* (2013.01)
(21) 출원번호 10-2022-7042785
(22) 출원일자(국제) 2021년05월07일
심사청구일자 없음
(85) 번역문제출일자 2022년12월06일
(86) 국제출원번호 PCT/US2021/031256
(87) 국제공개번호 WO 2021/226436
국제공개일자 2021년11월11일
(30) 우선권주장
63/021,319 2020년05월07일 미국(US)
*(뒷면에 계속)*

(71) 출원인
**트랜슬레이트 바이오 인코포레이티드**
미국 매사추세츠주 02451 월섬 웨스트 스트리트 200
(72) 발명자
**디아스, 아누샤**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**트랜, 캉, 앤**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
*(뒷면에 계속)*
(74) 대리인
**양영준, 김영**

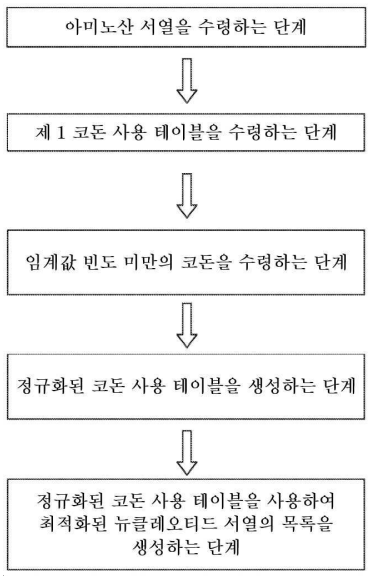
전체 청구항 수 : 총 108 항

(54) 발명의 명칭 **SARS-COV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열**

## *(57) 요 약*

본 발명은 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열에 관한 것이다. 이들 서열은 COVID-19 감염을 포함하는, β-코로나바이러스에 의해 야기되는 감염의 치료 또는 예방을 위해, 이러한 치료를 필요로 하는 인간 또는 동물 대상체에서 백신 조성물에 사용하기에 특히 적합하다.

*대 표 도* - 도1a



(52) CPC특허분류
***A61P 31/14*** (2018.01)
*A61K 2039/53* (2013.01)
*A61K 2039/545* (2013.01)
*A61K 2039/55555* (2013.01)
*C12N 2770/20022* (2013.01)
*C12N 2770/20034* (2013.01)

(72) 발명자
**자차리아, 민니**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**구, 샤오보**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**뵈글린, 리안**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**스케일스키, 조셉, 에이.**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**카르브, 쉬리랑**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**데로사, 프랑크**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내
**푸, 통-밍**
미국 18370 펜실베니아 스위프트워터 디스커버리 드라이브 1 사노피 파스퇴르 인크. 내
**칼닌, 키릴**
미국 18370 펜실베니아 스위프트워터 디스커버리 드라이브 1 사노피 파스퇴르 인크. 내
**치부쿨라, 수드하**
미국 18370 펜실베니아 스위프트워터 디스커버리 드라이브 1 사노피 파스퇴르 인크. 내
**플리트니크, 티모시**
미국 18370 펜실베니아 스위프트워터 디스커버리 드라이브 1 사노피 파스퇴르 인크. 내
**카시미로, 다닐로**
미국 18370 펜실베니아 스위프트워터 디스커버리 드라이브 1 사노피 파스퇴르 인크. 내
**두빈스, 제프리, 에스.**
미국 02421 매사추세츠 렉싱턴 하트웰 애비뉴 29 트랜슬레이트 바이오 인코포레이티드 내

(30) 우선권주장
63/032,825 2020년06월01일 미국(US)
63/076,729 2020년09월10일 미국(US)
63/076,718 2020년09월10일 미국(US)
63/088,739 2020년10월07일 미국(US)
63/143,604 2021년01월29일 미국(US)
63/143,612 2021년01월29일 미국(US)
63/146,807 2021년02월08일 미국(US)

# 명 세 서

## 청구범위

### 청구항 1

퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산으로서, 상기 최적화된 뉴클레오티드 서열은 10% 이상의 사용 빈도와 연관된 코돈으로 이루어지되, 상기 최적화된 뉴클레오티드 서열은,

(i) 다음의 뉴클레오티드 서열 중 하나를 갖는 종결 신호를 함유하지 않고,

5'-$X_1$ATCT$X_2$T$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, T 또는 G로부터 독립적으로 선택됨), 및 5'-$X_1$AUCU$X_2$U$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, U 또는 G로부터 독립적으로 선택됨);

(ii) 시스 조절 요소 및 음성 반복 요소를 함유하지 않으며;

(iii) 0.8을 초과하는 코돈 적용 인덱스를 가지되;

중첩되지 않는 30개의 뉴클레오티드-길이의 부분으로 나누어질 경우, 상기 최적화된 뉴클레오티드 서열의 각 부분은 30% 내지 70%의 구아닌 시토신 함량 범위를 갖는, 핵산.

### 청구항 2

제1항에 있어서, 최적화된 뉴클레오티드 서열은, TATCTGTT; TTTTTT; AAGCTT; GAAGAGC; TCTAGA; UAUCUGUU; UUUUUU; AAGCUU; GAAGAGC; UCUAGA의 서열 중 하나를 갖는 종결 신호를 함유하지 않는, 핵산.

### 청구항 3

제1항 또는 제2항에 있어서, 최적화된 서열에 의해 암호화된 전장 SARS-CoV-2 스파이크 단백질은 L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 함유하는, 핵산.

### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 핵산은 mRNA인, 핵산.

### 청구항 5

제1항 내지 제3항 중 어느 한 항에 있어서, 핵산은 DNA인, 핵산.

### 청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 11을 포함하는 아미노산 서열을 암호화하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는, 핵산.

### 청구항 7

제1항 내지 제5항 중 어느 한 항에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 167을 포함하는 아미노산 서열을 암호화하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166 또는 서열번호 173의 핵산 서열을 갖는, 핵산.

### 청구항 8

치료 요법에 사용하기 위한, 제1항 내지 제7항 중 어느 한 항의 핵산.

### 청구항 9

제1항 내지 제7항 중 어느 한 항에 있어서, β-코로나바이러스에 의해 야기되는 감염증의 예방에 사용하기 위한, 핵산을 포함하는 면역원성 조성물.

**청구항 10**

제9항에 따른 용도에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 면역원성 조성물.

**청구항 11**

제9항 또는 제10항에 따른 용도에 있어서, β-코로나바이러스는 SARS-CoV-2인, 면역원성 조성물.

**청구항 12**

제9항 내지 제11항 중 어느 한 항에 따른 용도에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 면역원성 조성물.

**청구항 13**

β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 제1항 내지 제7항 중 어느 한 항의 핵산을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

**청구항 14**

제13항에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 방법.

**청구항 15**

제13항 또는 제14항에 있어서, β-코로나바이러스는 SARS-CoV-2인, 방법.

**청구항 16**

제13항 내지 제15항에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 방법.

**청구항 17**

약학적 조성물로서, i) 제1항 내지 제7항 중 어느 한 항의 핵산, 및 ii) 지질 나노입자를 포함하는, 약학적 조성물.

**청구항 18**

제17항에 있어서, 핵산은 지질 나노입자 중 캡슐화되는, 약학적 조성물.

**청구항 19**

제17항 또는 제18항에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 및 PEG-변형 지질을 포함하는, 약학적 조성물.

**청구항 20**

제19항에 있어서,

a. 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고;

b. 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고;

c. 콜레스테롤계 지질은 콜레스테롤이고;

d. PEG-변형 지질은 DMG-PEG-2K인, 약학적 조성물.

**청구항 21**

제19항 또는 제20항에 있어서, 양이온성 지질은 약 30 내지 60%의 몰비로, 예를 들어, 약 35 내지 40%의 몰비로 지질 나노입자를 구성하는, 약학적 조성물.

**청구항 22**

제19항 내지 제21항 중 어느 한 항에 있어서, 양이온성 지질 대 비-양이온성 지질 대 콜레스테롤계 지질 대 PEG-변형 지질의 비율은 몰비로 대략 30 내지 60:25 내지 35:20 내지 30:1 내지 15인, 약학적 조성물.

**청구항 23**

제19항 내지 제22항 중 어느 한 항에 있어서, 지질 나노입자는 cKK-E12, DOPE, 콜레스테롤 및 DMG-PEG2K; cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-Deg-Lin, DOPE, 콜레스테롤 및 DMG-PEG2K; 또는 OF-02, DOPE, 콜레스테롤 및 DMG-PEG2K를 포함하는, 약학적 조성물.

**청구항 24**

제19항 내지 제23항 중 어느 한 항에 있어서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 몰비 40:30:28.5:1.5로 포함하는, 약학적 조성물.

**청구항 25**

제17항 내지 제24항 중 어느 한 항에 있어서, 지질 나노입자는 150 nm 미만, 예를 들어, 100 nm 미만의 평균 크기를 갖는, 약학적 조성물.

**청구항 26**

제25항에 있어서, 지질 나노입자는 약 50 내지 70 nm, 예를 들어, 약 55 내지 65 nm의 평균 크기를 갖는, 약학적 조성물.

**청구항 27**

제17항 내지 제26항 중 어느 한 항에 있어서, 핵산은 약 0.5 mg/mL 내지 약 1.0 mg/mL 농도의 mRNA인, 약학적 조성물.

**청구항 28**

제17항 내지 제27항 중 어느 한 항에 있어서, β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 데 사용하기 위한, 약학적 조성물.

**청구항 29**

제28항에 따른 용도에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 약학적 조성물.

**청구항 30**

제28항 또는 제29항에 따른 용도에 있어서, β-코로나바이러스는 SARS-CoV-2인, 약학적 조성물.

**청구항 31**

제28항 내지 제30항 중 어느 한 항에 따른 용도에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 약학적 조성물.

**청구항 32**

제28항 내지 제31항 중 어느 한 항에 따른 용도에 있어서, 약학적 조성물은 근육내 투여되는, 약학적 조성물.

**청구항 33**

제32항에 따른 용도에 있어서, 약학적 조성물은 적어도 1회 투여되는, 약학적 조성물.

**청구항 34**

제33항에 따른 용도에 있어서, 약학적 조성물은 적어도 2회 투여되는, 약학적 조성물.

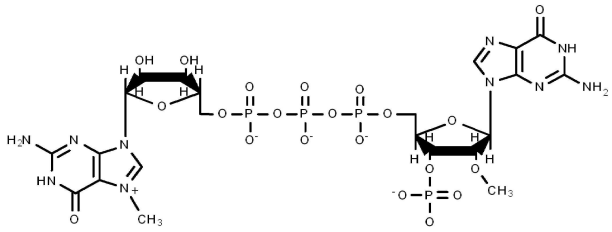
**청구항 35**

제34항에 따른 용도에 있어서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월인, 약학적 조성물.

**청구항 36**

mRNA 작제물로서, 다음의 구조 요소:

(i) 하기 구조를 갖는 5' 캡:



(ii) 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

(iii) 서열번호 148의 핵산 서열을 갖는 단백질 코딩 영역;

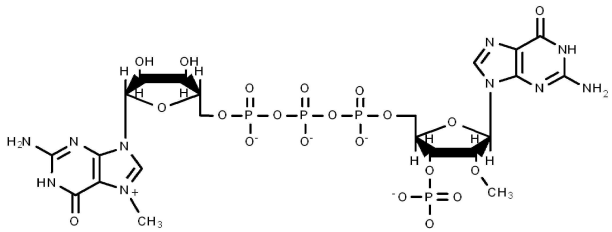
(iv) 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

폴리A 꼬리로 이루어지는, mRNA 작제물.

**청구항 37**

mRNA 작제물로서, 다음의 구조 요소:

(i) 다음의 구조를 갖는 5' 캡:



(ii) 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

(iii) 서열번호 173의 핵산 서열을 갖는 단백질 코딩 영역;

(iv) 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

폴리A 꼬리로 이루어지는, mRNA 작제물..

**청구항 38**

제36항의 mRNA 작제물을 캡슐화하는, 지질 나노입자.

**청구항 39**

제37항의 mRNA 작제물을 캡슐화하는, 지질 나노입자.

**청구항 40**

제36항의 mRNA 작제물 및 제37항의 mRNA 작제물을 캡슐화하는, 지질 나노입자.

**청구항 41**

제38항 내지 제40항 중 어느 한 항에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함하는, 지질 나노입자.

**청구항 42**

제41항에 있어서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K인, 지질 나노입자.

**청구항 43**

제42항에 있어서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 몰비 40:30:28.5:1.5로 포함하는, 지질 나노입자.

**청구항 44**

제36항의 mRNA 작제물 및/또는 제37항의 mRNA 작제물, 또는 제38항 내지 제43항 중 어느 한 항의 지질 나노입자를 포함하는, 면역원성 조성물.

**청구항 45**

제44항에 있어서, 5 μg 내지 200 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 46**

제45항에 있어서, 7 μg 내지 135 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 47**

제46항에 있어서, 적어도 10 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 48**

제46항에 있어서, 적어도 15 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 49**

제46항에 있어서, 적어도 20 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 50**

제46항에 있어서, 적어도 25 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 51**

제46항에 있어서, 적어도 35 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 52**

제46항에 있어서, 적어도 40 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 53**

제46항에 있어서, 적어도 45 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 54**

제46항에 있어서, 7.5 μg, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물(들)을 포함하는, 면역원성 조성물.

**청구항 55**

β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 제44항 내지 제54항 중 어느 한 항의 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

**청구항 56**

제55항에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 방법.

**청구항 57**

제55항 또는 제56항에 있어서, β-코로나바이러스는 SARS-CoV-2인, 방법.

**청구항 58**

제55항 내지 제57항 중 어느 한 항에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 방법.

**청구항 59**

제55항 내지 제58항 중 어느 한 항에 있어서, 면역원성 조성물은 대상체에게 적어도 2회 투여되는, 방법.

**청구항 60**

제59항에 있어서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월인, 방법.

**청구항 61**

적어도 2개의 핵산을 포함하는 면역원성 조성물로서,

1. 제1 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고;

2. 제2 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함하는, 면역원성 조성물.

**청구항 62**

제61항에 있어서, 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는, 면역원성 조성물.

**청구항 63**

제61항 또는 제62항에 있어서, 제2 핵산은 서열번호 167을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166 또는 서열번호 173의 핵산 서열을 갖는, 면역원성 조성물.

**청구항 64**

제61항 내지 제63항 중 어느 한 항에 있어서, 적어도 2개의 핵산은 mRNA 작제물인, 면역원성 조성물.

**청구항 65**

제64항에 있어서, 제1 핵산의 최적화된 뉴클레오티드 서열은 서열번호 148의 핵산 서열을 가지며, 제2 핵산의

최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 갖는, 면역원성 조성물.

**청구항 66**

제65항에 있어서, 제1 핵산은 제36항의 mRNA 작제물이고, 제2 핵산은 제37항의 mRNA 작제물인, 면역원성 조성물.

**청구항 67**

제61항 내지 제66항 중 어느 한 항에 있어서, 적어도 2개의 핵산은 지질 나노입자 중 캡슐화되는, 면역원성 조성물.

**청구항 68**

제67항에 있어서, 적어도 2개의 핵산은 동일한 지질 나노입자 중 캡슐화되는, 면역원성 조성물.

**청구항 69**

제67항에 있어서, 적어도 2개의 핵산은 개별적인 지질 나노입자 중 캡슐화되는, 면역원성 조성물.

**청구항 70**

제67항 내지 제69항 중 어느 한 항에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함하는, 면역원성 조성물.

**청구항 71**

제70항에 있어서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K인, 면역원성 조성물.

**청구항 72**

제71항에 있어서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 몰비 40:30:28.5:1.5로 포함하는, 면역원성 조성물.

**청구항 73**

제61항 내지 제72항 중 어느 한 항에 있어서, 적어도 2개의 핵산을 총 7.5 μg, 15 μg, 45 μg 또는 135 μg의 양으로 포함하는, 면역원성 조성물.

**청구항 74**

제61항 내지 제73항 중 어느 한 항에 있어서, β-코로나바이러스에 의해 야기되는 감염증의 예방에 사용하기 위한, 면역원성 조성물.

**청구항 75**

제74항에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 면역원성 조성물.

**청구항 76**

제74항 또는 제75항에 따른 용도에 있어서, β-코로나바이러스는 SARS-CoV-2인, 면역원성 조성물.

**청구항 77**

제74항 내지 제76항 중 어느 한 항에 따른 용도에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 면역원성 조성물.

**청구항 78**

β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

**청구항 79**

제78항에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 방법.

**청구항 80**

제78항 또는 제79항에 있어서, β-코로나바이러스는 SARS-CoV-2인, 방법.

**청구항 81**

제78항 내지 제80항에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 방법.

**청구항 82**

제78항 내지 제81항 중 어느 한 항에 있어서, 대상체는 상기 감염의 예방을 위한 면역원성 조성물을 이전에 투여받은 적이 없는, 방법.

**청구항 83**

제78항 내지 제81항 중 어느 한 항에 있어서, 대상체는 상기 감염의 예방을 위해 하나 이상의 면역원성 조성물(들)을 이전에 투여받은 적이 있는, 방법.

**청구항 84**

제83항에 있어서, 대상체는 상기 감염의 예방을 위해 적어도 2주 간격으로 2가지의 면역원성 조성물을 이전에 투여받은 적이 있는, 방법.

**청구항 85**

제83항 또는 제84항에 있어서, 하나 이상의 면역원성 조성물(들)은 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물과 상이한, 방법.

**청구항 86**

제83항 내지 제85항 중 어느 한 항에 있어서, 하나 이상의 면역원성 조성물(들)은,

a. 제9항 내지 제12항 중 어느 한 항에 따른 면역원성 조성물;

b. 제17항 내지 제35항 중 어느 한 항에 따른 약학적 조성물;

c. 제44항 내지 제54항 중 어느 한 항에 따른 면역원성 조성물; 및

d. 모데나(COVID-19 Vaccine Moderna, 예컨대, 예를 들어, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 또는 Novavax(NVX-CoV2373)로부터 선택되는, 방법.

**청구항 87**

제83항 내지 제86항 중 어느 한 항에 있어서, 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물은 하나 이상의 면역원성 조성물(들)의 투여 후 3 내지 18개월차에 투여되는, 방법.

**청구항 88**

제87항에 있어서, 제53항 내지 제65항 중 어느 한 항의 면역원성 조성물은 하나 이상의 면역원성 조성물(들)의

투여 후 적어도 9개월 또는 적어도 12개월차에 투여되는, 방법.

**청구항 89**

제83항 내지 제88항 중 어느 한 항에 있어서, 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물은 적어도 1회, 예를 들어 적어도 2회 투여되는, 방법.

**청구항 90**

β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 mRNA 작제물을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하되, 상기 mRNA 작제물은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함하는, 방법.

**청구항 91**

제90항에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 167을 포함하는 아미노산 서열을 암호화하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 갖는, 방법.

**청구항 92**

제90항 또는 제91항에 있어서, mRNA 작제물은 제31항의 mRNA 작제물인, 방법.

**청구항 93**

제90항 내지 제92항 중 어느 한 항에 있어서, mRNA 작제물은 지질 나노입자 중 캡슐화되는, 방법.

**청구항 94**

제93항에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함하는, 방법.

**청구항 95**

제94항에 있어서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K인, 방법.

**청구항 96**

제95항에 있어서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 몰비 40:30:28.5:1.5로 포함하는, 방법.

**청구항 97**

제90항 내지 제96항 중 어느 한 항에 있어서, 면역원성 조성물은 7.5 μg, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물을 포함하는, 방법.

**청구항 98**

제90항 내지 제97항에 있어서, β-코로나바이러스는 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는, 방법.

**청구항 99**

제90항 내지 제98항에 있어서, β-코로나바이러스는 SARS-CoV-2인, 방법.

**청구항 100**

제90항 내지 제99항에 있어서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는, 방법.

**청구항 101**

제90항 내지 제100항에 있어서, 대상체는 상기 감염의 예방을 위한 면역원성 조성물을 이전에 투여받은 적이 없는, 방법.

**청구항 102**

제90항 내지 제101항에 있어서, 대상체는 상기 감염의 예방을 위해 하나 이상의 면역원성 조성물(들)을 이전에 투여받은 적이 있는, 방법.

**청구항 103**

제102항에 있어서, 대상체는 상기 감염의 예방을 위해 적어도 2주 간격으로 2가지의 면역원성 조성물을 이전에 투여받은 적이 있는, 방법.

**청구항 104**

제102항 또는 제103항에 있어서, 하나 이상의 면역원성 조성물(들)은 제53항 내지 제65항 중 어느 한 항의 면역원성 조성물과 상이한, 방법.

**청구항 105**

제102항 내지 제104항 중 어느 한 항에 있어서, 제1 및/또는 제2 면역원성 조성물(들)은,

a. 제9항 내지 제12항 중 어느 한 항에 따른 면역원성 조성물;

b. 제17항 내지 제35항 중 어느 한 항에 따른 약학적 조성물; 또는

c. 제44항 내지 제54항 중 어느 한 항에 따른 면역원성 조성물; 및

d. 모데나(COVID-19 Vaccine Moderna, 예컨대, 예를 들어, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 또는 Novavax(NVX-CoV2373)로부터 선택되는, 방법.

**청구항 106**

제102항 내지 제105항 중 어느 한 항에 있어서, 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물은 하나 이상의 면역원성 조성물(들)의 투여 후 3 내지 18개월차에 투여되는, 방법.

**청구항 107**

제106항에 있어서, 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물은 하나 이상의 면역원성 조성물(들)의 투여 후 적어도 9개월 또는 적어도 12개월차에 투여되는, 방법.

**청구항 108**

제90항 내지 제107항 중 어느 한 항에 있어서, 제61항 내지 제73항 중 어느 한 항의 면역원성 조성물은 적어도 1회, 예를 들어 적어도 2회 투여되는, 방법.

## *발명의 설명*

### *기 술 분 야*

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은, 그 내용이 전체로서 본원에 통합되는, 2020년 5월 7일자로 출원된 미국 특허 가출원 일련번호 제63/021,319호, 2020년 6월 1일자로 출원된 미국 특허 가출원 일련번호 제63/032,825호, 2020년 9월 10일자로 출원된 미국 특허 가출원 일련번호 제63/076,718호, 2020년 9월 10일자로 출원된 미국 특허 가출원 일련번호 제

63/076,729호, 2020년 10월 7일자로 출원된 미국 특허 가출원 일련번호 제63/088,739호, 2021년 1월 29일자로 출원된 미국 특허 가출원 일련번호 제63/143,604호, 2021년 1월 29일자로 출원된 미국 특허 가출원 일련번호 제63/143,612호, 및 2021년 2월 8일자로 출원된 미국 특허 가출원 일련번호 제63/146,807호의 이익 및 이에 대한 우선권을 주장한다.

[0003] **서열 목록**

[0004] 본 명세서는 (2021년 5월 7일에 MRT-2161WO_SL이라는 명칭의 .txt 파일로 전자 제출된) 서열 목록을 참조로 한다. 전술한 일자에 생성된 .txt 파일은 757 KB 크기이다. 서열 목록의 전체 내용은 본원에 참조로서 포함된다.

[0005] **기술 분야**

[0006] 본 발명은 SARS-CoV-2 항원성 폴리펩티드 및 이들 SARS-CoV-2 항원성 폴리펩티드를 암호화하는 최적화된 뉴클레오티드 서열에 관한 것이다. 이들 항원성 폴리펩티드 및 최적화된 뉴클레오티드 서열은 COVID-19 감염을 포함하는, β-코로나바이러스에 의해 야기되는 감염의 치료 또는 예방을 위해, 이러한 치료를 필요로 하는 인간 또는 동물 대상체에서 백신 조성물에 사용하기에 특히 적합하다.

## *배 경 기 술*

[0007] 코로나바이러스 질병 2019(COVID-19) 팬데믹은 전 세계 공중 보건에 심각한 위협을 가한다. COVID-19의 원인 병원체는 새롭게 출현한 인간 병원균인 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2)이다.

[0008] 단백질 항원의 선택 및 설계 모두는 단백질-기반 또는 핵산-기반에 상관없이, 백신의 면역원성에 기여한다. 또한, mRNA-기반 백신과 같은 핵산-기반 면역원성 조성물과 관련하여, 하나 이상의 단백질 항원을 암호화하는 핵산으로부터 수득된 발현 수준은 그 효능에 상당한 영향을 미칠 수 있다.

[0009] 재조합 DNA 기술 및 핵산 서열화 및 합성의 진보는, 병원균의 게놈 서열이 결정되면, 단백질 항원을 신속하게 설계하는 것을 가능하게 하였다. 백신의 성공 또는 실패는 생체 내에서 중화 항체의 형태로 매우 효과적인 반응을 생성하는 항원성 폴리펩티드의 선택에 따라 달라질 수 있다. 따라서, COVID-19에 대한 예방을 제공하는 면역원성 조성물에 사용하기 위한 SARS-CoV-2 단백질로부터 유래된 새로운 항원성 폴리펩티드를 제공할 필요성이 존재한다.

[0010] 세포 내 mRNA로부터의 단백질의 효과적인 발현 또는 생성은 다양한 인자에 따라 달라진다. 단백질 코딩 뉴클레오티드 서열 내에서의 코돈의 조성물 및 순서의 최적화("코돈 최적화")는 mRNA 암호화 단백질의 보다 높은 발현을 야기할 수 있다. 코돈 최적화를 수행하는 다양한 방법이 당업계에 공지되어 있지만, 그 각각은 연산적 관점 및/또는 치료적 관점에서 상당한 단점 및 한계를 갖는다. 특히, 코돈 최적화의 공지된 방법은 종종 각각의 아미노산에 대해, 모든 코돈을 해당 아미노산에 대한 가장 높은 사용량을 갖는 코돈으로 대체하여, "최적화된" 서열이 각각의 아미노산을 암호화하는 하나의 코돈만을 함유하게 하는 것과 관련된다.

[0011] 따라서, 효과적인 mRNA 백신의 생산을 위해 선택되거나 설계된 단백질 항원을 암호화하는 mRNA의 발현을 증가시키기 위한 최적화된 뉴클레오티드 서열을 생성하는 개선된 코돈 최적화 방법이 필요하다.

[0012] 또한, SARS-CoV-2가 전 세계적으로 확산됨에 따라, 이 바이러스의 새로운 변이체가 출현하였다. 따라서, SARS-CoV-2의 자연적으로 발생하는 다수의 변이체에 대해 광범위한 중화 항체 반응을 효과적으로 유도할 수 있는 약학적 조성물(예를 들어, 면역원성 조성물)을 제공할 필요성이 존재한다.

## *발명의 내용*

[0013] 본 발명은 SARS-CoV-2에 대한 효과적인 면역 반응을 생성하는 단백질 항원의 선택 및/또는 설계에 대한 필요성을 다룬다. 이는 또한 최적화된 뉴클레오티드 서열을 갖는 핵산(예를 들어, mRNA)을 포함하는 백신의 제공을 통해, COVID-19 감염의 효과적인 치료 또는 예방을 위한 단백질 항원을 암호화하는 최적화된 뉴클레오티드 서열을 생성할 필요성을 다룬다. SARS-CoV-2에 대한 다양한 선택 및/또는 설계된 단백질 항원뿐만 아니라, 각각의 이러한 단백질 항원에 대한 적어도 하나의 최적화된 뉴클레오티드 서열이 본원에 제공된다.

[0014] 또한, 단백질 항원의 아미노산 서열을 분석하여 적어도 하나의 최적화된 뉴클레오티드 서열을 생성하는 방법이 제공된다. 각각의 선택 및/또는 설계된 단백질 항원에 대한 최적화된 뉴클레오티드 서열은 자연적으로 발생하는 뉴클레오티드 서열과 연관된 단백질의 발현과 비교하여 암호화된 단백질 항원의 발현을 증가시키도록 설계된다. 코돈 최적화는 유전자 코드의 중복으로 인한, 암호화된 단백질 항원의 번역된 아미노산의 서열을 변경하지 않고

다양한 기준에 기초하여 단백질 코딩 뉴클레오티드 서열을 생성한다. 또한, 본원에 개시된 최적화된 뉴클레오티드 서열은 시험관 내 합성 동안 고품질 전장 전사체를 생성하도록 설계되며, 따라서 종래 기술의 코돈 최적화 알고리즘으로 생성된 최적화된 뉴클레오티드 서열에 비해 보다 비용 효과적으로 제조될 수 있다. 특히, 시험관 내 합성 동안 불완전한 전사체를 초래할 수 있는 종결 서열 등은 본원에 기술된 서열 최적화 프로세스에 의해 효과적으로 제거된다.

[0015] 실시예에서 입증된 바와 같이, 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 본 발명의 LNP-캡슐화된 최적화된 뉴클레오티드 서열을 포함하는 면역원성 조성물은 효과적인 중화 항체 반응을 생성할 수 있고, 따라서 COVID-19 감염에 대한 보호 효능을 제공할 수 있다.

[0016] 본 발명은 또한, SARS-CoV-2의 자연 발생 변이체에 대해, 특히 중화 항체의 형태로, 광범위하게 효과적인 면역 반응을 유도할 수 있는 면역원성 조성물에 대한 필요성을 다룬다. 실시예에 나타낸 바와 같이, 발명자는 놀랍게도, 이전에 COVID-19 백신으로 면역화된 대상체에게 SARS-CoV-2 S 단백질의 남아공 변이체를 암호화하는 LNP-캡슐화된 최적화된 뉴클레오티드 서열을 포함하는 면역원성 조성물의 투여가, 우한, 남아공, 일본/브라질 및 캘리포니아에서 단리된 SARS-CoV-2의 자연 발생 변이체뿐만 아니라 계통발생학적으로 보다 먼 SARS-CoV-1 계통을 포함하는 광범위한 범위의 β-코로나바이러스에 대항하여 효과적인 중화 항체 반응을 유도할 수 있음을 발견하였다.

[0017] 특히, 본 발명은, 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공하며, 여기에서 최적화된 뉴클레오티드 서열은 10% 이상의 사용 빈도와 연관된 코돈으로 이루어지되, 최적화된 뉴클레오티드 서열은, 다음의 뉴클레오티드 서열 중 하나를 갖는 종결 신호를 함유하지 않고, 5'-$X_1$ATCTC$X_2$T$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, T 또는 G로부터 독립적으로 선택됨), 및 5'-$X_1$AUCUC$X_2$U$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, U 또는 G로부터 독립적으로 선택됨); 시스 조절 요소 및 음성 반복 요소를 함유하지 않으며; 0.8을 초과하는 코돈 적용 인덱스를 가지되; 중첩되지 않는 30개의 뉴클레오티드-길이의 부분으로 나누어질 경우, 최적화된 뉴클레오티드 서열의 각 부분은 30% 내지 70%의 구아닌 시토신 함량 범위를 갖는다. 특정 구현예에서, 핵산은 mRNA이다. 일부 구현예에서, 핵산은 DNA이다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 다음의 서열 중 하나를 갖는 종결 신호를 함유하지 않는다: TATCTGTT; TTTTTT; AAGCTT; GAAGAGC; TCTAGA; UAUCUGUU; UUUUUU; AAGCUU; GAAGAGC; UCUAGA.

[0018] 일부 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 11의 아미노산 서열을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고, 서열번호 11의 아미노산 서열을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는다.

[0019] 일부 구현예에서, 최적화된 서열에 의해 암호화된 전장 SARS-CoV-2 스파이크 단백질은 L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 함유한다. 이들 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 167의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드는 서열번호 166 또는 서열번호 173의 핵산 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고, 서열번호 167의 아미노산 서열을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 166 또는 서열번호 173의 핵산 서열을 갖는다.

[0020] 소정의 구현예에서, 본 발명의 핵산은 치료에 사용하기 위한 것이다. 예를 들어, 본 발명은 또한 β-코로나바이러스에 의해 야기되는 감염의 예방에 사용하기 위한, 본 발명의 핵산을 포함하는 면역원성 조성물을 제공한다. 이에 더하여, 본 발명은 β-코로나바이러스에 의해 야기되는 감염의 예방을 위한 의약의 제조에서의 본 발명의 핵산의 용도를 또한 제공한다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다.

[0021] 본 발명은 β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법을 추가로 제공하며, 전술한 방법은 본 발명의 핵산을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함한다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다.

[0022] 또한, 본 발명은 i) 본 발명의 핵산 및 ii) 지질 나노입자를 포함하는 약학적 조성물을 제공한다. 특정 구현예에서, 핵산은 mRNA이며, 이는 약 0.5 mg/mL 내지 약 1.0 mg/mL의 농도로 존재할 수 있다. 특정 구현예에서, 본 발명의 핵산(예를 들어, 본 발명에 따른 mRNA)은 지질 나노입자 중 캡슐화된다. 소정의 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 및 PEG-변형 지질을 포함한다. 특정 구현예에서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K이다. 소정의 구현예에서, 양이온성 지질은, 몰비로 지질 나노입자의 약 30 내지 60%, 예를 들어, 약 35 내지 40%를 구성한다. 소정의 구현예에서, 양이온성 지질 대 비-양이온성 지질 대 콜레스테롤계 지질 대 PEG-변형 지질의 비율은 몰비로 대략 30 내지 60:25 내지 35:20 내지 30:1 내지 15이다.

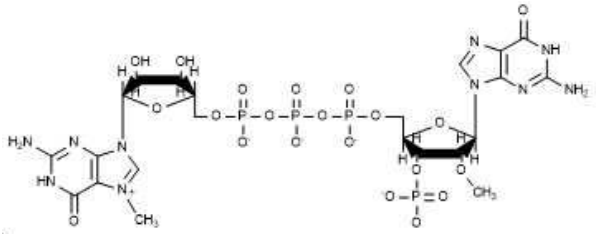
[0023] 소정의 구현예에서, 본 발명의 핵산(예를 들어, 본 발명에 따른 mRNA)을 캡슐화하는 지질 나노입자는, cKK-E12, DOPE, 콜레스테롤 및 DMG-PEG2K; cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-Deg-Lin, DOPE, 콜레스테롤 및 DMG-PEG2K; 또는 OF-02, DOPE, 콜레스테롤 및 DMG-PEG2K를 포함한다. 특정 구현예에서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 40:30:28.5:1.5의 몰비로 포함한다. 소정의 구현예에서, 지질 나노입자는 150 nm 미만, 예를 들어, 130 nm 미만, 110 nm 미만, 100 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 약 90 내지 110 nm의 평균 크기를 갖거나, 약 50 내지 70 nm, 예를 들어, 약 55 내지 65 nm의 평균 크기를 갖는다.

[0024] 소정의 구현예에서, 본 발명의 약학적 조성물은 β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 데 사용하기 위한 것이다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다.

[0025] 소정의 구현예에서, 본 발명의 약학적 조성물은 근육내 투여된다. 소정의 구현예에서, 본 발명의 약학적 조성물은 적어도 1회 투여된다. 일부 구현예에서, 약학적 조성물은 적어도 2회 투여된다. 특정 구현예에서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월이다. 일부 구현예에서, 투여 사이의 기간은 약 3주이다.

[0026] 특정 일 구현예에서, 본 발명은 다음의 구조적 요소로 이루어진 mRNA 작제물(mRNA 작제물 1)을 제공한다:

[0027] 다음의 구조를 갖는 5' 캡:

[0028] 

[0029] 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

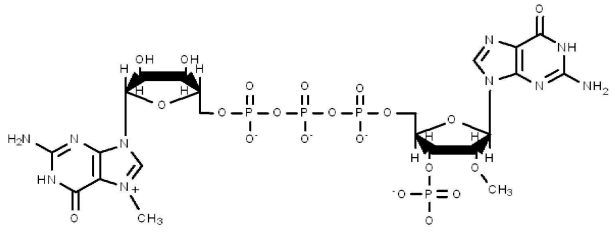
[0030] 서열번호 148의 핵산 서열을 갖는 단백질 코딩 영역;

[0031] 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

[0032] 폴리A 꼬리.

[0033] 또 다른 특정 구현예에서, 본 발명은 다음의 구조적 요소로 이루어진 mRNA 작제물(mRNA 작제물 2)을 제공한다:

[0034] 다음의 구조를 갖는 5' 캡:

[0035]  ;

[0036] 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

[0037] 서열번호 173의 핵산 서열을 갖는 단백질 코딩 영역;

[0038] 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

[0039] 폴리A 꼬리.

[0040] 특정 구현예에서, 본 발명은 본 발명의 mRNA 작제물을 캡슐화하는 지질 나노입자를 제공한다. 일부 구현예에서, 지질 나노입자는 본 발명의 하나 이상의 mRNA 작제물을 캡슐화한다; 즉, 예를 들어 지질 나노입자는 mRNA 작제물 1 및 mRNA 작제물 2 둘 모두를 캡슐화할 수 있다. 소정의 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함한다. 소정의 구현예에서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K이다. 특정 구현예에서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 40:30:28.5:1.5의 몰비로 포함한다.

[0041] 본 발명은 또한 본 발명의 mRNA 작제물, 또는 본 발명의 mRNA 작제물을 캡슐화하는 지질 나노입자를 포함하는 면역원성 조성물을 제공한다. 일부 구현예에서, 면역원성 조성물은 본 발명의 하나 이상의 mRNA 작제물, 예를 들어, mRNA 작제물 1 및 mRNA 작제물 2를 포함한다. 일부 구현예에서, 면역원성 조성물은 동일한 지질 나노입자 중 캡슐화된 하나 이상의 mRNA 작제물(예를 들어, mRNA 작제물 1 및 mRNA 작제물 2)을 포함한다. 다른 구현예에서, 하나 이상의 mRNA 작제물(예를 들어, mRNA 작제물 1 및 mRNA 작제물 2)은 개별적인 지질 나노입자 중 캡슐화된다. 소정의 구현예에서, 면역원성 조성물은 5 μg 내지 200 μg의 mRNA 작제물(들)을 포함한다.

[0042] 소정의 구현예에서, 면역원성 조성물은 7 μg 내지 135 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 10 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 15 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 20 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 25 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 35 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 40 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 적어도 45 μg의 mRNA 작제물(들)을 포함한다. 소정의 구현예에서, 면역원성 조성물은 7.5 μg, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물(들)을 포함한다. 일반적으로, 소정의 mRNA의 μg 양에 대한 언급은 면역원성 조성물 중의 mRNA의 총 투여량을 지칭한다.

[0043] 특정 구현예에서, 본 발명의 mRNA 작제물, 또는 본 발명의 mRNA 작제물을 캡슐화하는 지질 나노입자를 포함하는 면역원성 조성물은 β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 데 사용하기 위한 것이다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다.

[0044] 본 발명은 또한 β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법을 제공하며, 전술한 방법은 본 발명의 mRNA 작제물, 또는 본 발명의 mRNA 작제물을 캡슐화하는 지질 나노입자를 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함한다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다. 특정 구현예에서, 면역원성 조성물은 적어도 2회 대상체에게 투여된다. 소정의 구현예에서, 투여 사이의 기간은 적어도 2주이다. 일부 구현예에서, 투여 사이의 기간은 약 3주이다.

[0045] 특정 구현예에서, 본 발명은 적어도 2개의 핵산을 포함하는 면역원성 조성물을 제공하며, 여기에서 제1 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고; 제2 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함한다.

[0046] 일부 구현예에서, 제1 핵산은 서열번호 11의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한

다. 특정 구현예에서, 제1 핵산은 서열번호 44 또는 서열번호 148의 핵산 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고, 서열번호 11의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는다.

[0047] 일부 구현예에서, 제2 핵산은 서열번호 167의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 제2 핵산은 서열번호 168 또는 서열번호 173의 핵산 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고, 서열번호 167의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 166 또는 서열번호 169의 핵산 서열을 갖는다.

[0048] 특정 구현예에서, 적어도 2개의 핵산은 mRNA 작제물이다. 특정 구현예에서, 제1 핵산의 최적화된 뉴클레오티드 서열은 서열번호 148의 핵산 서열을 가지며, 제2 핵산의 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 갖는다. 특정 구현예에서, 제1 핵산은 mRNA 작제물 1이고, 제2 핵산은 mRNA 작제물 2이다. 소정의 구현예에서, 적어도 2개의 핵산은 지질 나노입자 중 캡슐화된다. 소정의 구현예에서, 적어도 2개의 핵산은 동일한 지질 나노입자 중 캡슐화된다. 소정의 구현예에서, 적어도 2개의 핵산은 개별적인 지질 나노입자 중 캡슐화된다.

[0049] 일부 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함한다. 소정의 구현예에서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K이다. 특정 구현예에서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 40:30:28.5:1.5의 몰비로 포함한다. 추가의 특정 구현예에서, 면역원성 조성물은 적어도 2개의 핵산의 총 7.5 μg, 15 μg, 45 μg 또는 135 μg을 포함한다.

[0050] 위의 네개의 단락(단락 [0030] 내지 [0034])에 기술된 면역원성 조성물은 β-코로나바이러스에 의해 야기되는 감염의 예방에 사용될 수 있다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%(예를 들어, 적어도 80%, 90%, 95% 또는 99%) 동일한 스파이크 단백질을 갖는다.

[0051] 특정 구현예에서, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위한 면역원성 조성물을 이전에 투여받지 않았다; 즉, 위의 네개의 단락(단락 [0030] 내지 단락 [0034])에 기술된 면역원성 조성물은 이러한 목적을 위해 대상체에게 투여되는 최초의 면역원성 조성물이다. 보다 일반적으로, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위해 하나 이상의 면역원성 조성물(들)을 이전에 투여받았다. 명확성을 위해, "대상체는 이전에 하나 이상의 면역원성 조성물(들)을 투여받았다"는 것은 해당 대상체가 이전에 동일한 면역원성 조성물의 1회 이상의 투여량 또는 상이한 면역원성 조성물(들)의 1회 이상의 투여량으로 투여되었다는 것을 의미한다. 예를 들어, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위해 적어도 2주 간격으로 2개의 면역원성 조성물을 이전에 투여받았을 수 있다. 일부 구현예에서, 이들 하나 이상의 면역원성 조성물(들)은 전술한 네개의 단락(단락 [0030] 내지 단락 [0034])에 기술된 면역원성 조성물(들)과 상이하다. 특정 구현예에서, 전술한 하나 이상의 면역원성 조성물(들)은 본원에 개시된 약학적 조성물(예를 들어, 본원에 개시된 면역원성 조성물 또는 백신), 및 모데나(Moderna)(COVIDCOVID-19 Vaccine Moderna, 예컨대, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(AstraZeneca)(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 또는 Novavax(NVX-CoV2373)에서 제조한 COVID-19 백신으로부터 선택된다. 특정 구현예에서, 위의 네개의 단락(단락 [0030] 내지 단락 [0034])에 기술된 면역원성 조성물은 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위해 대상체에게 이전에 투여된 하나 이상의 면역원성 조성물(들)의 투여 후 3 내지 18개월차에 투여된다. 특정 구현예에서, 위의 네개의 단락(단락 [0030] 내지 단락 [0034])에 기술된 면역원성 조성물은 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위해 대상체에게 이전에 투여된 하나 이상의 면역원성 조성물(들)의 투여 후 적어도 9개월 내지 적어도 12개월차에 투여된다. 특정 구현예에서, 위의 네개의 단락(단락 [0030] 내지 단락 [0034])에 기술된 면역원성 조성물은 적어도 1회, 예를 들어 적어도 2회 투여된다.

[0052] 또 다른 특정 구현예에서, 본 발명은 β-코로나바이러스에 의해 야기되는 감염을 치료하거나 예방하는 방법을 추가로 제공하며, 전술한 방법은 mRNA 작제물을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하며, 여기에서 전술한 mRNA 작제물은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연

변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 167의 아미노산 서열을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 166 또는 서열번호 173의 핵산 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고, 서열번호 167의 아미노산 서열을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 갖는다. 특정 구현예에서, mRNA 작제물은 mRNA 작제물 2이다. 소정의 구현예에서, mRNA 작제물은 지질 나노입자 중 캡슐화된다. 소정의 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함한다. 소정의 구현예에서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K이다. 특정 구현예에서, 지질 나노입자는 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 40:30:28.5:1.5의 몰비로 포함한다. 소정의 구현예에서, 면역원성 조성물은 7.5 μg, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물을 포함한다. 특정 구현예에서, β-코로나바이러스는 안지오텐신 전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현한다. 특정 구현예에서, β-코로나바이러스는 SARS-CoV-2이다. 다른 구현예에서, β-코로나바이러스는 서열번호 1과 적어도 75%, 80%, 90%, 95% 또는 99% 동일한 스파이크 단백질을 갖는다.

[0053] 바로 위의 단락(단락 [0037])에 기술된 방법에서, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위한 면역원성 조성물을 이전에 투여받지 않았을 수 있다. 보다 일반적으로, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염의 예방을 위해 하나 이상의 면역원성 조성물(들), 예를 들어, 2가지의 면역원성 조성물을 적어도 2주 간격으로 이전에 투여받았다. 소정의 구현예에서, 하나 이상의 면역원성 조성물(들)은 본 발명의 면역원성 조성물과 상이하다. 소정의 구현예에서, 대상체에게 이전에 투여된 전술한 하나 이상의 면역원성 조성물(들)은 본원에 개시된 약학적 조성물(예를 들어, 본원에 개시된 면역원성 조성물 또는 백신), 및 모데나(Moderna)(COVIDCOVID-19 Vaccine Moderna, 예컨대, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(AstraZeneca)(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 또는 Novavax(NVX-CoV2373)에서 제조한 COVID-19 백신으로부터 선택된다. 특정 구현예에서, 위의 단락(단락 [0037])에 기술된 방법은 대상체에게 이전에 투여된 하나 이상의 면역원성 조성물(들)의 투여 후 약 3 내지 18개월차에 해당 단락에 기술된 면역원성 조성물을 투여하는 단계를 포함한다. 특정 구현예에서, 위의 단락(단락 [0037])에 기술된 방법은 하나 이상의 면역원성 조성물(들)의 투여 후 적어도 9개월 또는 적어도 12개월차에 해당 단락에 기술된 면역원성 조성물을 투여하는 단계를 포함한다. 특정 구현예에서, 위의 단락(단락 [0037])에 기술된 방법은 해당 단락에 기술된 면역원성 조성물을 적어도 1회, 예를 들어 적어도 2회 투여하는 단계를 포함한다.

## *도면의 간단한 설명*

[0054] 도 1a 및 도 1b는 본 발명에 따른, 최적화된 뉴클레오티드 서열을 생성하기 위한 프로세스를 도시한다. 도 1a에 도시된 바와 같이, 프로세스는 관심 아미노산 서열 및 주어진 유기체(예를 들어, 포유동물 또는 인간)에서의 각 코돈의 빈도를 반영하는 제1 코돈 사용 테이블을 수령한다. 그런 다음, 프로세스는 임의의 코돈이 임계 빈도(예를 들어, 10%) 미만인 코돈 사용 빈도와 연관되는 경우, 제1 코돈 사용 테이블로부터 해당 코돈을 제거한다. 제1 단계에서 제거되지 않은 코돈의 코돈 사용 빈도는 정규화되어 정규화된 코돈 사용 테이블을 생성한다. 프로세스는 정규화된 코돈 사용 테이블을 사용하여 최적화된 뉴클레오티드 서열의 목록을 생성한다. 최적화된 뉴클레오티드 서열 각각은 관심 아미노산 서열을 암호화한다. 도 1b에 도시된 바와 같이, 최적화된 뉴클레오티드 서열의 목록은 최적화된 뉴클레오티드 서열의 업데이트된 목록을 생성하기 위해, 모티프 스크린 필터, 구아닌-시토신(GC) 함량 분석 필터, 및 코돈 적용 인덱스(CAI) 분석 필터를 이 순서로 적용함으로써 추가로 처리된다.

도 2는 EPO에 대한 ELISA 검정에 의해 결정된, 다양한 코돈 최적화된 뉴클레오티드 서열로부터 생성된 단백질의 수율을 도시하는 예시적인 막대 차트를 도시한다.

도 3은 SARS-CoV-2의 스파이크 단백질의 구조를 도시한다. SS = 신호 서열; NTD = N-말단 도메인; RBD = 수용체 결합 도메인; FP = 융합 펩티드; HR1 = 헵타드 반복-N; CH, 중심 나선; CTD, 커넥터 도메인; HR2, 헵타드 반복 2; TM, 막관통 도메인; CT, 세포질 꼬리. S2', S2' 프로테아제 절단 부위는 화살표로 표시되어 있다. PP 및 GSAS 돌연변이는 스파이크 단백질의 융합-전 형태를 유도한다. 이러한 이미지는 Wrapp 등의 문헌[Science 367, 6483, 1260-1263(2020)]의 도 1에 기초한다.

도 4는 본원에 개시된 약학적 조성물의 일부를 형성할 수 있거나, 예를 들어, 본원에 개시된 핵산-기반 백신에 사용하기 위해 본원에 개시된 최적화된 뉴클레오티드 서열에 의해 암호화될 수 있는 SARS-CoV-2의 스파이크 단백질 및 이의 변이체를 도시한다. 도메인 및 서브유닛, 퓨린 절단 부위를 제거하고 잔기 985, 986 및 987을 프롤린(P, PP, PPP 및 GSAS 돌연변이)으로 대체하기 위한 돌연변이, 및 관련 서열번호가 표시되어 있다. 도 3에서 사용된 동일한 약어가 사용된다.

도 5 내지 도 7은 전장 미처리 SARS-CoV-2 S 단백질(작제물 A) 및 SARS-CoV-2 S 단백질의 3개의 안정한 융합-전 형태(작제물 B 내지 D)를 암호화하는 최적화된 핵산 서열을 발현하는 핵산 벡터 작제물의 단백질 생성을 입증한다. 작제물 B는, 퓨린 절단 부위가 결여되고(따라서, S1 및 S2 서브유닛을 형성하도록 절단되지 않음), 프롤린을 잔기 986 및 987로서 함유하도록(따라서, 이의 융합-전 형태에서 단백질을 안정화함), 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 암호화한다. 작제물 C는, 프롤린을 잔기 986 및 987로서 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 암호화하고, 작제물 D는 퓨린 절단 부위가 결여되도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 암호화한다. 도 5 및 도 6은 작제물 A 및 B가 당화된 성숙 단백질(대략 225 kDa 밴드) 및 전처리된 전장 S 단백질(대략 170 내지 180 kDa 밴드)을 생성할 수 있음을 나타낸다. 도 5는 또한 작제물 A에서의 S1 및 S2 서브유닛 밴드의 존재를 나타내며, 이는 미처리 전장 SARS-CoV-2 S 단백질이 세포에 의해 올바르게 처리됨을 입증한다. 도 7은 4개의 작제물 모두가 전장 S 단백질을 생성할 수 있음을 입증한다. S1 및 S2 서브유닛 밴드는 작제물 A 및 작제물 C에서 검출되었다. 완전히 당화된 성숙 S 단백질의 강한 밴드는 작제물 B 및 작제물 D에서 검출되었다.

도 8은 본원에 개시된 약학적 조성물의 일부를 형성할 수 있거나, 예를 들어, 본원에 개시된 핵산-기반 백신에 사용하기 위해 본원에 개시된 최적화된 뉴클레오티드 서열에 의해 암호화될 수 있는 SARS-CoV-2의 스파이크 단백질 및 이의 변이체를 도시한다. 도메인, 서브유닛, 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린(P, PP, PPP, PPPPP 및 GSAS)으로 돌연변이화하기 위한 돌연변이, D614G 돌연변이, ER 회수 신호 및 연장된 N-말단 신호 펩티드의 제거, 및 관련 서열번호가 표시되어 있다. 도 3에서 사용된 동일한 약어가 사용된다.

도 9는, 전장 절단 불가능 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 지질 나노입자(LNP)-캡슐화된 mRNA의 면역원성 조성물이 마우스에서 강력한 결합 및 중화 항체 반응을 생성하였음을 도시한다. 도 9a는 0.2 μg, 1 μg, 5 μg 또는 10 μg의 LNP-캡슐화된 mRNA의 2회 투여로 면역화한 후의 마우스에서 유도된 ELISA 역가를 도시한다. mRNA-LNP 조성물의 희석제가 투여된 마우스 군은 음성 대조군으로서 기능한다. 도 9b는 슈도바이러스 기반 검정에 의해 결정했을 때, 0.2 μg, 1 μg, 5 μg 또는 10 μg의 LNP-캡슐화 mRNA 중 어느 하나의 2회 투여로 면역화한 후 마우스에서 생성된 중화 항체의 역가를 도시한다. 가볍고, 강하고 심각한 증상(Conv Sera)을 가진 COVID-19 환자로부터의 39개의 개별 전환기 혈청 샘플(전환기 혈청)은 양성 대조군으로서 기능한다. 도 9c에 도시된 바와 같이, 면역원성 조성물은 0일차 및 21일차에 투여되었다. 혈액을 -7일차(베이스라인), 14일차, 21일차, 28일차 및 35일차에 샘플링하였다.

도 10은, 전장 절단 불가능 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물이 마우스에서 Th1-편향 T-세포 반응을 생성하였음을 도시한다. 도 10a는 35일차에 단리된 비장세포가 높은 수준의 Th1 사이토카인 인터페론-γ(IFN-γ)를 분비하였음을 도시한다. 도 10b는 이들 비장세포가 검출 가능한 양의 Th2 사이토카인 IL-5를 분비하지 않았음을 도시한다. 도 10c에 도시된 바와 같이, 마우스를 0일차 및 21일차에 5 μg 또는 10 μg의 LNP-캡슐화된 mRNA의 2회 투여로 면역화하고, -4일차, 14일차, 21일차, 28일차 및 35일차에 혈액을 샘플링하고, 35일차에 비장을 채취하여 ELISPOT 검정으로 IFN-γ 및 IL-5 수준을 결정하였다.

도 11은, 전장 절단 불가능 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물이 시노몰구스 원숭이에서 강력한 결합 및 중화 항체 반응을 생성하였음을 도시한다. 도 11a는 15 μg, 45 μg, 또는 135 μg의 LNP-캡슐화된 mRNA의 2회 투여로 면역화한 후의 시노몰구스 원숭이에서 유도된 ELISA 역가를 도시한다. 도 11b는 15 μg, 45 μg, 또는 135 μg의 LNP-캡슐화된 mRNA의 2회 투여로 면역화한 후, 슈도바이러스 기반 검정으로 결정했을 때의 시노몰구스 원숭이에서 생성된 중화 항체의 역가를 도시한다. 도 11c는 15 μg, 45 μg, 또는 135 μg의 LNP-캡슐화된 mRNA 중 하나의 2회 투여로 면역화한 후의 시노몰구스 원숭이에서 생성된 마이크로중화 역가를 도시한다. 가볍고, 강하고 심각한 증상(Conv Sera)을 가진 COVID-19 환자로부터의 39개의 개별 전환기 혈청 샘플(전환기 혈청)은 도 11b 및 도 11c에서 도시된 검정에서의 양성 대조군으로서 기능한다. 도 11d에 도시된 바와 같이, 면역원성 조성물은 0일차 및

21일차에 투여되었다. 혈액을 -4일차(베이스라인), 2일차, 7일차, 14일차, 21일차, 23일차, 28일차. 35일차 및 42일차에 샘플링하였다. 시험 조성물에 의해 유도된 세포 매개 면역(CMI)을 결정하기 위해 42일차에 말초 혈액 단핵 세포(PMC)를 단리하였다.

도 12는, 전장 절단 불가능 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물이 시노몰구스 원숭이에서 Th1-편향 T-세포 반응을 생성하였음을 도시한다. 원숭이에게 0일차 및 21일차에의 5 μg 또는 10 μg의 LNP-캡슐화된 mRNA 중 하나의 2회 투여로 면역화시켰다. 도 12a 및 도 12c는 42일차에 단리된 PMBC가 펩티드 풀 S1 및 S2(SARS-CoV-2 S 단백질 유래 펩티드) 각각으로 자극 후, 높은 수준의 Th1 사이토카인 인터페론-γ(IFN-γ)를 분비하였음을 나타낸다. 도 12b 및 도 12d는 이들 PMBC가 펩티드 자극에 반응하여 Th2 사이토카인 IL-13의 베이스라인 수준만을 분비하였음을 나타낸다. 미처리(비활성화 및 비자극) 비장세포는 IFN-γ 및 IL-13의 베이스라인 수준(점선)을 설정하기 위한 대조군으로서 기능하였다.

도 13은 도 9 및 도 11에 요약된 데이터의 통계적 분석을 도시한다. 시험된 면역원성 조성물의 1 μg, 5 μg 및 10 μg 투여량 수준에 대한 마우스에서의 슈도바이러스(PsV) 역가는 대조군 인간 회복기 혈청 PsV 역가와 유의하게 상이하였다(도 13a). ELISA(IgG), 슈도바이러스(PsV) 및 마이크로중화(MN) 역가 간의 스피어만 상관 계수(Spearman Correlation Coefficient, SCC)를 도 11에 요약된 시노몰구스 원숭이 실험에 대해 계산하였다. SCC를 개별 동물 별로 수행하였고, 평균(± 표준 오차)을 투여량(N = 4) 또는 모든 시험 동물(N = 12) 별로 계산하였다. 도 13b는 이러한 분석의 결과를 나타낸다. 도 13c 및 도 13d는 시노몰구스 원숭이에서의 마이크로중화(MN) 및 슈도바이러스(PsV) 역가가 대조군으로서 역할을 한 인간 회복기 혈청의 MN 및 PsV 역가보다 유의하게 더 높았음을 나타낸다.

도 14는 전장 융합-전 안정화 SARS-CoV-2 S 단백질을 암호화하는 최적화된 mRNA를 포함하는 LNP 제형으로 면역화함으로써 마우스 및 NHP에서 유도된 중화 항체 역가를 도시한다. 각각의 5가지 제형(WT, 2P, GSAS, 2P/GSAS, 2P/GSAS/ALAYT)의 투여량 당 0.4 μg으로 3주 간격으로 2회의 면역화를 마우스에게 수행하였다. 6가지의 제형(2P, GSAS, 2P/GSAS, 2P/GSAS/ALAYT, 6P 및 6P/GSAS)의 투여량 당 5 μg으로 동일한 면역화 일정을 사용하여 비인간 영장류(NHP)를 면역화하였다. 면역화 전 동물(-4일차)로부터 투여 후 14, 21, 28, 35 및 42일차에 혈청 샘플을 채취하였다. 각각의 점은 개별 혈청 샘플을 나타내고, 선은 해당 군에 대한 기하 평균을 나타낸다. 각 패널에 대한 아래의 점선은 검정 판독의 하한을 나타낸다.

도 15는 시리아 골든 햄스터에서의 실시예 5의 LNP 제형의 보호 효능을 도시한다. (a) 단일 또는 2회 투여 요법 중 하나로 투여된 햄스터에서의 체중 감소; (b) 0.15 μg(-●-), 1.5 μg(-■-), 4.5 μg(-▲-), 13.5 μg(-▼-)의 투여량 중 하나, 샴(Sham)(-◆-)을 투여받은 햄스터 또는 미접종(-○-) 동물의 폐에서의 H&E 염색; (c) 단일 또는 2회의 투여 요법으로 면역화된 햄스터의 4일차 및 7일차 투여 후 병원성 점수; (d) 감염 후(DPI) 4일차 및 7일차에 대조군(샴 및 미처리)과 비교 시, 실시예 5의 LNP 제형의 2회 투여로 면역화된 햄스터의 폐 및 비강 조직에서의 SARS-CoV-2 하위게놈 mRNA(sgmRNA)의 정량화.

도 16은 실시예 14에 기술된 중화 검정에 사용된 슈도바이러스(PsV)의 제조를 위해 S 단백질을 유래한 계통을 제공한다. SARS-CoV-2 계통의 경우, 우한(Wuhan) 인덱스 계통으로부터의 SARS-CoV-2 S 단백질과 비교한 돌연변이는 D614G 돌연변이의 존재와 함께 나타난다. 가능한 경우, S-단백질 아미노산 서열의 GenBank 번호가 제공된다. PsV는 Integral Molecular로부터 수득하였으며, 또한 각 PsV에 대한 카탈로그 번호 및 로트 번호 둘 모두가 표시되어 있다.

도 17은, 실시예 5의 LNP 제형의 2회 투여로 이전에 면역화된 비인간 영장류(NHP)가, 남아공, 일본/브라질 및 캘리포니아에서 관찰된 S 단백질의 자연 발생 변이체, 및 SARS-CoV-2 S 단백질의 남아공 변이체를 암호화하는 부스터 mRNA 백신으로 면역화한 후 SARS-CoV-1 계통으로부터 유래된 S 단백질뿐만 아니라, 원래 우한 균주로부터 유래된 S 단백질에 대한 효과적인 중화 항체 반응을 나타냄을 도시한다. 실시예 5에 기술된 바와 같이, 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 mRNA를 포함하는 LNP 제형으로 0일차 및 35일차에 NHP에게 2회의 면역화를 수행하였다. 자연 발생 남아공 계통에서 관찰된 돌연변이를 갖는 상응하는 S 단백질을 암호화하는 mRNA를 포함하는 부스터 LNP 제형을 305일차에 주입하였다. 35, 308, 329 및 343일차에 혈청 샘플을 채취하였다. 각각의 점은 개별 혈청 샘플을 나타내고, 선은 해당 군에 대한 기하 평균을 나타낸다. 점선은 검출 하한을 나타낸다.

## *발명을 실시하기 위한 구체적인 내용*

[0055] **정의**

[0056] 본 발명을 보다 쉽게 이해하기 위하여, 먼저 특정 용어를 아래와 같이 정의한다. 다음의 용어 및 기타 용어에 대한 추가적인 정의가 본 명세서 전반에 걸쳐 기재되어 있다.

[0057] 본 명세서 및 첨부된 청구범위에서 사용된 바와 같이, 문맥에 의해 명백히 달리 표시되지 않는 한, 단수형은 복수의 지시 대상을 포함한다.

[0058] 본원에서 사용되는 바와 같이, 구체적으로 언급되거나 문맥으로부터 명백하지 않는 한, 용어 "또는"은 포괄적인 것으로 이해되며 "또는" 및 "및" 둘 모두를 포함한다.

[0059] 본원에서 사용되는 용어 "예를 들어" 및 "즉"은 단지 예시로서 의도된 제한 없이 사용되며, 본 명세서에 명시적으로 열거된 항목들만을 지칭하는 것으로 해석되어서는 안 된다.

[0060] 본원에서 사용되는 바와 같이, 문맥으로부터 구체적으로 언급되거나 명백하지 않는 한, 용어 "약"은 당업계에서 정상적인 허용 오차의 범위 이내, 예를 들어 평균의 2개의 표준 편차 이내인 것으로 이해된다. "약"은 언급된 값의 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, 0.01%, 또는 0.001% 이내인 것으로 이해될 수 있다. 문맥으로부터 달리 명백하지 않는 한, 본원에 제공된 모든 수치는 당업자가 이해할 수 있는 정상적인 변동을 반영한다.

[0061] 본원에서 사용되는 바와 같이, 용어 "불현 전사체" 또는 "불현-전 전사체" 등은 DNA 템플릿에 의해 암호화된 전장 mRNA 분자보다 짧은 임의의 전사체이며, RNA 중합효소가 템플릿 DNA로부터 서열 독립적인 방식으로 조기 방출됨으로써 생성된다. 일부 구현예에서, 불현 전사체는 표적 DNA 분자로부터 전사되는 전장 mRNA 분자의 길이의 90% 미만, 예를 들어 전장 mRNA 분자의 길이의 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5%, 1% 미만일 수 있다.

[0062] 본원에서 사용되는 용어 "코돈" 및 "코돈들"은 유전자 코드의 단위를 함께 형성하는 3개의 뉴클레오티드의 서열을 지칭한다. 각각의 코돈은 번역 또는 단백질 합성의 과정에서의 특이적 아미노산 또는 정지 신호에 상응한다. 유전자 코드는 퇴화되고, 둘 이상의 코돈이 특이적 아미노산 잔기를 암호화할 수 있다. 예를 들어, 코돈은 DNA 또는 RNA 뉴클레오티드를 포함할 수 있다.

[0063] 본원에서 사용되는 용어 "코돈 최적화" 및 "코돈 최적화된"은 펩티드, 폴리펩티드, 또는 단백질을 암호화하는 자연 발생 또는 야생형 핵산의 코돈 조성물을 이의 아미노산 서열을 변경시키지 않고 변형시켜, 상기 핵산의 단백질 발현을 개선시키는 것을 지칭한다. 본 발명의 맥락에서, "코돈 최적화"는 또한 구아닌-시토신 함량, 코돈 적용 인덱스, 불안정화 핵산 서열 또는 모티프의 존재, 및/또는 일시 정지 부위 및/또는 종결자 신호의 존재와 같은 뉴클레오티드 서열의 목록으로부터 최적의 뉴클레오티드 서열보다 적은 필터로 제거함으로써 하나 이상의 최적화된 뉴클레오티드 서열이 도달하는 프로세스를 지칭할 수 있다.

[0064] 본원에서 사용되는 "전장 mRNA"는 모세관 전기 영동에 의해 구분된 특정 검정, 예를 들어, 겔 전기영동 및 UV를 사용하는 검출 및 UV 흡수 분광법을 사용할 때 특성화된 것과 같다. 전장 폴리펩티드를 암호화하는 mRNA 분자의 길이는 표적 DNA로부터 전사되는 전장 mRNA 분자의 길이의 적어도 50%, 예를 들어 표적 DNA로부터 전사되는 전장 mRNA 분자의 길이의 적어도 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.01%, 99.05%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%이다.

[0065] 본원에서 사용되는 용어 "시험관 내"는 다세포 유기체 내가 아니라 예컨대, 시험관 또는 반응 용기, 세포 배양 등과 같은 인공적인 환경에서 발생하는 사건을 말한다.

[0066] 본원에서 사용되는 용어 "생체 내"는 인간 및 비인간 동물과 같은 다세포 유기체 내에서 발생하는 사건을 말한다. 세포-기반 시스템의 맥락에서, 전술한 용어는 (예를 들어, 생체 외 시스템에 반대되는) 활세포 내에서 발생하는 사건을 지칭하도록 사용될 수 있다.

[0067] 본원에서 사용되는 용어 "전령 RNA(mRNA)"는 적어도 하나의 폴리펩티드를 암호화하는 폴리리보뉴클레오티드를 지칭한다. 본원에서 사용되는 mRNA는 변형 및 비변형 RNA 둘 모두를 망라한다. mRNA는 하나 이상의 코딩 및 비코딩 영역을 함유할 수 있다. mRNA는 천연 공급원으로부터 정제되고, 재조합 발현 시스템을 사용하여 생산되고 임의로 정제되고, 시험관 내 전사되거나, 화학적으로 합성될 수 있다. 적절한 경우, 예컨대, 화학적으로 합성된 분자의 경우, mRNA는 화학적으로 변형된 염기 또는 당, 골격 변형 등을 갖는 유사체와 같은 뉴클레오시드 유사체를 포함할 수 있다. mRNA 서열은 달리 표시하지 않는 한, 5'에서 3' 방향으로 제시된다.

[0068] 본원에서 사용되는 용어 "핵산"은 가장 넓은 의미로 폴리뉴클레오티드 사슬에 혼입되거나 혼입될 수 있는 임의의 화합물 및/또는 물질을 말한다. 일부 구현예에서, 핵산은 인산디에스테르 연결을 통해 폴리뉴클레오티드 사슬에 혼입되거나 혼입될 수 있는 화합물 및/또는 물질이다. 일부 구현예에서, "핵산"은 개별 핵산 잔기(예를 들어, 뉴클레오티드 및/또는 뉴클레오시드)를 지칭한다. 일부 구현예에서, "핵산"은 개별 핵산 잔기를 포함하는 폴리뉴클레오티드 사슬을 지칭한다. 일부 구현예에서, "핵산"은 RNA뿐만 아니라 단일 및/또는 이중 가닥 DNA 및/또는 cDNA를 망라한다. 또한, 용어 "핵산", "DNA", "RNA", 및/또는 유사한 용어는 핵산 유사체, 즉, 포스포디에스테르 골격 이외의 것을 갖는 유사체를 포함한다. 핵산 서열은 달리 표시하지 않는 한, 5'에서 3' 방향으로 제시된다.

[0069] 본원에서 사용되는 용어 "뉴클레오티드 서열"은, 가장 넓은 의미에서, 핵산 내의 핵염기의 순서를 지칭한다. 일부 구현예에서, "뉴클레오티드 서열"은 유전자 내의 개별 핵염기의 순서를 지칭한다. 일부 구현예에서, "뉴클레오티드 서열"은 단백질-코딩 유전자 내의 개별 핵염기의 순서를 지칭한다. 일부 구현예에서, "뉴클레오티드 서열"은 단일 및/또는 이중 가닥 DNA 및/또는 cDNA 내의 개별 핵염기의 순서를 지칭한다. 일부 구현예에서, "뉴클레오티드 서열"은 RNA 내의 개별 핵염기의 순서를 지칭한다. 일부 구현예에서, "뉴클레오티드 서열"은 mRNA 내의 개별 핵염기의 순서를 지칭한다. 특정 구현예에서, "뉴클레오티드 서열"은 RNA 또는 DNA의 단백질-코딩 서열 내의 개별 핵염기의 순서를 지칭한다. 뉴클레오티드 서열은 달리 표시하지 않는 한, 일반적으로 5'에서 3' 방향으로 제시된다.

[0070] 본원에서 사용되는 용어 "조기 종결"은 DNA 템플릿의 전체 길이가 전사되기 전에 전사가 종결되는 것을 지칭한다. 본원에서 사용되는 바와 같이, 조기 종결은 DNA 템플릿 내에 뉴클레오티드 서열 모티프(본원에서는 또한 단순히 "모티프"로도 지칭됨), 예를 들어 종결 신호의 존재에 의해 야기되고, 이는 전장 mRNA보다 짧은 mRNA 전사체("조기 종결된 전사체" 또는 "절단된 mRNA 전사체")를 생성한다. 종결 신호의 예는 본원에 기술된 것과 같은 대장균 rrnB 종결자 t1 신호(컨센서스 서열: ATCTGTT) 및 이의 변이체를 포함한다.

[0071] 본원에서 사용되는 용어 "템플릿 DNA"(또는 "DNA 템플릿")는 시험관 내 전사에 의해 합성될 mRNA 전사체를 암호화하는 핵산 서열을 포함하는 DNA 분자에 관한 것이다. 템플릿 DNA는, 템플릿 DNA에 의해 암호화된 mRNA 전사체를 생산하기 위해, 시험관 내 전사를 위한 템플릿으로서 사용한다. 템플릿 DNA는 시험관 내 전사에 필요한 모든 요소, 특히 원하는 mRNA 전사체를 암호화하는 DNA 서열에 작동 가능하게 연결된 DNA-의존성 RNA 중합효소, 예를 들어 T3, T7, 및 SP6 RNA 중합효소의 결합을 위한 프로모터 요소를 포함한다. 또한, 템플릿 DNA는, mRNA 전사체를 암호화하는 DNA 서열의 동일성을 예를 들어 PCR 또는 DNA 서열화에 의해 결정하기 위해, mRNA 전사체를 암호화하는 DNA 서열의 5' 및/또는 3'에 프라이머 결합 부위를 포함할 수 있다. 본 발명의 맥락에서 "템플릿 DNA"는 선형 또는 원형 DNA 분자일 수 있다. 본원에서 사용되는 용어 "템플릿 DNA"는 원하는 mRNA 전사체를 암호화하는 핵산 서열을 포함하는 DNA 벡터, 예컨대 플라스미드 DNA를 지칭할 수 있다.

[0072] 본원에서 사용되는 "예방하는"은 특정 감염, 질환, 장애 및/또는 병태의 하나 이상의 증상 또는 특징의 개시를 부분적으로 또는 완전히 억제하는 것을 지칭한다.

[0073] 본원에서 사용되는 "예방"은 특정 감염, 질환, 장애 및/또는 병태의 하나 이상의 증상 또는 특징의 개시를 부분적으로 또는 완전히 억제하는 것을 지칭한다.

[0074] 본원에서 사용되는 용어 "치료"는 부분적으로 또는 완전하게 특정 감염, 질환, 장애 및/또는 병태의 하나 이상의 증상 또는 특징을 경감시키고, 회복시키고, 개선시키고, 완화시키고, 이의 발병을 지연시키고, 이의 진행을 억제하고, 이의 중증도를 감소시키고/시키거나 이의 발생 빈도를 감소시키는 것을 지칭한다.

[0075] 본원에서 사용되는 용어 "면역원성 조성물"은 대상체에게 투여될 때 면역 반응을 유도하는 핵산 또는 단백질을 포함하는 조성물을 의미한다. 일부 구현예에서, "면역원성 조성물"은 핵산을 포함한다. 일부 구현예에서, 핵산은 mRNA이다. 일부 구현예에서, 핵산은 DNA이다. 용어 "면역원성 조성물" 및 "백신"은 본원에서 상호 교환적으로 사용되며, 따라서 동등한 의미를 갖는 것을 의미한다는 것을 이해해야 한다.

[0076] 2개의 뉴클레오티드(또는 아미노산) 서열 간의 서열 동일성 백분율은 2개의 서열의 정렬 후에 결정된다. 이러한 정렬 및 서열 동일성 백분율은 당업계에 공지된 소프트웨어 프로그램, 예를 들어, Current Protocols in Molecular Biology(F.M. Ausubel 등 편집, 1987) 부록 30의 7.7.18절에 기술된 내용을 사용하여 결정될 수 있다. 본 발명의 맥락에서, 정렬은 12의 갭 개방 페널티 및 2의 갭 연장 페널티, 62개의 BLOSUM 매트릭스를 갖는 아핀(affine) 갭 검색을 사용하는 스미스-워터만(Smith-Waterman) 상동성 검색 알고리즘에 의해 결정된다. 스미스-워터만 상동성 검색 알고리즘은 Smith & Waterman의 문헌[Adv. Appl. Math. 2: 482-489(1981)]에 개시되어

있다. 이어서, 2개의 뉴클레오티드(또는 아미노산) 서열 내의 동일한 위치에 위치한 각각의 뉴클레오티드(또는 아미노산) 간의 비교를 수행한다. 주어진 위치가 2개의 뉴클레오티드(또는 아미노산) 서열에서 동일한 뉴클레오티드(또는 아미노산)에 의해 점유되는 경우, 이들 서열은 이 위치에 대해 서로 동일하다. 이어서, 서열 동일성 백분율은, 비교가 이루어지는 뉴클레오티드(또는 아미노산) 서열 내의 뉴클레오티드(또는 아미노산)의 총 수에 대해, 각각의 뉴클레오티드(또는 아미노산)가 동일한 위치의 수로부터 결정된다.

[0077] 본원에 사용되는 모든 기술적 및 과학적 용어는 본 출원이 속하는 기술 분야의 당업자가 공통적으로 이해하고, 본 출원이 속하는 기술 분야에서 공통적으로 사용되는 것과 같은 의미를 가진다. 본 발명의 배경기술을 기술하고 그의 실행에 관한 추가적인 자세한 사항을 제공하도록 본원에서 참조된 출판물 및 기타 다른 참고물은 참조로서 본원에 포함된다.

[0078] **발명을 실시하기 위한 구체적인 내용**

[0079] 본 발명은 최적화된 뉴클레오티드 서열을 갖는 mRNA를 포함하는 백신의 제공을 통해, 감염성 질환의 효과적인 치료 또는 예방을 위한 단백질 항원을 암호화하는 최적화된 뉴클레오티드 서열을 생성할 필요성을 다룬다. 단백질 항원을 암호화하는 자연적으로 발생하는 뉴클레오티드 서열을 처리하여 적어도 하나의 최적화된 뉴클레오티드 서열을 생산하기 위한 방법이 제공된다. 최적화된 뉴클레오티드 서열은 자연적으로 발생하는 뉴클레오티드 서열과 연관된 단백질의 발현과 비교하여 암호화된 단백질 항원의 발현을 증가시키도록 설계된다. 코돈 최적화는 유전자 코드의 중복으로 인한, 암호화된 단백질 항원의 번역된 아미노산의 서열을 변경하지 않고 다양한 기준에 기초하여 단백질 코딩 뉴클레오티드 서열의 조성물을 변형시킬 수 있다.

[0080] mRNA 코돈의 사용과 동족 tRNA의 풍부함 사이의 불균형을 피하기 위해, 코돈 최적화는 숙주 세포에서 전달 RNA(tRNA)의 자연적으로 발생하는 풍부함과 보다 잘 일치하고 특이적 tRNA의 고갈을 피하는 뉴클레오티드 서열 내의 코돈의 조성물을 제공할 수 있다. tRNA 풍부함은 단백질 번역 속도에 영향을 미치므로, 뉴클레오티드 서열의 코돈 최적화는 단백질 번역의 효율 및 암호화된 단백질에 대한 수율을 증가시킬 수 있다. 희귀 tRNA의 부족은 단백질 번역을 중단시키거나 종료시킬 수 있기 때문에, 예를 들어, 낮은 코돈 사용을 특징으로 하는 희귀 코돈을 사용하지 않음으로써, 단백질 번역 및 단백질 수율의 효율이 증가될 수 있다.

[0081] 코돈 최적화는, 단백질의 번역을 제어하고 초기 폴리펩티드 사슬의 적절한 접힘을 보장하는 데 중요한 뉴클레오티드 서열 중의 암호화된 정보를 제거할 수 있기 때문에, 암호화된 단백질의 기능적 활성 감소 및 관련 효능 손실을 희생시킬 수 있다(Mauro & Chappell, Trends Mol Med. 2014; 20(11):604-13). 본 발명자는 일부 다양성을 유지하는 최적화된 서열, 즉 각각의 아미노산을 암호화하는 단 하나의 코돈을 반드시 포함할 필요가 없는 서열이, 암호화된 단백질의 기능적 활성을 유지하면서 증가된 단백질 수율을 달성할 수 있음을 발견하였다.

[0082] **최적화된 뉴클레오티드 서열의 생성**

[0083] 도 1a 및 도 1b는 본 발명에 따른, 최적화된 뉴클레오티드 서열을 생성하기 위한 프로세스를 도시한다. 프로세스는 먼저 코돈 최적화 서열의 목록을 생성한 다음, 해당 목록에 3개의 필터를 적용한다. 구체적으로, 이는 모티프 스크린 필터, 구아닌-시토신(GC) 함량 분석 필터, 및 코돈 적용 인덱스(CAI) 분석 필터를 적용하여 최적화된 뉴클레오티드 서열의 업데이트된 목록을 생성한다. 업데이트된 목록은 암호화된 단백질 항원의 효과적인 전사 및/또는 번역을 방해할 것으로 예상되는 특징을 함유하는 뉴클레오티드 서열을 더 이상 포함하지 않는다.

[0084] *코돈 최적화*

[0085] 유전자 코드는 64개의 가능한 코돈을 갖는다. 각각의 코돈은 3개의 뉴클레오티드의 서열을 포함한다. 게놈의 단백질 코딩 영역 내의 각 코돈에 대한 사용 빈도는, 특정 코돈이 게놈의 단백질 코딩 영역 내에 나타나는 경우의 수를 결정하고, 후속하여 수득된 값을 게놈의 단백질 코딩 영역 내에서 동일한 아미노산을 암호화하는 코돈의 총 수로 나눔으로써 계산될 수 있다.

[0086] 코돈 사용 테이블은, 테이블이 생성된 특정 생물학적 공급원에 대해, 얼마나 자주, 특정 아미노산을 암호화하기 위해 각 코돈이 사용되는지에 관한 실험적으로 도출된 데이터를 포함한다. 이러한 정보는, 코돈이 해당 아미노산을 암호화하는 총 횟수에 대해 특정 아미노산을 암호화하는 데 사용되는 빈도의 백분율(0 내지 100%), 또는 분율(0 내지 1)로서, 각 코돈에 대해 표현된다.

[0087] 코돈 사용 테이블은, 코돈 사용 데이터베이스(Codon Usage Database)(Nakamura 외 (2000) *Nucleic Acids Research* 28(1), 292; https://www.kazusa.or.jp/codon/에서 온라인으로 이용 가능함) 및 고성능 통합 가상 환경-코돈 사용 테이블(High-performance Integrated Virtual Environment-Codon Usage Tables)(HIVE-CUT) 데이

터베이스(Adie 등 (2017), *BMC Bioinformatic*s 18(1), 391; http://hive.biochemistry.gwu.edu/review/codon에서 온라인으로 이용 가능함)와 같은 공개적으로 이용 가능한 데이터베이스에 저장되어 있다.

[0088] 코돈 최적화의 제1 단계 동안, 코돈이 임계 빈도(예를 들어, 10%) 미만인 코돈 사용 빈도와 연관되는 경우, 주어진 유기체(예를 들어, 포유동물 또는 인간)에서의 각 코돈의 빈도를 반영하는 제1 코돈 사용 테이블로부터 해당 코돈이 제거된다. 제1 단계에서 제거되지 않은 코돈의 코돈 사용 빈도는 정규화되어 정규화된 코돈 사용 테이블을 생성한다. 관심 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열은 정규화된 코돈 사용 테이블에서 주어진 아미노산과 연관된 하나 이상의 코돈의 사용 빈도에 기초하여 아미노산 서열 내의 각 아미노산에 대한 코돈을 선택함으로써 생성된다. 주어진 아미노산에 대한 특정 코돈을 선택할 확률은 정규화된 코돈 사용 테이블에서의 이러한 아미노산과 연관된 코돈과 연관된 사용 빈도와 동일하다.

[0089] 본 발명의 코돈 최적화된 서열은 최적화된 뉴클레오티드 서열을 생성하기 위한 컴퓨터 구현 방법에 의해 생성된다. 방법은, (i) 아미노산 서열을 수령하는 단계(여기에서, 아미노산 서열은 펩티드, 폴리펩티드, 또는 단백질을 암호화); (ii) 제1 코돈 사용 테이블을 수령하는 단계(여기에서, 제1 코돈 사용 테이블은 아미노산의 목록을 포함하되, 테이블 내의 각각의 아미노산은 적어도 하나의 코돈과 연관되고, 각각의 코돈은 사용 빈도와 연관됨); (iii) 임계 빈도 미만인 사용 빈도와 연관된 임의의 코돈을 코돈 사용 테이블로부터 제거하는 단계; (iv) 단계 (iii)에서 제거되지 않은 코돈의 사용 빈도를 정규화함으로써 정규화된 코돈 사용 테이블을 생성하는 단계; 및 (v) 정규화된 코돈 사용 테이블 내의 아미노산과 연관된 하나 이상의 코돈의 사용 빈도에 기초하여 아미노산 서열 내의 각각의 아미노산에 대한 코돈을 선택함으로써 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 생성하는 단계를 포함한다. 임계 빈도는 5% 내지 30%의 범위, 특히 5%, 10%, 15%, 20%, 25%, 또는 30%일 수 있다. 본 발명의 맥락에서, 임계 빈도는 통상적으로 10%이다.

[0090] 정규화된 코돈 사용 테이블을 생성하는 단계는, (a) 제1 아미노산과 연관되고 단계 (iii)에서 제거된 각각의 코돈의 사용 빈도를 제1 아미노산과 연관된 나머지 코돈에 분배하는 단계; 및 (b) 각각의 아미노산에 대해 단계 (a)를 반복하여 정규화된 코돈 사용 테이블을 생성하는 단계를 포함한다. 일부 구현예에서, 제거된 코돈의 사용 빈도는 나머지 코돈 중에 균등하게 분포된다. 일부 구현예에서, 제거된 코돈의 사용 빈도는 나머지 코돈 중에 균등하게 분배된다. 이러한 맥락에서, "분배된"은 특정 아미노산과 연관된 제거된 코돈의 사용 빈도의 합쳐진 양을 취하고, 이러한 합쳐진 빈도의 일부를 특정 아미노산을 암호화하는 나머지 각각의 코돈에 배분하는 것으로 정의될 수 있다.

[0091] 각각의 아미노산에 대한 코돈을 선택하는 단계는, (a) 정규화된 코돈 사용 테이블에서, 아미노산 서열의 제1 아미노산과 연관된 하나 이상의 코돈을 식별하는 단계; (b) 제1 아미노산과 연관된 코돈을 선택하는 단계(여기에서, 특정 코돈을 선택할 확률은 정규화된 코돈 사용 테이블에서 제1 아미노산과 연관된 코돈과 연관된 사용 빈도와 동일함); 및 (c) 코돈이 아미노산 서열 내의 각각의 아미노산에 대해 선택될 때까지 단계 (a) 및 (b)를 반복하는 단계를 포함한다.

[0092] 아미노산 서열 내에서 각각의 아미노산에 대한 코돈을 선택하는 단계에 의해 최적화된 뉴클레오티드 서열을 생성하는 단계(전술한 방법 중 단계 (v))를 n회 수행하여 최적화된 뉴클레오티드 서열의 목록을 생성한다.

[0093] *모티프 스크린*

[0094] 모티프 스크린 필터는 최적화된 뉴클레오티드 서열의 목록에 적용된다. 임의의 공지된 음성 시스 조절 요소 및 음성 반복 요소를 암호화하는 최적화된 뉴클레오티드 서열은 목록으로부터 제거되어 업데이트된 목록을 생성한다.

[0095] 목록 내의 각각의 최적화된 뉴클레오티드 서열에 대해, 종결 신호를 함유하는지의 여부 또한 결정된다. 하나 이상의 종결 신호를 함유하는 임의의 뉴클레오티드 서열은 업데이트된 목록을 생성하는 목록으로부터 제거된다. 일부 구현예에서, 종결 신호는 다음의 뉴클레오티드 서열을 갖는다: 5'-$X_1$ATCT$X_2$T$X_3$-3'(여기에서, $X_1$, $X_2$, 및 $X_3$은 A, C, T, 또는 G로부터 독립적으로 선택됨). 일부 구현예에서, 종결 신호는 다음의 뉴클레오티드 서열 중 하나를 갖는다: TATCTGTT; 및/또는 TTTTTT; 및/또는 AAGCTT; 및/또는 GAAGAGC; 및/또는 TCTAGA. 일부 구현예에서, 종결 신호는 다음의 뉴클레오티드 서열을 갖는다: 5'-$X_1$AUCU$X_2$U$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, U 또는 G로부터 독립적으로 선택됨). 일부 구현예에서, 종결 신호는 다음의 뉴클레오티드 서열 중 하나를 갖는다: UAUCUGUU; 및/또는 UUUUUU; 및/또는 AAGCUU; 및/또는 GAAGAGC; 및/또는 UCUAGA.

[0096] *구아닌-시토신(GC) 함량*

[0097] 방법은, 최적화된 뉴클레오티드 서열의 업데이트된 목록에서 각각의 최적화된 뉴클레오티드의 구아닌-시토신(GC) 함량을 결정하는 단계를 추가로 포함한다. 서열의 GC 함량은 구아닌 또는 시토신인 뉴클레오티드 서열 내 염기의 백분율이다. 최적화된 뉴클레오티드 서열의 목록은, 그의 GC 함량이 사전에 결정된 GC 함량 범위를 벗어나는 경우, 해당 목록으로부터 임의의 뉴클레오티드 서열을 제거하는 단계에 의해 추가로 업데이트된다.

[0098] 각각의 최적화된 뉴클레오티드 서열의 GC 함량을 결정하는 단계는, 각각의 뉴클레오티드 서열에 대해, 뉴클레오티드 서열의 하나 이상의 추가 부분의 GC 함량을 결정하는 단계(여기에서, 추가 부분은 서로 중첩되지 않고 제1 부분과 중첩되지 않음)를 포함하며, 여기에서, 최적화된 서열의 목록을 업데이트하는 단계는, 임의의 부분의 GC 함량이 사전에 결정된 GC 함량 범위를 벗어나는 경우, 해당 뉴클레오티드 서열을 제거하는 단계를 포함하되, 선택적으로, 뉴클레오티드 서열의 GC 함량을 결정하는 단계는 임의의 부분의 GC 함량이 사전에 결정된 GC 함량 범위를 벗어나는 것으로 결정될 경우 중단된다. 일부 구현예에서, 뉴클레오티드 서열의 제1 부분 및/또는 하나 이상의 추가 부분은 사전에 결정된 수의 뉴클레오티드를 포함하되, 선택적으로, 사전에 결정된 수의 뉴클레오티드는 5 내지 300개의 뉴클레오티드, 또는 10 내지 200개의 뉴클레오티드, 또는 15 내지 100개의 뉴클레오티드, 또는 20 내지 50개의 뉴클레오티드의 범위이다. 본 발명의 맥락에서, 사전에 결정된 수의 뉴클레오티드는 통상적으로 30개의 뉴클레오티드이다. 사전에 결정된 GC 함량 범위는 15% 내지 75%, 또는 40% 내지 60%, 또는 30% 내지 70%일 수 있다. 본 발명의 맥락에서, 사전에 결정된 GC 함량 범위는 통상적으로 30% 내지 70%이다.

[0099] 본 발명의 맥락에서 적절한 GC 함량 필터는 먼저 최적화된 뉴클레오티드 서열의 첫번째 30개의 뉴클레오티드, 즉 최적화된 뉴클레오티드 서열의 뉴클레오티드 1 내지 30을 분석할 수 있다. 분석은, G 또는 C 중 어느 하나를 갖는 부분에서의 뉴클레오티드의 수를 결정하는 단계를 포함할 수 있고, 해당 부분의 GC 함량을 결정하는 단계는 해당 부분에서의 G 또는 C 뉴클레오티드의 수를 해당 부분에서의 뉴클레오티드의 총 수로 나누는 단계를 포함할 수 있다. 이러한 분석의 결과는, G 또는 C인 부분에서의 뉴클레오티드의 비율, 예를 들어, 50%와 같은 백분율, 또는 예를 들어, 0.5와 같은 소수일 수 있는 값을 제공할 것이다. 제1 부분의 GC 함량이 사전에 결정된 GC 함량 범위를 벗어나는 경우, 해당 최적화된 뉴클레오티드 서열은 최적화된 뉴클레오티드 서열의 목록에서 제거될 수 있다.

[0100] 제1 부분의 GC 함량이 사전에 결정된 GC 함량 범위 내에 속하는 경우, 이에 이어서 GC 함량 필터는 해당 최적화된 뉴클레오티드 서열의 제2 부분을 분석할 수 있다. 본 실시예에서, 이는 최적화된 뉴클레오티드 서열의 두번째 30개의 뉴클레오티드, 즉, 뉴클레오티드 31 내지 60일 수 있다. 부분의 분석은, GC 함량이 사전에 결정된 GC 함량 범위를 벗어나는 부분이 발견될 때까지 각 부분에 대해 반복될 수 있으며, 발견되는 경우 최적화된 뉴클레오티드 서열은 목록에서 제거될 수 있고, 또는 전체 최적화된 뉴클레오티드 서열이 분석되고 함량 범위를 벗어나는 부분이 발견되지 않은 경우, GC 함량 필터는 해당 목록에서의 최적화된 뉴클레오티드 서열을 유지하고 해당 목록에서의 그 다음 최적화된 뉴클레오티드 서열로 이동할 수 있다.

[0101] *코돈 적용 인덱스(CAI)*

[0102] 방법은, 최적화된 뉴클레오티드 서열의 가장 최근에 업데이트된 목록에서 각각의 최적화된 뉴클레오티드 서열의 코돈 적용 인덱스를 결정하는 단계를 추가로 포함한다. 서열의 코돈 적용 인덱스는 코돈 사용 편향의 척도이며, 0과 1 사이의 값일 수 있다. 최적화된 뉴클레오티드 서열의 가장 최근에 업데이트된 목록은, 이의 코돈 적용 인덱스가 사전에 결정된 코돈 적용 인덱스 임계값 이하인 경우, 임의의 뉴클레오티드 서열을 제거함으로써 추가로 업데이트된다. 코돈 적용 인덱스 임계값은 0.7, 또는 0.75, 또는 0.8, 또는 0.85, 또는 0.9일 수 있다. 본 발명자는 코돈 적용 인덱스가 0.8 이상인 최적화된 뉴클레오티드 서열이 매우 높은 단백질 수율을 야기한다는 것을 발견하였다. 따라서, 본 발명의 맥락에서, 코돈 적용 인덱스 임계값은 통상적으로 0.8이다.

[0103] 코돈 적용 인덱스는, 예를 들어, "*The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications*"((Sharp 및 Li, 1987 *Nucleic Acids Research* 15(3), p.1281-1295); https://www.ncbi.nlm.nih.gov/pmc/articles/PMC340524/에서 온라인 상으로 사용 가능함)에 기술된 바와 같이, 당업자에게 명백할 수 있는 임의의 방식으로, 각각의 최적화된 뉴클레오티드 서열에 대해 계산될 수 있다.

[0104] 코돈 적용 인덱스 계산을 구현하는 단계는 다음에 따라, 또는 다음과 유사한 방법을 포함할 수 있다. 서열 내의 각 아미노산에 대해, 서열 내의 각각의 코돈의 중량은 상대 적용도($w_i$)로 명명된 파라미터로 표현될 수 있다. 상대 적용도는, 코돈 $f_i$의 관찰된 빈도와 해당 아미노산에 대한 가장 빈번한 동종 코돈 $f_j$의 빈도 사이의 비율로서, 참조 서열 세트로부터 연산될 수 있다. 그런 다음, 서열의 코돈 적용 인덱스는 서열의 길이에 걸쳐

(코돈에서 측정된) 각 코돈에 연관된 중량의 기하 평균으로서 계산될 수 있다. 코돈 적용 인덱스를 계산하는 데 사용되는 참조 서열 세트는 본 발명의 방법과 함께 사용되는 코돈 사용 테이블이 유래되는 참조 서열 세트와 동일할 수 있다.

[0105] **최적화된 뉴클레오티드 서열의 합성**

[0106] 일단 최적화된 뉴클레오티드 서열의 목록이 생성되면, 시험관 내 합성(통상적으로 "시험관 내 전사"로도 지칭됨)은, 프로모터, 리보뉴클레오티드 삼인산염의 풀, DTT 및 마그네슘 이온을 포함할 수 있는 완충제 시스템, 및 적절한 RNA 중합효소(예를 들어, T3, T7, 또는 SP6 RNA 중합효소)와 같은 핵산 벡터, DNase I, 파이로포스파타아제 및/또는 RNase 억제제로 수행될 수 있다. 정확한 조건은 특정 응용예에 따라 달라질 것이다.

[0107] 핵산 벡터는 통상적으로 플라스미드이다. 용어 '플라스미드' 또는 '플라스미드 핵산 벡터'는 원형 핵산 분자, 예를 들어 인공 핵산 분자를 지칭한다. 본 발명의 맥락에서의 플라스미드 DNA는 원하는 핵산 서열, 예컨대 mRNA 전사체를 암호화하는 서열 및/또는 적어도 하나의 단백질 항원을 암호화하는 개방 해독 프레임을 포함하는 핵산 서열을 혼입하거나 보유하는 데 적합하다. 이러한 플라스미드 DNA 작제물/벡터는 발현 벡터, 클로닝 벡터, 전달 벡터 등일 수 있다.

[0108] 핵산 벡터는 일반적으로, 원하는 mRNA 전사체에 상응하는 (이를 암호화하는) 서열 또는 이의 일부분, 예컨대 단백질 항원을 암호화하는 최적화된 뉴클레오티드 서열 및 mRNA의 5'- 및/또는 3' UTR에 상응하는 서열을 포함한다. 일부 구현예에서, 원하는 mRNA 전사체에 상응하는 서열은 또한 3' UTR 뒤에 있는 폴리A-꼬리를 암호화하여 폴리A-꼬리를 mRNA 전사체와 함께 포함할 수 있다. 본 발명의 맥락에서 보다 일반적으로, 원하는 mRNA 전사체에 상응하는 서열은 5'/3' UTR 및 개방 해독 프레임으로 이루어진다. 본 발명의 일부 구현예에서, 시험관 내 전사 동안 핵산 벡터로부터 합성된 mRNA 전사체는 폴리A 꼬리를 함유하지 않는다. 폴리A 꼬리는 합성 후 프로세싱 단계에서 mRNA 전사체에 첨가될 수 있다.

[0109] **최적화된 뉴클레오티드 서열의 스크리닝**

[0110] 단백질 항원을 암호화하는 최적화된 뉴클레오티드 서열을 암호화하는 개별적인 시험관 내 전사되고, 캡핑되고 꼬리화된 mRNA는, 최적화된 뉴클레오티드 서열에 의해 암호화된 단백질의 발현 수준을 결정할 수 있도록 생체 내 또는 시험관 내에서 세포 내로 형질감염될 수 있다. mRNA 암호화, 예를 들어, 단백질 항원을 암호화하는 자연 발생 뉴클레오티드 서열, 또는 본원에 기술된 최적화된 뉴클레오티드 서열을 생성하기 위한 프로세스 이외의 방법으로 제조된 단백질 항원을 암호화하는 코돈 최적화된 뉴클레오티드 서열은 대조군 mRNA로서 기능할 수 있다. 각각의 mRNA 및 대조군 mRNA는 별도의 세포 또는 유기체와 접촉되며, 여기에서 해당 세포 또는 유기체는 접촉한다. 본 발명에 따라 생성된 최적화된 뉴클레오티드 서열을 포함하는 mRNA는, 대조군 mRNA와 접촉된 세포 또는 유기체에 의해 생성된 단백질의 수율과 비교하여 단백질 항원의 증가된 수율을 생성하는 경우, 본 발명에 따른 면역원성 조성물에 사용하기 위해 선택된다.

[0111] 웨스턴 블롯팅과 같은 당업계에 공지된 방법은 최적화된 뉴클레오티드 서열이 암호화된 단백질 항원의 발현 및 생성을 증가시킨다는 것을 실험적으로 검증하기에 적합하다. 또한, 본 발명의 방법에 의해 생성된 다수의 최적화된 뉴클레오티드 서열을 스크리닝하여 가장 높은 단백질 수율을 생성하는 서열 또는 서열을 식별할 수 있다. 일부 구현예에서, 최적화된 뉴클레오티드 서열에 의해 암호화된 단백질의 발현 수준은 적어도 2배, 예를 들어 적어도 3배 또는 4배 증가된다.

[0112] 일부 구현예에서, 최적화된 뉴클레오티드 서열에 의해 암호화된 단백질 항원의 기능적 활성이 결정된다. 최적화된 뉴클레오티드 서열에 의해 암호화된 단백질의 기능적 활성은 양호하게 확립된 방법의 범위를 사용하여 결정될 수 있다. 이들 방법은 암호화된 단백질 항원의 특성에 따라 달라질 수 있다. 예를 들어, 단백질 항원 상의 형태적 에피토프를 인식하는 항체는 최적화된 뉴클레오티드 서열로부터 발현된 단백질 항원의 적절한 접힘을 확인하는 데 사용될 수 있다. 대안적으로 또는 추가적으로, SARS-CoV-2의 스파이크 단백질과 관련된 본 발명의 구현예에서, 스파이크 단백질은 인간 안지오텐신 전환 효소 2(ACE2)와 접촉되어 이의 수용체 결합 활성을 확인할 수 있다. 결합 활성은 일반적으로 자연 발생 코딩 서열로부터 발현된 SARS-CoV-2의 스파이크 단백질과 같은 대조군에 대해 상대적으로 평가된다.

[0113] **SARS-CoV-2 단백질**

[0114] 코로나바이러스(CoV)는 코로나바이러스(*Coronaviridae*), 아테리바이러스(*Arteriviridae*) 및 로니바이러스(*Nidovirales*) 계열을 포함하는니도바이러스(*Roniviridae*) 목에 속하는 가장 큰 바이러스 군이다. CoV는 양성-감지 단일 가닥 RNA 게놈 및 대략 125 nm의 직경을 갖는 나선형 대칭의 헥캡시드를 갖는 구형 외피형 바이러스

이다.

[0115] SARS-CoV-2는 메르스-CoV 및 SARS-CoV와 같은 인간을 감염시키는 다른 코로나바이러스와 마찬가지로 β-코로나바이러스이다. 주로 ORF1a/b 영역으로 명명된 바이러스 30kb RNA 게놈의 처음 2/3는 주요 비구조 단백질을 구성하는 2개의 폴리단백질(pp1a 및 pp1ab)을 암호화한다. 나머지 게놈은 부속 단백질 및 4개의 필수 구조 단백질, 즉 스파이크(S) 당단백질, 작은 외피(E) 단백질, 매트릭스/막(M) 단백질, 및 핵캡시드(N) 단백질을 암호화한다(Kang 등 (2020) https://doi.org/10.1101/2020.03.06.977876). SARS-CoV-2는 숙주 세포 수용체(인간 내 ACE 2)에 결합하고 세포 진입을 매개하기 위해 이의 S 단백질을 사용한다. 이는, 아래에서 상세히 논의되는 바와 같이, S 단백질을 중화 항체의 주요 표적으로 만든다.

[0116] *스파이크 당단백질(S 단백질)*

[0117] 세포 진입은 세포 표면 상의 수용체에 대한 S 단백질의 결합 및 숙주 세포 프로테아제에 의한 S 단백질 프라이밍에 의존한다. S 단백질은 숙주 세포 수용체(S1 서브유닛)에 대한 결합 및 바이러스 및 세포 막(S2 서브유닛)의 융합을 담당하는 2개의 기능적 서브유닛을 포함한다(도 3). S 단백질은 바이러스의 표면 상에 독특한 스파이크 구조를 생성하는 동종이량체를 형성한다. S1 서브유닛은 큰 수용체 결합 도메인(RBD)을 갖는 반면, S2는 스파이크 분자의 줄기를 형성한다. 전장 SARS-CoV-2 S 당단백질의 아미노산 서열은 서열번호 1(Gen Bank QHD43416.1)에 의해 제공된다. S1 서브유닛은 잔기 1 내지 681에 위치하고, S2 서브유닛은 잔기 686 내지 1208에 위치하며, S2' 서브유닛은 잔기 816 내지 1208에 위치한다. S 단백질의 C-말단부는 막관통 도메인을 함유하고, 세포질 꼬리의 마지막 19개의 아미노산은 소포체(ER)-보유 신호를 함유한다.

[0118] 자연 발생 SARS-CoV-2 S 단백질에 대한 기준은 서열번호 1에 의해 제공되는 전장 SARS-CoV-2 S 당단백질을 참조한다. 자연 발생 SARS-CoV-2 S 단백질에 대한 임의의 변형은 서열번호 1의 잔기에 기초하여 번호가 매겨진다.

[0119] 팬데믹 SARS-CoV-2 서열 간의 관찰된 다양성은 낮지만, 이의 급속한 글로벌 확산은 이 바이러스에 희귀하지만 유리한 돌연변이에 작용할 자연선택에 대한 충분한 기회를 제공한다. 단지 우한으로부터의 인덱스 계열(즉, 서열번호 1)보다는 유행하는 SARS-CoV-2 바이러스의 서열을 표적화하는 것이 유리하다.

[0120] SARS-CoV-2 S 당단백질의 아미노산 변화인 D614G는 2020년 COVID-19 팬데믹 기간의 조기에 출현했으며, 2020년 7월 현재 전 세계적으로 가장 흔한 형태의 바이러스가 되었다. G614에 감염된 환자는 D614에 감염된 환자에 비해 보다 많은 바이러스 핵산을 배출하고, G614-보유 바이러스는 시험관 내에서 그의 D614 대응체보다 유의하게 보다 높은 감염성 역가를 나타낸다(Korber 등, 2020, Cell 182, 1-16). 따라서, D614G 돌연변이를 포함하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열은 본원에 기술된 바와 같은 면역원성 조성물에 사용하기에 특히 적합할 수 있다.

[0121] SARS-CoV-2 S 단백질에서 식별된 다른 희귀 돌연변이는 아래의 표에 요약되어 있다(Korber 등, 2020 - https://doi.org/10.1101/2020.04.29.069054):

| 스파이크 돌연변이 | 영향을 미칠 수 있는 스파이크 위치 |
|---|---|
| L5F | 신호 펩티드 |
| L8V/W | 신호 펩티드 |
| H49Y | S1 NTD 도메인 |
| Y145H/del | S1 NTD 도메인 |
| Q239K | S1 NTD 도메인 |
| V367F | 업/다운 형태 |
| G476S | 직접적으로 RBD 중 |
| V483A | 업/다운 형태 |
| V6151/F | SARS-CoV ADE 에피토프 중 |
| A831V | S2 중 잠재적 융합 펩티드 |
| D839Y/N/E | S2 서브유닛 |
| S943P | HR1 의 융합 코어 |
| P1263L | 세포질 꼬리 |

[0122]

[0123] 추가의 SARS-CoV-2 S 당단백질 돌연변이는 다음을 포함한다: L18F, HV 69-70 결실, Y144 결실, E154Q, Q218E,

A222V, S447N, F490S, S494P, N501Y, A570D, E583D, T618E, P681H, A701V, T716I, T723I, I843V, S982A 및 D1118H. 2020년 말, 새로운 SARS-CoV-2 변이체가 다수의 돌연변이를 포함하며 UK, 남아공, 브라질 및 캘리포니아에서 출현하였다. UK 변이체에서의 SARS-CoV-2 S 당단백질에 존재하는 돌연변이(계통 B.1.1.7로 명명됨)는 H69 결실(H69), V70 결실(V70), Y144 결실(Y144), N501Y, A570D, P681H, T716I, S982A 및 D1118H을 포함한다(Rambaut 등, 2020 https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563). 2020년 10월, 남아공 변이체(계통 B.1.351로 명명됨)는 SARS-CoV-2 S 당단백질 단백질 중 6개의 돌연변이 - D80A, K417N, E484K, N501Y, D614G 및 A701V를 포함한다. 11월 말까지, 3개의 추가적인 SARS-CoV-2 S 당단백질 돌연변이(L18F, R246I 및 K417N)가 출현하였고, 이는 L242(ΔL242), A243(ΔA243) 및 L244(ΔL244)에서의 3개의 아미노산의 결실을 가졌다(Tegally 등 (2020) https://doi.org/10.1101/2020.12.21.20248640). 브라질 변이체(계통 P.1로 명명됨)의 SARS-CoV-2 S 당단백질에 존재하는 돌연변이는 L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I 및 V1176F를 포함한다. 캘리포니아 변이체(CAL.20C로 알려짐)의 SARS-CoV-2 S 당단백질에 존재하는 돌연변이는 S13I, W152C 및 L452R을 포함한다(Zhang 등 (2021) https://doi.org/10.1101/2021.01.18.21249786).

[0124] 일부 구현예에서, 전장 SARS-CoV-2 S 당단백질의 아미노산 서열은 다수의 돌연변이를 가질 수 있다. 예를 들어, 서열번호 1의 아미노산 서열에 대해 2개 이상, 3개 이상, 4개 이상, 5개 이상, 6개 이상, 7개 이상, 8개 이상, 9개 이상, 또는 10개 이상의 돌연변이를 가질 수 있다. SARS-CoV-2 S 당단백질에서의 돌연변이는 아미노산 결실 또는 아미노산 치환일 수 있다. 돌연변이의 가능한 조합은 다음을 포함한다: (a) L18F, A222V, D614G; (b) A222V, D614G; (c) A222V, E583D, D614G; (d) S447N, D614G; (e) E154Q, F490S, D614G, I834V; (f) D614G, A701V; (g) Q218E, D614G; (h) D614G, T618R; (i) ΔL242, ΔA243, ΔL244; (j) A222V, E583D, A701V; (k) ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H (UK 변이체 + D614G); (l) D80A, K417N, E484K, N501Y, D614G 및 A701V(남아공 고정 돌연변이 + D614G); (m) D80A, K417N, E484K, N501Y 및 A701V(남아공 고정 돌연변이); (n) D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G, A701V(남아공 변이체 1 + D614G); (o) L18F , D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V(남아공 변이체 2 + D614G); (p) L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F(브라질 변이체 + D614G) 및 (q) S13I, W152C, L452R 및 D614G(캘리포니아 변이체 + D614G).

[0125] 일부 구현예에서, 전장 SARS-CoV-2 S 당단백질의 아미노산 서열은 서열번호 1의 아미노산 서열에 대한 돌연변이 중 하나 이상을 가질 수 있다. 이는 다음의 돌연변이 중 하나 이상을 포함할 수 있다: D614G 돌연변이, L5F 돌연변이, L8V/W 돌연변이, H49Y 돌연변이, Y145H/del 돌연변이, Q239K 돌연변이, V367F 돌연변이, G476S 돌연변이, V483A 돌연변이, V6151/F 돌연변이, A831V 돌연변이, D839Y/N/E 돌연변이, S943P 돌연변이, P1263L 돌연변이. 따라서, 특정 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 D614G 돌연변이를 포함한다. 예를 들어, 특정 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 D614G 돌연변이를 포함한다.

[0126] 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 L5F 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 L8V/W 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 H49Y 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 Y145H/del 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 Q239K 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 V367F 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 G476S 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 V483A 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 V6151/F 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 A831V 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 D839Y/N/E 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 S943P 돌연변이를 포함한다. 일부 구현예에서, 본원에 기술된 S 단백질 중 어느 하나 또는 이의 항원성 단편은 P1263L 돌연변이를 포함한다.

[0127] 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 전장 SARS-CoV-2 S 단백질을 암호화한다. 전장 SARS-

CoV-2 S 단백질은 서열번호 1을 포함하는 아미노산 서열 또는 서열번호 1과 적어도 95%, 96%, 97%, 98%, 또는 99% 동일한 아미노산 서열을 가질 수 있다. 일부 구현예에서, 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 29의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 29와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 1의 아미노산 서열을 암호화한다.

[0128] 일부 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 1의 아미노산 서열에 대해 하나 이상의 돌연변이를 포함하는 SARS-CoV-2 S 단백질을 암호화한다. 예를 들어, 일부 구현예에서, 최적화된 뉴클레오티드 서열은 다음의 돌연변이 중 하나 이상을 포함하는 SARS-CoV2 S 단백질을 암호화한다: D614G 돌연변이, L5F 돌연변이, L8V/W 돌연변이, H49Y 돌연변이, Y145H/del 돌연변이, Q239K 돌연변이, V367F 돌연변이, G476S 돌연변이, V483A 돌연변이, V6151/F 돌연변이, A831V 돌연변이, D839Y/N/E 돌연변이, S943P 돌연변이, P1263L 돌연변이. 따라서, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 D614G 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 예를 들어, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 D614G 돌연변이를 포함하는 SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편을 암호화한다.

[0129] 일부 구현예에서, 최적화된 뉴클레오티드 서열은 L5F 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 L8V/W 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 H49Y 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 Y145H/del 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 Q239K 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 V367F 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 G467S 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 V483A 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 V6151/F 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 A831V 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 D839Y/N/E 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 S943P 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 P1263L 돌연변이를 포함하는 SARS-CoV2 S 단백질을 암호화한다.

[0130] 대안적으로, 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질의 항원성 단편을 암호화할 수 있다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 2의 아미노산 서열 또는 서열번호 2와 적어도 95%, 96%, 97%, 98%, 또는 99% 동일한 아미노산 서열을 가질 수 있는, SARS-CoV-2 S 단백질의 엑토도메인을 암호화할 수 있다. 엑토도메인은 막관통 도메인 및 세포질 꼬리를 포함하는 전장 SARS-CoV-2 S 단백질의 잔기 1209 내지 1273을 함유하지 않는다. 일부 구현예에서, SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 30의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 30과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 2의 아미노산 서열을 암호화한다.

[0131] 다른 구현예에서, SARS-CoV-2 S 단백질의 항원성 단편은 SARS-CoV-2 S 단백질의 S1 서브유닛, S2 서브유닛 및/또는 S2' 서브유닛 중 하나 이상을 포함할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 3의 아미노산 서열을 갖는 S1 서브유닛을 암호화할 수 있다. 따라서, 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 3을 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2 S 단백질의 S1 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 31의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 31과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 3의 아미노산 서열을 암호화한다. 대안적인 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 4의 아미노산 서열을 갖는 S2 서브유닛을 암호화할 수 있다. 따라서, 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 4를 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 32의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 32와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 4의 아미노산 서열을 암호화한다. 대안적인 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 5의 아미노산 서열을 갖는 S2' 서브유닛을 암호화할 수 있다. 따라서, 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 5를 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2

S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 33의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 33과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 5의 아미노산 서열을 암호화한다.

[0132] 일부 구현예에서, SARS-CoV-2 S 단백질의 항원성 단편은 SARS-CoV-2 S 단백질의 전장 S2 서브유닛 또는 S2' 서브유닛을 포함할 수 있다. 전장 S2 서브유닛 또는 S2' 서브유닛은 막관통 도메인 및 세포질 꼬리를 포함한다. 전장 S2 서브유닛은 SARS-CoV-2 S 단백질의 잔기 686 내지 1273을 포함하고, S2' 서브유닛은 SARS-CoV-2 S 단백질의 잔기 816 내지 1273을 포함한다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 72의 아미노산 서열을 갖는 전장 S2 서브유닛을 암호화할 수 있다. 따라서, 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 72를 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2 S 단백질의 전장 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 71의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 71과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 72의 아미노산 서열을 암호화한다. 대안적인 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 98의 아미노산 서열을 갖는 전장 S2' 서브유닛을 암호화할 수 있다. 따라서, 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 98을 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 97의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 적어도 81개이며 서열번호 98의 아미노산 서열을 암호화한다.

[0133] SARS-CoV-2 S 단백질은 S1 서브유닛에 위치한 수용체 결합 도메인(RBD)을 통해 안지오텐신 전환 효소 2(ACE2) 수용체에 먼저 결합하고, 이어서 S2 서브유닛을 통해 바이러스 및 숙주 막을 융합함으로써 숙주 세포 내로의 바이러스 진입을 매개한다(Tai 등 (2020) Cellular and Molecular immunology, doi.org/10.1038/s41423-020-0400-4). Tai 등은 S 단백질의 잔기 331 내지 524에서의 SARS-CoV-2의 RBD의 영역을 식별하였다. SARS-CoV-2 S 단백질의 잔기 331 내지 521로부터의 추정 RBD는 아래의 표 2의 서열번호 6으로 제공된다. SARS-CoV의 193-아미노산 RBD(잔기 318-510) 및 인간 IgG1 Fc 단편을 함유하는 재조합 융합 단백질은 이로 면역화된 토끼에서 매우 강력한 항체 반응을 유도하는 것으로 나타났다(He 등 (2004) *Biochem Biophys Res Commun*; 324(2): 773-781.). 따라서, SARS-CoV-2 S 단백질의 RBD는 또한 항체 반응을 고도로 유도할 수도 있다. SARS-CoV의 RBD 및 SARS-CoV-2의 RBD 둘 모두는 ACE2에 결합한다. 따라서, SARS-CoV-2 S 단백질의 항원 단편은 RBD를 암호화할 수 있을 것으로 고려된다. 이에 따라, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 6의 아미노산 서열을 갖는, SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, SARS-CoV-2 S 단백질의 RBD를 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 34의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 34와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 6의 아미노산 서열을 암호화한다.

[0134] 특정 구현예에서, SARS-CoV-2 S 단백질의 항원성 단편은 외인성 N-말단 신호 펩티드와 융합된다. 신호 펩티드는 단백질을 ER 및 분비 경로로 표적화하여, 단백질이 발현되는 숙주 세포의 분비 경로에 진입한다. 특정 구현예에서, 본 발명은 N-말단 신호 펩티드와 작동가능하게 연결된 SARS-CoV-2 S 단백질의 항원성 단편을 제공한다. 예를 들어, SARS-CoV-2 S 단백질의 RBD는 N-말단 신호 펩티드에 작동가능하게 연결될 수 있으며, 이는 생성된 단백질이 이를 발현하는 숙주 세포로부터 분비될 수 있게 한다.

[0135] 특정 구현예에서, N-말단 신호 펩티드는 자연 발생 SARS-CoV-2 S 단백질의 천연 신호 펩티드인 서열 MFVFLVLLPLVSSQC(서열번호 7)를 가질 수 있다. 일부 구현예에서, 신호 펩티드는 뉴클레오티드 서열 ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAGTGT(서열번호 37)에 의해 암호화된다. 다수의 다른 신호 펩티드가 당업계에 공지되어 있으며, 이는 숙주 세포, 예를 들어, Freudl의 문헌[*Microbial Cell Factories* 17: 52(2018)]에서의 리뷰에서 언급된 것들로부터 단백질을 분비하는 데 사용될 수 있다. 본 발명의 일부로서 사용될 수 있는 대안적인 신호 펩티드는 MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLS(서열번호 38)이다. 일부 구현예에서, 신호 펩티드는 뉴클레오티드 서열에 의해 암호화된다.

[0136] AUGGCCACUGGAUCAAGAACCUCACUGCUGCUCGCUUUUGGACUGCUUUGCCUGCCCUGGUUGCAAGAAGGAUCGGCUUUCCCGACCAUCCCACUCUCC (서열번호 39).

[0137] 본 발명의 일부로서 사용될 수 있는 또 다른 신호 펩티드는 MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLS(서열번호 40)이다. 일부 구현예에서, 신호 펩티드는 뉴클레오티드 서열에 의해 암호화된다.

[0138] AUGGCAACUGGAUCAAGAACCUCCCUCCUGCUCGCAUUCGGCCUGCUCUGUCUCCCAUGGCUCCAAGAAGGAAGCGCGUUCCCCACUAUCCCCCUCUCG (서열번호 41).

[0139] SARS-CoV-2 게놈의 원래의 주석은 SARS-CoV-2 S 단백질의 신호 펩티드 서열을 서열번호 7로 식별하였다. SARS-CoV-2 게놈의 대안적인 주석은 보다 긴 천연 N-말단 신호 펩티드 서열인 MFLLTTKRTMFVFLVLLPLVSSQC(서열번호 142)를 식별하였으며, 이는 9개의 아미노산만큼 더 길다. 특정 구현예에서, N-말단 신호 펩티드는 서열번호 142의 서열을 가질 수 있다. 일부 구현예에서, 신호 펩티드는 뉴클레오티드 서열 ATGTTCCTGCTGACAACAAAAAGAACCATGTTTGTGTTCCTGGTGCTGCTGCCTCTGGTGTCCTCACAGTGT(서열번호 143)에 의해 암호화된다.

[0140] 특정 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 8을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화할 수 있다. 일 구현예에서, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 35의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 35와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 8의 아미노산 서열을 암호화한다.

[0141] 특정 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 74를 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화할 수 있다. 일 구현예에서, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 73의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 73과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 74의 아미노산 서열을 암호화한다.

[0142] 특정 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 66을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화할 수 있다. 일 구현예에서, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 65의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 65와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 66의 아미노산 서열을 암호화한다.

[0143] 특정 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 68을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 전장 S2 서브유닛을 암호화할 수 있다. 일 구현예에서, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 전장 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 67의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 67과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 68의 아미노산 서열을 암호화한다. 특정 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 96을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화할 수 있다. 일 구현예에서, N-말단 신호 펩티드에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 95의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 95와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 96의 아미노산 서열을 암호화한다.

[0144] CoV S 단백질은 일반적인 클래스 I 바이러스 융합 단백질이며, 이는 S 단백질의 융합 가능성의 활성화를 위해 프로테아제 절단을 필요로 한다. 2단계 순차적 프로테아제 절단 모델은 SARS-CoV-2 S 단백질의 S 단백질의 활성화, (1) S1과 S2 서브유닛 사이의 프라이밍 절단 및 (2) S2' 부위 상에서의 활성화 절단에 대해 제안된 바 있다(Ou 등 (2020) *Nature communications*, 11, 1620). SARS-CoV-2 S 단백질은, 이 바이러스를 SARS-CoV 및 SARS-관련 CoV와 구별시키고, 이러한 바이오생성 동안 가공되는, S1/S2 서브유닛 사이의 경계에서의 퓨린 절단 부위를 보유한다(Walls 등 (2020) Cell doi.org/10.1016/j.cell.2020.02.058).

[0145] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되었으며, 확장된 N-말단 신호 펩티드를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 123을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 122의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 122와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, 또는 99% 동일하고, 서열번호 123의 아미노산 서열을 암호화한다.

[0146] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되었으며, 확장된 N-말단 신호 펩티드를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 137을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 136의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 136과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 137의 아미노산 서열을 암호화한다.

[0147] 융합-전 안정화는, 아마도 이러한 단백질이 보다 안정적인 융합-후 구조를 채택하는 경향으로부터 기인하는 접힘 오류를 방지함으로써, 바이러스 융합 당단백질의 재조합 발현을 증가시키는 경향이 있다. 융합-전 안정화된 바이러스 당단백질은 이의 야생형 대응체에 비해 우월한 면역원으로 간주된다.

[0148] S1 및 S2 서브유닛의 절단을 방지하기 위해 퓨린 절단 부위를 돌연변이시킴으로써, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태가 생성될 수 있다. 예를 들어, 퓨린 절단 부위(위치 682 내지 685)에서의 RRAR 잔기는 GSAS 잔기(즉, R682G R683S A684A R685S)로 돌연변이될 수 있다. 따라서, 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위의 제거, 즉 퓨린에 의해 인식되는 아미노산 잔기를 퓨린 절단 부위를 형성하지 않지만 S 단백질의 구조를 유지하는 대체 아미노산으로 대체함으로써, 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 특정 구현예에서, RRAR 퓨린 절단 부위 잔기 682 내지 685는 퓨린 절단 부위를 제거하기 위해 잔기 GSAS로 돌연변이될 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 9를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열은 서열번호 42의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 42와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 9의 아미노산 서열을 암호화한다.

[0149] SARS-CoV-2 S 단백질은 프롤린으로 잔기 985, 986 및 987(즉, D985P) 중 하나 이상을 치환함으로써 그의 융합-전 형태로 안정화될 수 있다. 예를 들어, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태는, 잔기 985(즉, D985P)에서의 하나의 안정화 프롤린 돌연변이; 잔기 986 및 987(즉, K986P, V987P)에서의 2개의 안정화 프롤린 돌연변이; 또는 잔기 985, 986 및 987(즉, D985P, K986P, V987P)에서의 3개의 안정화 프롤린 돌연변이를 만듦으로써 생성될 수 있다.

[0150] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 10을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 43의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 43과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 10의 아미노산 서열을 암호화한다. 추가의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 118을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 이러한 아미노산 서열은 D614G 돌연변이를 포함한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 119의 서열을 갖는다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 119와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 118의 아미노산 서열을 암호화한다.

[0151] 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 78을 포함하는 아미노산 서열을 갖는, SARS-CoV-2 S 단백질의 S2 서브유닛의 융합-전 안정화된 변이체를 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 77의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 77과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 78의 아미노산 서열을 암호화한다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 70을 포함하는 아미노산 서열을 갖는, SARS-CoV-2 S 단백질의 전장 S2 서브유닛의 융합-전 안정화된 변이체를 암호화할 수 있다. 일

구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 69의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 69와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 70의 아미노산 서열을 암호화한다.

[0152] 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 82를 포함하는 아미노산 서열을 갖는, SARS-CoV-2 S 단백질의 S2' 서브유닛의 융합-전 안정화된 변이체를 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 81의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 81과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 82의 아미노산 서열을 암호화한다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 86을 포함하는 아미노산 서열을 갖는, SARS-CoV-2 S 단백질의 전장 S2' 서브유닛의 융합-전 안정화된 변이체를 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 85의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 85와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 86의 아미노산 서열을 암호화한다.

[0153] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 985를 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 88을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 87의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 87과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 88의 아미노산 서열을 암호화한다.

[0154] 일부 구현예에서, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태는 잔기 985, 986 및 987(즉, D985P, K986P, V987P)에서 S2 서브유닛의 C-말단에 3개의 안정화 프롤린 돌연변이를 만듦으로써 생성될 수 있다. 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 985, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 92를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 91의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 91과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 92의 아미노산 서열을 암호화한다.

[0155] 일부 구현예에서, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태는, S1 및 S2 서브유닛의 절단을 방지하기 위해 퓨린 절단 부위를 돌연변이시킴으로써 생성될 수 있고, (a) 잔기 986 및 987(즉, K986P, V987P)에서의 2개의 안정화 프롤린 돌연변이를 만들고/만들거나; (b) 잔기 985에서의 안정화 프롤린 돌연변이를 만듦으로써 생성될 수 있다. 예를 들어, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 일부 구현예에서, 잔기 682 내지 685에서 퓨린 절단 부위를 형성하는 잔기는 잔기 GSAS로 돌연변이된다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 11을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 44의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 44와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 11의 아미노산 서열을 암호화한다. 추가의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 120을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 이러한 아미노산 서열은 D614G 돌연변이를 포함한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 121의 서열을 갖는다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 121과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 120의 아미노산 서열을 암호화한다. 대안적으로, 최적화된 뉴클레오티드 서열은 서열 번호 12를 포함하는 아미노산 서열을 갖는, 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 45의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 45와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 12의 아미노산 서열을 암호화한다.

[0156] 소정의 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하

고 잔기 985를 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 일부 구현예에서, 잔기 682 내지 685에서 퓨린 절단 부위를 형성하는 잔기는 잔기 GSAS로 돌연변이된다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 90을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 89의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 89와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 90의 아미노산 서열을 암호화한다.

[0157] 소정의 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 985, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 일부 구현예에서, 잔기 682 내지 685에서 퓨린 절단 부위를 형성하는 잔기는 잔기 GSAS로 돌연변이된다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 94를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 93의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 93과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 94의 아미노산 서열을 암호화한다.

[0158] SARS-CoV-2 S 단백질은 프롤린으로 잔기 817, 892, 899 및 942(즉, F817P, A892P, A899P 및 A942P) 중 하나 이상을 치환함으로써 그의 융합-전 형태로 추가로 안정화될 수 있다. 예를 들어, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태는, 잔기 817(즉, F817P)에서의 하나의 안정화 프롤린 돌연변이; 잔기 817 및 892(즉, F817P, A892P)에서의 2개의 안정화 프롤린 돌연변이; 또는 817, 892, 899(즉, F817P, A892P, A899P)에서의 3개의 안정화 프롤린 돌연변이; 또는 817, 892, 899 및 942(즉, F817P, A892P, A899P, A942P)에서의 4개의 안정화 프롤린 돌연변이를 만듦으로써 생성될 수 있다. 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 817, 892, 899 및 942를 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다.

[0159] 바람직한 구현예에서, SARS-CoV-2 S 단백질의 융합-전 안정화된 형태는 잔기 817, 892, 899, 942, 986에서 안정화 프롤린 돌연변이를 만듦으로써 생성될 수 있다. 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 129를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 128의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 128과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 129의 아미노산 서열을 암호화한다.

[0160] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 131을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 130의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 130과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 131의 아미노산 서열을 암호화한다.

[0161] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되고, D614G 돌연변이를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 133을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 132의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호

132와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 133의 아미노산 서열을 암호화한다.

[0162] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되고, D614G 돌연변이를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 135를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 134의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 134와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 135의 아미노산 서열을 암호화한다.

[0163] T4 박테리오파지 피브리틴 폴돈(Foldon)은 삼량체 형성을 유도하는 것을 돕기 위해 SARS-CoV-2 S 단백질의 항원성 단편의 C-말단에 위치할 수 있다. 폴돈은 인플루엔자 백신에 사용하기 위한 삼량체 인플루엔자 헤마글루티닌 줄기 도메인을 생성하는 데 사용된다(Lu 등 (2014) *PNAS*, 111, 1, 124-130). 폴돈은 GYIPEAPRDGQAYVRKDGEWVLLSTFL(서열번호 13)의 아미노산 서열을 가질 수 있다. 따라서, 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질의 엑토도메인, 또는 이의 항원성 단편, 및 C-말단 폴돈을 암호화할 수 있다. 특정 구현예에서, 폴돈은 SARS-CoV-2 S 단백질의 엑토도메인 또는 SARS-CoV-2 S 단백질의 S2' 서브유닛의 C-말단에 위치한다. 일 구현예에서, 본 발명은, C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 엑토도메인을 포함하는 아미노산 서열을 암호화하는, 서열번호 14를 포함하는 아미노산 서열을 갖는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 46의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 46과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 14의 아미노산 서열을 암호화한다. 본 발명은 또한, C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 S2 서브유닛을 포함하는 아미노산 서열을 암호화하는, 서열번호 76을 포함하는 아미노산 서열을 갖는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 75의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 75와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 76의 아미노산 서열을 암호화한다. 본 발명은 또한, C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 S2' 서브유닛을 포함하는 아미노산 서열을 암호화하는, 서열번호 15를 포함하는 아미노산 서열을 갖는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 47의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 47과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 15의 아미노산 서열을 암호화한다.

[0164] 일부 구현예에서, 최적화된 뉴클레오티드 서열은 C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인을 암호화하며, 여기에서 엑토도메인은 활성화에 필요한 퓨린 절단 부위를 제거하고/하거나 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질의 엑토도메인에 비해 변형되었다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은, 서열번호 16을 포함하는 아미노산 서열을 갖는, C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 48의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 48과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 16의 아미노산 서열을 암호화한다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인을 암호화하며, 여기에서 엑토도메인은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되었다. 따라서, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 17을 포함하는 아미노산 서열을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 49의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 49와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 17의 아미노산 서열을 암호화한다.

[0165] 일부 구현예에서, 최적화된 뉴클레오티드 서열은 C-말단 폴돈을 갖는 SARS-CoV-2 S 단백질의 S2 또는 S2' 서브유닛의 융합-전 안정화된 엑토도메인을 암호화하며, 여기에서 엑토도메인은 자연 발생 SARS-CoV-2 S 단백질에 비해 잔기 986 및 987이 프롤린으로 돌연변이되었다. 따라서, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 80을 포함하는 아미노산 서열을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 S2 서브유닛을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 79의 서열을 갖는다. 다른 구현예에서, 최적화된

뉴클레오티드 서열은 서열번호 79와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 80의 아미노산 서열을 암호화한다. 따라서, 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 84를 포함하는 아미노산 서열을 갖는 SARS-CoV-2 S 단백질의 융합-전 안정화된 S2 서브유닛을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 83의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 83과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 84의 아미노산 서열을 암호화한다.

[0166] 단백질 중 Fc 도메인의 존재는 단백질의 혈장 반감기를 현저하게 증가시킴으로써 분자의 치료 활성을 연장시킨다. Fc 도메인은 또한 혈류로부터의 단백질의 신장 제거를 느리게 할 수 있고, 백신에 사용하기에 유리할 수 있는 특징인 면역 세포에서 발견되는 Fc-수용체(FcR)와 단백질의 상호작용을 가능하게 한다. 또한, Fc 도메인은 독립적으로 접히고, 시험관 내 및 생체 내 파트너 분자의 용해도 및 안정성을 개선할 수 있다(Czajkowsky 등 (2012) EMBO Mol Med. (10): 1015-1028). 따라서, 본 발명은 또한, C-말단 Fc 도메인으로 SARS-CoV-2 S 단백질의 엑토도메인 또는 이의 항원성 단편을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. Fc 도메인은 다음의 아미노산 서열을 포함할 수 있다:

[0167] PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (서열번호 18). 특정 구현예에서, 항원성 단편은 SARS-CoV-2 S 단백질의 RBD이다. 일부 구현예에서, 최적화된 뉴클레오티드 서열은, 서열번호 19를 포함하는 아미노산 서열을 갖는 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 50의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 50과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 19의 아미노산 서열을 암호화한다.

[0168] 본 발명은 또한, N-말단 신호 펩티드 및 C-말단 Fc 도메인에 작동 가능하게 연결된, SARS-CoV-2 S 단백질의 엑토도메인 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. Fc는 서열번호 18의 아미노산 서열을 가질 수 있다. 신호 펩티드는 서열번호 7의 아미노산 서열을 가질 수 있다. 특정 구현예에서, 항원성 단편은 SARS-CoV-2 S 단백질의 RBD이다. 일부 구현예에서, 본 발명은 서열번호 20을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드 및 C-말단 Fc 도메인에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 36의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 36과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 20의 아미노산 서열을 암호화한다.

[0169] 항체의 약동학적 특성은 주로 신생아 Fc 수용체(FcRn)에 대한 Fc 도메인의 pH-의존적 결합에 의해 결정된다. 예를 들어, 아미노산 치환 M428L/N434S(LS 돌연변이체), M252Y/S254T/T256E(YTE 돌연변이체), 또는 H433K/N434F(KF 돌연변이체)를 함유하는 Fc 도메인은 pH 5.8에서 FcRn에 대해 10 내지 12배 더 높은 친화도를 부여한다. 이는 항체 반감기를 크게 증가시킨다(2 내지 4배 더 긴 순환 시간). 따라서, 본 발명의 융합 단백질에 포함된 Fc 영역을 변형시키는 것은 이의 혈청에서의 반감기를 연장할 수 있다. L309D/Q311H/N434S(DHS) 치환을 함유하는 Fc 변이체는 천연 IgG1 및 전술한 변이체 둘 모두에 비해 항체의 약동학을 추가로 개선하는 것으로 나타났다(Lee 등 (2019) Nature communications, 10, 5031). 따라서, 특정 구현예에서, Fc 영역은 인간 IGHG에 기초한 EU 넘버링 시스템을 사용하여, 야생형과 비교하여 돌연변이되었다. 예를 들어, 위치 309에서의 L 잔기, 311에서의 Q 잔기 및 434에서의 N 잔기는 각각 D, H 및 S로(즉, L309D; Q311H 및 N434S로) 돌연변이될 수 있다. 돌연변이된 Fc 도메인은 다음의 서열을 포함할 수 있다:

[0170] PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTV**D**H**H**DWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALH**S**HYTQKSLSLSPGK (서열번호 100).

[0171] 다른 구현예에서, 위치 428에서의 M 잔기 및 434에서의 N 잔기는 각각 L 및 S로(즉, M428L 및 N434S로) 돌연변이될 수 있다. 돌연변이된 Fc 도메인은 다음의 서열을 포함할 수 있다:

[0172] PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSV**L**HEALH**S**HYTQKSLSLSPGK (서열번호 101).

[0173] 다른 구현예에서, 위치 252에서의 M 잔기, 위치 254에서의 S 잔기 및 256에서의 T 잔기는 각각 Y, T 및 E로(즉, M252Y, S254T 및 T256E로) 돌연변이될 수 있다. 돌연변이된 Fc 도메인은 다음의 서열을 포함할 수 있다:

[0174] PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTL**Y**I**T**R**E**PEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKE YKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQG NVFSCSVMHEALHNHYTQKSLSLSPGK (서열번호 102).

[0175] 다른 구현예에서, 위치 433에서의 H 잔기 및 434에서의 N 잔기는 각각 K 및 F로(즉, H433K 및 N434F로) 돌연변이될 수 있다. 돌연변이된 Fc 도메인은 다음의 서열을 포함할 수 있다:

[0176] PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKE YKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQG NVFSCSVMHEAL**KF**HYTQKSLSLSPGK

[0177] (서열번호 103).

[0178] 따라서, 본 발명은 또한, N-말단 신호 펩티드 및 C-말단 Fc 도메인에 작동 가능하게 연결된, SARS-CoV-2 S 단백질의 항원성 단편 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. Fc는 서열번호 100, 서열번호 101, 서열번호 102 또는 서열번호 103의 아미노산 서열을 가질 수 있다. 신호 펩티드는 서열번호 7의 아미노산 서열을 가질 수 있다. 특정 구현예에서, 항원성 단편은 SARS-CoV-2 S 단백질의 RBD이다.

[0179] 일부 구현예에서, 본 발명은 서열번호 104를 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드 및 C-말단 돌연변이된 Fc 도메인에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 105의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 105와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 104의 아미노산 서열을 암호화한다.

[0180] 일부 구현예에서, 본 발명은 서열번호 106을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드 및 C-말단 돌연변이된 Fc 도메인에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 107의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 107과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 106의 아미노산 서열을 암호화한다.

[0181] 일부 구현예에서, 본 발명은 서열번호 108을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드 및 C-말단 돌연변이된 Fc 도메인에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 109의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 109와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 108의 아미노산 서열을 암호화한다.

[0182] 일부 구현예에서, 본 발명은 서열번호 110을 포함하는 아미노산 서열을 갖는, N-말단 신호 펩티드 및 C-말단 돌연변이된 Fc 도메인에 작동가능하게 연결된 SARS-CoV-2 S 단백질의 RBD를 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 111의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 111과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 110의 아미노산 서열을 암호화한다.

[0183] 코로나바이러스는 소포체(ER)-골지 중간 구획(ERGIC)의 내강에서 어셈블되고 그 내부로 퍼진다. SARS-CoV-2 S 단백질의 세포질 꼬리는 S 단백질을 골지로부터 ER로 이동시킬 수 있는 ER 회수 신호(ERRS)를 함유한다. 이러한 프로세스는 바이러스 입자 내로의 S 단백질 혼입을 용이하게 하는, ERGIC에서의 S 단백질 축적으로 여겨진다. SARS CoV S 단백질의 ER 회수 신호는 세포질 꼬리의 이염기성 모티프(KxHxx)이며, 이는 표준 딜리신 ER 회수 신호와 유사하다(McBride 등 (2007) Journal Of Virology, 81, 5, 2418-2428).

[0184] ER 회수 신호를 돌연변이시키는 것은 바이러스가 바이러스 입자를 형성하는 것을 방지할 수 있다. 임의의 특정 이론에 구속되고자 함이 없이, 본 발명자는 백신에 포함되도록 의도된 SARS-CoV-2 S 단백질로부터 ER 회수 신호를 제거하는 것이 유리하다고 믿는다. 따라서, 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, ER 회수 신호를 돌연변이시킴으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, SARS-CoV-2 S 단백질의 KLHYT ER 회수 신호는 잔기 1268 및 1270을 알라

닌으로 돌연변이시킴으로써(즉, ALAYT) 제거될 수 있다.

[0185] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고, 잔기 986 및 987을 프롤린으로 돌연변이시키고, ER 회수 신호를 제거함으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 125를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 124의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 124와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 125의 아미노산 서열을 암호화한다.

[0186] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고, 잔기 986 및 987을 프롤린으로 돌연변이시키고, ER 회수 신호를 제거함으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되었으며, 확장된 N-말단 신호 펩티드를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 127을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 126의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 126과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 127의 아미노산 서열을 암호화한다.

[0187] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고, 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고, ER 회수 신호를 제거함으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 139를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 138의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 138과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 139의 아미노산 서열을 암호화한다.

[0188] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고, 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고, ER 회수 신호를 제거함으로써 자연 발생 SARS-CoV-2 S 단백질에 비해 변형되었으며, 확장된 N-말단 신호 펩티드를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 141을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 140의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 140과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 141의 아미노산 서열을 암호화한다.

[0189] 전술한 단락(단락 0 및 [0110])에 열거된 돌연변이의 특정 조합은 본원에 개시된 SARS-CoV-2 S 단백질 중 어느 하나에 도입될 수 있다. 예를 들어, 특정 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하도록 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 151을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 150의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 150과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 151의 아미노산 서열을 암호화한다.

[0190] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이(UK 변이체 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하

나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 153을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 152의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 152와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 153의 아미노산 서열을 암호화한다.

[0191] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거함으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이(UK 변이체 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 155를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 154의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 154와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 155의 아미노산 서열을 암호화한다.

[0192] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이(UK 변이체 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 157을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 156의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 156과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 157의 아미노산 서열을 암호화한다.

[0193] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 817, 892, 899 및 942, 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, ΔH69, ΔV70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이(UK 변이체 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 159를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 158의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 158과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 159의 아미노산 서열을 암호화한다.

[0194] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 1 + D614G)를 함유하도록 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 161을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 160의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 160과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 161의 아미노산 서열을 암호화한다.

[0195] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 1 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 163을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 162

의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 162와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 163의 아미노산 서열을 암호화한다.

[0196] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, L18F, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 2 + D614G)를 함유하도록 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 165를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 164의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 164와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 165의 아미노산 서열을 암호화한다.

[0197] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, L18F, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 2 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 167을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 166의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 166과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 167의 아미노산 서열을 암호화한다.

[0198] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이(브라질 변이체 + D614G)를 함유하도록 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형된, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 169를 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 168의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 168과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 169의 아미노산 서열을 암호화한다.

[0199] 일부 구현예에서, 본 발명에 따른 최적화된 뉴클레오티드 서열은, 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시킴으로써 우한으로부터의 인덱스 계열의 SARS-CoV-2 S 단백질(서열번호 1)에 대해 변형되고, L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이(브라질 변이체 + D614G)를 함유하는, 융합-전 안정화된 SARS-CoV-2 S 단백질, SARS-CoV-2 S 단백질의 융합-전 안정화된 엑토도메인, 또는 이들 중 어느 하나의 항원성 단편을 암호화할 수 있다. 따라서, 소정의 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 171을 포함하는 아미노산 서열을 갖는, 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화할 수 있다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 170의 서열을 갖는다. 다른 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 170과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하고, 서열번호 171의 아미노산 서열을 암호화한다.

[0200] *SARS-CoV-2 S 단백질 및 항원성 단편을 암호화하는 예시적인 최적화된 뉴클레오티드 서열*

[0201] 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 일 구현예에서, 본 발명은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서열은 효율적인 발현 인간 세포에 대해 최적화된 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 뉴클레오티드 서열을 포함한다. 본 발명에 따른 최적화된 뉴클레오티드 서열 및 이에 상응하는 아미노산 서열을 생성하는 프로세스로 생성된 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 예시적인 최적화된 뉴클레오티드 서열이 표 1에 제시된다. 굵은체로 표시되는 잔기는 자연 발생 SARS-CoV-2 S 단백질과 비교하여 돌연변이된 아미노

산을 나타내고, 밑줄이 쳐진 잔기는 신호 펩티드를 나타내며, 이탤릭체로 표시되는 잔기는 Fc 영역 또는 폴돈의 존재를 나타낸다.

**[표 1] 예시적인 SARS-CoV-2 S 서열.**

| | |
|---|---|
| SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 29)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA |

[0202]

[0203]

| | |
|---|---|
| | GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |

| | |
|---|---|
| SARS-CoV-2 S 단백질 서열 | (서열번호 1)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 30)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG |

[0204]

[0205]

| | |
|---|---|
| | CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG |

[0206]

| | |
|---|---|
| | CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTGA |
| SARS-CoV-2 S<br>단백질의 엑토도메인 | **(서열번호 2)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF |

| | |
|---|---|
| | TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQ |
| SARS-CoV-2 S 단백질의 S1 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 31)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC |

[0207]

| | |
|---|---|
| | ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTTGA |
| SARS-CoV-2 S 단백질의 S1 서브유닛 | (서열번호 3)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP |
| SARS-CoV-2 S 단백질의 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 32)<br>ATGTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC<br>GGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGA<br>TGGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCT<br>CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATAAGGTGGAGGCTGAAGTCCAGATTGACCGCCTGA<br>TTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAG |

[0208]

| | |
|---|---|
| | CAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATC<br>TGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTC<br>CAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATG<br>AGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCA<br>CGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTG<br>CTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGG<br>GAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGAC<br>CCAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTGA |
| SARS-CoV-2 S 단백질의 S2 서브유닛 | **(서열번호 4)**<br>MSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPV<br>SMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV<br>EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS<br>FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLT<br>VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDK<br>VEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM<br>SECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPA<br>QEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEP<br>QIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYF<br>KNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL<br>QELGKYEQ |
| SARS-CoV-2 S 단백질의 전장 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 71)**<br>ATGTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC<br>GGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGA<br>TGGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCT |

[0209]

| | |
|---|---|
| | CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATAAGGTGGAGGCTGAAGTCCAGATTGACCGCCTGA<br>TTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAG<br>CAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATC<br>TGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTC<br>CAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATG<br>AGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCA<br>CGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTG<br>CTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGG<br>GAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGAC<br>CCAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGAAGCTGCATTATACCTGA |
| SARS-CoV-2 S 단백질의 전장 S2 서브유닛 | **(서열번호 72)**<br>MSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPV<br>SMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV<br>EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS<br>FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLT<br>VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDK<br>VEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM<br>SECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPA<br>QEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEP<br>QIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYF<br>KNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL<br>QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSC<br>LKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 73)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA |

[0210]

| | |
|---|---|
| | GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT<br>GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATAAGGTGGAGGCTGAAGTCCAGATTGACCGCCTGATT<br>ACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTGA |
| 신호 서열을 갖는<br>SARS-CoV-2 S<br>단백질의 S2 서브유닛 | **(서열번호 74)**<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAE<br>IRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQ |

[0211]

| | |
|---|---|
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 전장 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 67)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA<br>GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT<br>GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATAAGGTGGAGGCTGAAGTCCAGATTGACCGCCTGATT<br>ACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGAAGCTGCATTATACCTGA |

[0212]

| | |
|---|---|
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 전장 S2 서브유닛 | (서열번호 68)<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAE<br>IRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVT<br>IMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |
| SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 33)<br>ATGAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTAA |

[0213]

| | |
|---|---|
| SARS-CoV-2 S 단백질의 S2' 서브유닛 | (서열번호 5)<br>MSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNG<br>LTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM<br>QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTAS<br>ALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD<br>KVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATK<br>MSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVP<br>AQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYE<br>PQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKY<br>FKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLID<br>LQELGKYEQ |
| SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 97)<br>ATGAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATC<br>GCCGGACTGATTGCCATCGTCATGGTGACCATCATGCTGTG<br>TTGCATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTA<br>GTTGCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAG<br>CCCGTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |

[0214]

| | |
|---|---|
| SARS-CoV-2 S 단백질의 전장 S2' 서브유닛 | (서열번호 98)<br>MSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNG<br>LTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM<br>QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTAS<br>ALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD<br>KVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATK<br>MSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVP<br>AQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYE<br>PQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKY<br>FKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLID<br>LQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCS<br>CLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 65)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTAA |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2' 서브유닛 | (서열번호 66)<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG |

[0215]

| | |
|---|---|
| | AISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQQCSFIEDLLFNKVTLADAGFI<br>KQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSAL<br>LAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYEN<br>QKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTL<br>VKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQT<br>YVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYH<br>LMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFP<br>REGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIV<br>NNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASV<br>VNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 95)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATC<br>GCCGGACTGATTGCCATCGTCATGGTGACCATCATGCTGTG<br>TTGCATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTA |

[0216]

| | |
|---|---|
| | GTTGCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAG<br>CCCGTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 전장 S2' 서브유닛 | (서열번호 96)<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVM<br>VTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLH<br>YT |
| SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 34)<br>ATGCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTT<br>CAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGG<br>AAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTA<br>TAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGA<br>GCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTAC<br>GCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGA<br>TCGCACCAGGACAGACAGGCAAGATTGCTGACTACAACTA<br>TAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGA<br>ACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAA<br>TTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCT<br>TCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTC<br>CACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCC<br>CCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGT<br>ACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTG<br>CATGCTCCATAA |
| SARS-CoV-2 S 단백질의 수용체 결합 도메인 | (서열번호 6)<br>MPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYN<br>SASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPG<br>QTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYR<br>LFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQ<br>PTNGVGYQPYRVVVLSFELLHAP |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 35)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT |

[0217]

| | |
|---|---|
| | ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCATAA |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | **(서열번호 8)**<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP |
| 퓨린 절단 부위를 제거하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 42)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT |

[0218]

[0219]

| | |
|---|---|
| | CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG |

| | |
|---|---|
| | AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하도록 돌연변이된 SARS-CoV-2 S 단백질 | (서열번호 9)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 43)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC |

[0220]

[0221]

| | |
|---|---|
| | TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG |

| | |
|---|---|
| | AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖는 SARS-CoV-2 S 단백질 | **(서열번호 10)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ |

[0222]

| | |
|---|---|
| | AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| **퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열** | **(서열번호 44)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT |

[0223]

[0224]

| | |
|---|---|
| | ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT |

| | |
|---|---|
| | TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질 | (서열번호 11)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |

[0225]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 45)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC |

[0226]

| | |
|---|---|
| | GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTGA |
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖고, D614G 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | **(서열번호 118)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC |

[0227]

| | |
|---|---|
| | TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖고, D614G 돌연변이*를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열<br>*밑줄이 그어진 잔기는 D614G 돌연변이 위치에 상응함 | **(서열번호 119)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT |

[0228]

[0229]

| | |
|---|---|
| | GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGGCGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGC<br>AGACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGA<br>TCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGC<br>CGAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATT<br>GGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTC<br>TCCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTG<br>CCTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTAC<br>TCCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCT<br>GTGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAG<br>CGTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAAT<br>GTTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAG<br>CTGAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACA<br>AGAACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTA<br>TAAGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCT<br>CACAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAG<br>CTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAG<br>ACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGAC<br>ATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGG<br>CCTCACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCG<br>CTCAATACACTAGCGCACTGCTGGCCGGAACCATCACATCA<br>GGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATT<br>CGCCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCA<br>CACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAA<br>CCAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCA<br>GCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTC<br>AACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGC<br>TGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGAC<br>ATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGAT<br>TGACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACAT |

| | |
|---|---|
| | ACGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGC<br>ATCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTG<br>CTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCT<br>ACCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTT<br>GTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAA<br>CTTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCC<br>ACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACAC<br>TGGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCAT<br>CACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCG<br>TGATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAG<br>CCAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTT<br>TAAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATC<br>TCCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGA<br>TTGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCT<br>CTGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATAT<br>CAAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGAC<br>TGATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATG<br>ACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGG<br>CTCTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGC<br>TGAAGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하고, D614G 돌연변이를 함유하도록 돌연변이된 SARS-CoV-2 S 단백질 | **(서열번호 120)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE |

[0230]

| | |
|---|---|
| | QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하고, D614G 돌연변이*를 함유하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열<br>*밑줄이 그어진 잔기는 D614G 돌연변이 위치에 상응함 | (서열번호 121)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGGCGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGC |

[0231]

| | |
|---|---|
| | AGACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGA<br>TCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGC<br>CGAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATT<br>GGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTC<br>TCCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |

[0232]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질의 엑토도메인 | (서열번호 12)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>Q |
| 폴돈을 갖는 SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 46)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA |

[0233]

[0234]

| | |
|---|---|
| | GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA |

[0235]

| | |
|---|---|
| | CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGGGGTA<br>CATTCCCGAGGCTCCTAGGGACGGCCAGGCATACGTGCGC<br>AAAGACGGCGAGTGGGTGCTGCTGTCCACATTCCTGTAA |
| 폴돈을 갖는 SARS-CoV-2 S 단백질의 엑토도메인 | **(서열번호 14)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL |

| | |
|---|---|
| | GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQ*GYIPEAPRDGQAYVRKDGEWVLLSTFL* |
| 폴돈 및 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 47)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGGGGTACATTCCCGAGGCTCCTAGGGACGGCCAGGCATA<br>CGTGCGCAAAGACGGCGAGTGGGTGCTGCTGTCCACATTCC<br>TGTAA |
| 폴돈 및 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2' 서브유닛 | **(서열번호 15)**<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL |

[0236]

| | |
|---|---|
| | NEVAKNLNESLIDLQELGKYEQ*GYIPEAPRDGQAYVRKDGEWV LLSTFL* |
| 폴돈을 갖는, 잔기 986 및 987 을 프롤린으로 돌연변이시킴으로써 변형된, SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 48)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA |

[0237]

[0238]

| | |
|---|---|
| | GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGGGGTA<br>CATTCCCGAGGCTCCTAGGGACGGCCAGGCATACGTGCGC<br>AAAGACGGCGAGTGGGTGCTGCTGTCCACATTCCTGTAA |

| | |
|---|---|
| 폴돈을 갖는, 잔기 986 및 987 을 프롤린으로 돌연변이시킴으로써 변형된, SARS-CoV-2 S 단백질의 엑토도메인 | (서열번호 16)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>*QGYIPEAPRDGQAYVRKDGEWVLLSTFL* |
| C-말단 폴돈을 갖는, 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 변형된, SARS-CoV-2 S 단백질의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 49)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA |

[0239]

[0240]

| | |
|---|---|
| | GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA |

| | |
|---|---|
| | GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGGGGTACAT<br>TCCCGAGGCTCCTAGGGACGGCCAGGCATACGTGCGCAAA<br>GACGGCGAGTGGGTGCTGCTGTCCACATTCCTGTAA |
| C-말단 폴돈을 갖는, 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987을 대체하도록 변형된, SARS-CoV-2 S 단백질의 엑토도메인 | **(서열번호 17)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG |

[0241]

| | |
|---|---|
| | QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>*QGYIPEAPRDGQAYVRKDGEWVLLSTFL* |
| Fc 영역을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 50)**<br>ATGCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTT<br>CAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGG<br>AAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTA<br>TAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGA<br>GCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTAC<br>GCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGA<br>TCGCACCAGGACAGACAGGCAAGATTGCTGACTACAACTA<br>TAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGA<br>ACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAA<br>TTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCT<br>TCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTC<br>CACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCC<br>CCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGT<br>ACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTG<br>CATGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCC<br>ACCATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTT<br>TCCTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCTCTC<br>GCACACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCAC<br>GAGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAG<br>TGGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAAC<br>AATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTG<br>CTGCACCAGGATTGGCTGAATGGAAAAGAATATAAGTGTA<br>AGGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGAC<br>AATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTG<br>TACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATC<br>AGGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGT<br>GACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAA<br>ATAACTACAAGACCACACCACCAGTGCTCGATAGCGACGG<br>GTCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCG<br>GTGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACG<br>AAGCTCTGCACAATCACTATACACAGAAATCCCTGTCCCTG<br>TCTCCAGGCAAATAA |
| Fc 영역을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | **(서열번호 19)**<br>MNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNS<br>ASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQ<br>TGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRL<br>FRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQP<br>TNGVGYQPYRVVVLSFELLHAP*PKSCDKTHTCPPCPAPELLGG*<br>*PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGV*<br>*EVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNK*<br>*ALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGF*<br>*YPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW*<br>*QQGNVFSCSVMHEALHNHYTQKSLSLSPGK* |

[0242]

| | |
|---|---|
| 신호 서열 및 Fc 영역을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 36)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT<br>ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCCAC<br>CATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCTCTCG<br>CACACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCACG<br>AGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAGT<br>GGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAACA<br>ATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTGC<br>TGCACCAGGATTGGCTGAATGGAAAAGAATATAAGTGTAA<br>GGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGACA<br>ATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTGT<br>ACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATCA<br>GGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGTG<br>ACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGACGGG<br>TCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCGG<br>TGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACGA<br>AGCTCTGCACAATCACTATACACAGAAATCCCTGTCCCTGT<br>CTCCAGGCAAATAA |
| 신호 서열 및 Fc 영역을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | (서열번호 20)<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP*PKSCDKT*<br>*HTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSH*<br>*EDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD*<br>*WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE*<br>*LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG*<br>*SFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK* |

[0243]

| | |
|---|---|
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 69)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA<br>GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT<br>GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGATTA<br>CCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGCA<br>GCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCTG<br>GCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCCA<br>AGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGAG<br>CTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCACG<br>TGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGCT<br>CCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGGA<br>GGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACCC<br>AGAGGAACTTCTATGAACCCCAGATCATCACCACTGACAAT<br>ACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATCGT<br>TAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGACT<br>CCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACAC<br>AAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAAC<br>GCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTAA<br>ATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCTG<br>CAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCCT<br>GGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCATC<br>GTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTTG<br>TTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGTA<br>AATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCGT<br>GAAGCTGCATTATACCTGA |

[0244]

| | |
|---|---|
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2 서브유닛 | (서열번호 70)<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEI<br>RASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVT<br>IMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |
| 신호 서열 및 폴돈을 갖는 SARS-CoV-2 S 단백질의 S2 서브유닛의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 75)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA<br>GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT<br>GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATAAGGTGGAGGCTGAAGTCCAGATTGACCGCCTGATT<br>ACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA |

[0245]

| | |
|---|---|
| | ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGGGGTACATTCCCGAG<br>GCTCCTAGGGACGGCCAGGCATACGTGCGCAAAGACGGCG<br>AGTGGGTGCTGCTGTCCACATTCCTGTGA |
| 신호 서열 및 폴돈을 갖는 SARS-CoV-2 S 단백질의 S2 서브유닛 | **(서열번호 76)**<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAE<br>IRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQ*GYIPEAPRDGQAYVRKDGEWVLL<br>STFL* |
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 77)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA<br>GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT<br>GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGATTA |

[0246]

| | |
|---|---|
| | CCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGCA<br>GCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCTG<br>GCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCCA<br>AGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGAG<br>CTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCACG<br>TGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGCT<br>CCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGGA<br>GGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACCC<br>AGAGGAACTTCTATGAACCCCAGATCATCACCACTGACAAT<br>ACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATCGT<br>TAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGACT<br>CCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACAC<br>AAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAAC<br>GCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTAA<br>ATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCTG<br>CAGGAACTGGGCAAGTATGAGCAGTGA |
| 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2 서브유닛 | **(서열번호 78)**<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRAAEI<br>RASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQ |
| 프롤린 및 폴돈으로 잔기 986 및 987 을 제거하도록 변형된, 신호 서열을 갖는 SARS-CoV-2 S 단백질의 S2 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 79)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAGC<br>CTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCAT<br>CGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAGA<br>TCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTACC<br>ATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGCT<br>GCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCCC<br>TGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGGA<br>GGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCTA<br>TTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCCA<br>GACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGATCT<br>CCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTATTA<br>AGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGAGA<br>CCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCTGC<br>CACCTCTGCTGACCGACGAGATGATCGCTCAATACACTAGC<br>GCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTCGG<br>GGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGATG<br>GCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTGCT |

[0247]

| | |
|---|---|
| | GTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTCCG<br>CAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCTCT<br>GCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGCTC<br>AGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACTTT<br>GGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGCCT<br>GGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGATTA<br>CCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGCA<br>GCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCTG<br>GCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCCA<br>AGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGAG<br>CTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCACG<br>TGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGCT<br>CCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGGA<br>GGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACCC<br>AGAGGAACTTCTATGAACCCCAGATCATCACCACTGACAAT<br>ACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATCGT<br>TAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGACT<br>CCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACAC<br>AAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAAC<br>GCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTAA<br>ATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCTG<br>CAGGAACTGGGCAAGTATGAGCAGGGGTACATTCCCGAGG<br>CTCCTAGGGACGGCCAGGCATACGTGCGCAAAGACGGCGA<br>GTGGGTGCTGCTGTCCACATTCCTGTGA |
| 신호 서열 및 폴돈을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2 서브유닛 | **(서열번호 80)**<br>MFVFLVLLPLVSSQCSVASQSIIAYTMSLGAENSVAYSNNSIAI<br>PTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGS<br>FCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFN<br>FSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>ARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF<br>GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSA<br>IGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAI<br>SSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRAAEI<br>RASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHG<br>VVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTH<br>WFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE<br>LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE<br>VAKNLNESLIDLQELGKYEQ*GYIPEAPRDGQAYVRKDGEWVLL<br>STFL* |
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 81)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT |

[0248]

| | |
|---|---|
| | CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTGA |
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2' 서브유닛 | **(서열번호 82)**<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQ |
| 신호 서열 및 폴돈을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 83)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG |

[0249]

| | |
|---|---|
| | CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGGGGTACATTCCCGAGGCTCCTAGGGACGGCCAGGCATA<br>CGTGCGCAAAGACGGCGAGTGGGTGCTGCTGTCCACATTCC<br>TGTGA |
| 신호 서열 및 폴돈을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 S2' 서브유닛 | **(서열번호 84)**<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQ*GYIPEAPRDGQAYVRKDGEWV<br>LLSTFL* |
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 전장 S2' 서브유닛을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 85)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA |

[0250]

| | |
|---|---|
| | AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATC<br>GCCGGACTGATTGCCATCGTCATGGTGACCATCATGCTGTG<br>TTGCATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTA<br>GTTGCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAG<br>CCCGTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |
| 신호 서열을 갖는, 프롤린으로 잔기 986 및 987 을 제거하도록 변형된, SARS-CoV-2 S 단백질의 전장 S2' 서브유닛 | **(서열번호 86)**<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVM<br>VTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLH<br>YT |
| 프롤린으로 잔기 985 를 제거하도록 변형된, SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 87)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA |

[0251]

[0252]

| | |
|---|---|
| | GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT<br>CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA |

| | |
|---|---|
| | CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGCCCAAGGTGGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |
| 프롤린으로 잔기 985를 제거하도록 변형된, SARS-CoV-2 S 단백질 | **(서열번호 88)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF |

[0253]

| | |
|---|---|
| | NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLPKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 985를 대체하도록 변형된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 89)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT |

[0254]

[0255]

| | |
|---|---|
| | AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGCCCAAGGTGGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA |

| | |
|---|---|
| | TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 985 를 대체하도록 변형된 SARS-CoV-2 S 단백질 | **(서열번호 90)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRL**P**K<u>V</u>EAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| 프롤린으로 잔기 985, 986 및 987 을 대체하도록 변형된, SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 91)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA |

[0256]

[0257]

| | |
|---|---|
| | GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTT |

| | |
|---|---|
| | CATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACG<br>CCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATT<br>GCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCT<br>CACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTC<br>AATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGG<br>CTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCG<br>CCATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACA<br>CAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACC<br>AGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGC<br>TCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGCCTCCACCCGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |
| 프롤린으로 잔기 985, 986 및 987 을 대체하도록 변형된, SARS-CoV-2 S 단백질 | **(서열번호 92)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN |

[0258]

| | |
|---|---|
| | TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRL**PPP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 985, 986 및 987을 대체하도록 변형된 SARS-CoV-2 S 단백질 서열을 암호화하는 최적화된 뉴클레오티드 | **(서열번호 93)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG |

[0259]

| | |
|---|---|
| | GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGCCTCCACCCGAGGCTGAAGTCCAGATTGACC<br>GCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTG<br>ACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCG<br>CAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGG<br>CCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACC<br>TGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTT<br>CTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTAC<br>AACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCC<br>CACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTC |

[0260]

| | |
|---|---|
| | GTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACCAC<br>TGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCG<br>GCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAG<br>CTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAA<br>CCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGA<br>ATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACC<br>GCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATT<br>GATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAAT<br>GGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATT<br>GCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTC<br>CTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTG<br>CTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAG<br>GGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 985, 986 및 987 을 대체하도록 변형된 SARS-CoV-2 S 단백질 | **(서열번호 94)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRL**PPP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |

[0261]

| | |
|---|---|
| 신호 서열을 갖는 SARS-CoV-2 S S2' 서브유닛 단백질 서열을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 95)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTAGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCT<br>GGCAGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGG<br>GCGACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTC<br>AATGGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGAT<br>GATCGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCA<br>CATCAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGAT<br>TCCATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTG<br>GCGTCACACAGAACGTGCTGTACGAAAACCAGAAGCTCAT<br>CGCTAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATT<br>CACTCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGAC<br>GTGGTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCA<br>AGCAGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTG<br>AACGACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAG<br>TCCAGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTG<br>CAAACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGA<br>TCCGGGCATCCGCAAATCTGGCAGCAACTAAGATGAGCGA<br>ATGCGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCA<br>AGGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACAT<br>GGCGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGA<br>AAAGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCA<br>AGGCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGC<br>ACACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCA<br>GATCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCG<br>ACGTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCT<br>CTCCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATA<br>AGTATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGG<br>GGACATCTCCGGAATTAACGCCTCCGTGGTGAATATCCAGA<br>AGGAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAA<br>TGAGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGC<br>AGTATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATC<br>GCCGGACTGATTGCCATCGTCATGGTGACCATCATGCTGTG<br>TTGCATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTA<br>GTTGCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAG<br>CCCGTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |
| 신호 서열을 갖는 SARS-CoV-2 S 단백질의 전장 SARS-CoV-2 S2' 서브유닛 | (서열번호 96)<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGDCLGDI<br>AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWT<br>FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS<br>AIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG<br>AISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA<br>AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAP<br>HGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNG<br>THWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQ<br>PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL<br>NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVM<br>VTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLH<br>YT |

[0262]

| | |
|---|---|
| 신호 서열 및 돌연변이된 Fc 영역(L309D/Q311H/N434S)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 105)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT<br>ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCCAC<br>CATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCTCTCG<br>CACACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCACG<br>AGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAGT<br>GGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAACA<br>ATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTGG<br>ATCACCATGATTGGCTGAATGGAAAAGAATATAAGTGTAA<br>GGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGACA<br>ATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTGT<br>ACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATCA<br>GGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGTG<br>ACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGACGGG<br>TCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCGG<br>TGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACGA<br>AGCTCTGCACTCTCACTATACACAGAAATCCCTGTCCCTGT<br>CTCCAGGCAAATAA |
| 신호 서열 및 돌연변이된 Fc 영역(L309D/Q311H/N434S)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | **(서열번호 104)**<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP*PKSCDKT*<br>*HTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSH*<br>*EDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTV**D**H**H**D*<br>*WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE*<br>*LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG*<br>*SFFLYSKLTVDKSRWQQGNVFSCSVMHEALH**S**HYTQKSLSLSPGK* |

[0263]

| | |
|---|---|
| 신호 서열 및 돌연변이된 Fc 영역(M428L/N434S)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 107)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT<br>ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCCAC<br>CATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCTATATCACTCG<br>CGAACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCACG<br>AGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAGT<br>GGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAACA<br>ATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTGC<br>TGCACCAGGATTGGCTGAATGGAAAAGAATATAAGTGTAA<br>GGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGACA<br>ATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTGT<br>ACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATCA<br>GGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGTG<br>ACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGACGGG<br>TCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCGG<br>TGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACGA<br>AGCTCTGCACAATCACTATACACAGAAATCCCTGTCCCTGT<br>CTCCAGGCAAATAA |
| 신호 서열 및 돌연변이된 Fc 영역(M428L/N434S)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | **(서열번호 106)**<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP*PKSCDKT*<br>*HTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSH*<br>*EDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD*<br>*WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE*<br>*LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG*<br>*SFFLYSKLTVDKSRWQQGNVFSCSVLHEALHSHYTQKSLSLSPGK* |

[0264]

| | |
|---|---|
| 신호 서열 및 돌연변이된 Fc 영역(M252Y/S254T/T256E)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 109)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT<br>ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCCAC<br>CATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCTATATCACTCG<br>CGAACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCACG<br>AGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAGT<br>GGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAACA<br>ATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTGC<br>TGCACCAGGATTGGCTGAATGGAAAAGAATATAAGTGTAA<br>GGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGACA<br>ATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTGT<br>ACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATCA<br>GGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGTG<br>ACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGACGGG<br>TCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCGG<br>TGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACGA<br>AGCTCTGCACAATCACTATACACAGAAATCCCTGTCCCTGT<br>CTCCAGGCAAATAA |
| 신호 서열 및 돌연변이된 Fc 영역(M252Y/S254T/T256E)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | **(서열번호 108)**<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP*PKSCDKT*<br>*HTCPPCPAPELLGGPSVFLFPPKPKDTL**YITRE**PEVTCVVVDVSH*<br>*EDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD*<br>*WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE*<br>*LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG*<br>*SFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK* |

[0265]

| | |
|---|---|
| 신호 서열 및 돌연변이된 Fc 영역(H433K/N434F)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 111)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTC<br>AACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGA<br>AGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTAT<br>AACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAG<br>CCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCTACG<br>CCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATC<br>GCACCAGGACAGACAGGCAAGATTGCTGACTACAACTATA<br>AGCTGCCTGACGACTTCACAGGATGTGTGATCGCATGGAAC<br>TCAAACAATCTGGACTCCAAAGTCGGGGGCAACTATAATT<br>ACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTC<br>GAGAGGGACATCAGTACAGAGATCTATCAGGCTGGCTCCA<br>CCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCC<br>TGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGGGTAC<br>CAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCA<br>TGCTCCACCTAAGTCCTGCGACAAAACCCATACATGTCCAC<br>CATGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCTCTCG<br>CACACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTCACG<br>AGGATCCTGAAGTGAAGTTTAACTGGTATGTCGACGGAGT<br>GGAAGTGCACAACGCCAAGACAAAGCCAAGAGAAGAACA<br>ATACAATTCTACTTATAGGGTGGTGTCTGTGCTGACAGTGC<br>TGCACCAGGATTGGCTGAATGGAAAAGAATATAAGTGTAA<br>GGTCTCTAACAAGGCCCTGCCCGCTCCAATTGAGAAGACA<br>ATTTCCAAGGCCAAGGGGCAGCCTCGGGAACCTCAGGTGT<br>ACACACTGCCCCCATCCAGGGATGAACTGACTAAAAATCA<br>GGTGTCTCTGACATGCCTGGTGAAAGGGTTTTATCCAAGTG<br>ACATTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGACGGG<br>TCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTCGG<br>TGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGATGCACGA<br>AGCTCTGAAATTTCACTATACACAGAAATCCCTGTCCCTGT<br>CTCCAGGCAAATAA |
| 신호 서열 및 돌연변이된 Fc 영역(H433K/N434F)을 갖는 SARS-CoV-2 S 단백질의 수용체 결합 도메인 | (서열번호 110)<br>MFVFLVLLPLVSSQCPNITNLCPFGEVFNATRFASVYAWNRKR<br>ISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF<br>VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD<br>SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG<br>FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAP*PKSCDKT*<br>*HTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSH*<br>*EDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQD*<br>*WLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE*<br>*LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG*<br>*SFFLYSKLTVDKSRWQQGNVFSCSVMHEAL**KF**HYTQKSLSLSPGK* |

[0266]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되며, 연장된 신호 서열을 함유하는 SARS-CoV-2 S 단백질 | (서열번호 122)<br>ATGTTCCTGCTGACAACAAAAAGAACCATGTTTGTGTTCCT<br>GGTGCTGCTGCCTCTGGTGTCCTCACAGTGTGTCAACCTGA<br>CAACAAGAACTCAGCTGCCACCAGCCTACACCAACTCCTTC<br>ACCAGAGGCGTGTATTACCCAGACAAGGTGTTTAGAAGCA<br>GCGTGCTGCACTCTACCCAGGACCTCTTTCTGCCCTTTTTCA<br>GCAACGTGACATGGTTTCACGCAATTCACGTGTCCGGCACT<br>AATGGCACAAAGCGGTTCGACAATCCAGTCCTGCCTTTCAA<br>CGATGGCGTCTACTTTGCATCTACTGAGAAATCCAATATCA<br>TTAGGGGATGGATCTTCGGCACAACCCTGGATTCTAAGACC<br>CAGAGCCTGCTGATCGTCAACAACGCCACAAACGTGGTCA<br>TTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCCTTTTCTGG<br>GCGTGTATTATCATAAGAACAATAAGAGCTGGATGGAGTC<br>CGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACCTTTG<br>AGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGGAAA<br>ACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTCAAA<br>AACATCGACGGCTATTTCAAGATCTATAGCAAGCATACCCC<br>AATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGCGCAC<br>TGGAGCCACTGGTTGACCTGCCTATCGGCATTAATATCACA<br>AGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATCTGAC<br>CCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCCGCTG<br>CCTACTATGTGGGCTATCTGCAGCCACGGACATTCCTGCTG<br>AAATACAATGAGAACGGGACAATCACAGATGCTGTTGATT<br>GCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCTCAAG<br>AGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCAAACTT<br>CAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCCTAATA<br>TCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACGCCACC<br>AGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAGGATCT<br>CTAACTGCGTCGCCGACTATTCCGTGCTGTATAACAGCGCC<br>TCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCGACAAA<br>ACTGAACGATCTCTGCTTTACAAATGTCTACGCCGACTCTT<br>TTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCACCAGG<br>ACAGACAGGCAAGATTGCTGACTACAACTATAAGCTGCCT<br>GACGACTTCACAGGATGTGTGATCGCATGGAACTCAAACA<br>ATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTGTAT<br>CGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAGGG<br>ACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTTGC<br>AATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAGTCT<br>TACGGGTTTCAGCCTACTAATGGAGTTGGGTACCAGCCATA<br>CAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTCCAG<br>CTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGTGAA<br>GAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCGGCA<br>CCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCCATTT<br>CAACAGTTTGGACGGGACATTGCCGACACCACCGATGCCG<br>TTCGGGATCCACAGACCCTGGAAATTCTGGACATTACACCG<br>TGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAACCA<br>ATACAAGCAACCAGGTTGCCGTCCTGTATCAGGATGTCAAT<br>TGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAGCTGA<br>CTCCCACATGGCGGGTGTATAGCACCGGATCCAACGTGTTT |

[0267]

| | |
|---|---|
| | CAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACGTGA<br>ATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGGCATT<br>TGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCTCCGC<br>CTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC<br>GGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGA<br>TGGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCT<br>CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGAT<br>TACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGAAGCTGCATTATACCTGA |

[0268]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987을 대체하도록 돌연변이되며, 연장된 신호 서열을 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 123)<br>MFLLTTKRTMFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPD<br>KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPV<br>LPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVV<br>IKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFE<br>YVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINL<br>VRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSG<br>WTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET<br>KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN<br>ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT<br>KLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPD<br>DFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS<br>TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV<br>VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLT<br>ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSV<br>ITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTG<br>SNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GS**<br>**AS**SVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPV<br>SMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV<br>EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS<br>FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLT<br>VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**<br>EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS<br>ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQ<br>EKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQI<br>ITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFK<br>NHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQ<br>ELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCL<br>KGCCSCGSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 124)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG |

[0269]

[0270]

| | |
|---|---|
| | CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTC<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCC |

| | |
|---|---|
| | ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGGCCCTGGCTTATACCTGA |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이된 SARS-CoV-2 S 단백질 | **(서열번호 125)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL |

[0271]

| | |
|---|---|
| | LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGV**ALA**YT |
| **퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이되며, 연장된 신호 서열을 함유하는, SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열** | **(서열번호 126)**<br>ATGTTCCTGCTGACAACAAAAAGAACCATGTTTGTGTTCCT<br>GGTGCTGCTGCCTCTGGTGTCCTCACAGTGTGTCAACCTGA<br>CAACAAGAACTCAGCTGCCACCAGCCTACACCAACTCCTTC<br>ACCAGAGGCGTGTATTACCCAGACAAGGTGTTTAGAAGCA<br>GCGTGCTGCACTCTACCCAGGACCTCTTTCTGCCCTTTTTCA<br>GCAACGTGACATGGTTTCACGCAATTCACGTGTCCGGCACT<br>AATGGCACAAAGCGGTTCGACAATCCAGTCCTGCCTTTCAA<br>CGATGGCGTCTACTTTGCATCTACTGAGAAATCCAATATCA<br>TTAGGGGATGGATCTTCGGCACAACCCTGGATTCTAAGACC<br>CAGAGCCTGCTGATCGTCAACAACGCCACAAACGTGGTCA<br>TTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCCTTTTCTGG<br>GCGTGTATTATCATAAGAACAATAAGAGCTGGATGGAGTC<br>CGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACCTTTG<br>AGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGGAAA<br>ACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTCAAA<br>AACATCGACGGCTATTTCAAGATCTATAGCAAGCATACCCC<br>AATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGCGCAC<br>TGGAGCCACTGGTTGACCTGCCTATCGGCATTAATATCACA<br>AGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATCTGAC<br>CCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCCGCTG<br>CCTACTATGTGGGCTATCTGCAGCCACGGACATTCCTGCTG<br>AAATACAATGAGAACGGGACAATCACAGATGCTGTTGATT<br>GCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCTCAAG<br>AGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCAAACTT<br>CAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCCTAATA<br>TCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACGCCACC<br>AGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAGGATCT<br>CTAACTGCGTCGCCGACTATTCCGTGCTGTATAACAGCGCC<br>TCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCGACAAA<br>ACTGAACGATCTCTGCTTTACAAATGTCTACGCCGACTCTT<br>TTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCACCAGG<br>ACAGACAGGCAAGATTGCTGACTACAACTATAAGCTGCCT<br>GACGACTTCACAGGATGTGTGATCGCATGGAACTCAAACA<br>ATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTGTAT<br>CGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAGGG<br>ACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTTGC<br>AATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAGTCT<br>TACGGGTTTCAGCCTACTAATGGAGTTGGGTACCAGCCATA |

[0272]

| | |
|---|---|
| | CAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTCCAG<br>CTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGTGAA<br>GAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCGGCA<br>CCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCCATTT<br>CAACAGTTTGGACGGGACATTGCCGACACCACCGATGCCG<br>TTCGGGATCCACAGACCCTGGAAATTCTGGACATTACACCG<br>TGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAACCA<br>ATACAAGCAACCAGGTTGCCGTCCTGTATCAGGATGTCAAT<br>TGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAGCTGA<br>CTCCCACATGGCGGGTGTATAGCACCGGATCCAACGTGTTT<br>CAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACGTGA<br>ATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGGCATT<br>TGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCTCCGC<br>CTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC<br>GGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGCAGA<br>TGGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCGCCT<br>CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGAT<br>TACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA |

[0273]

| | |
|---|---|
| | CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGGCCCTGGCTTATACCTGA |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이되며, 연장된 신호 서열을 함유하는, SARS-CoV-2 S 단백질 | **(서열번호 127)**<br>MFLLTTKRTMFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPD<br>KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPV<br>LPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVV<br>IKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFE<br>YVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINL<br>VRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSG<br>WTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET<br>KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN<br>ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT<br>KLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPD<br>DFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS<br>TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV<br>VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLT<br>ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSV<br>ITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTG<br>SNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GS**<br>**AS**SVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPV<br>SMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV<br>EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS<br>FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLT<br>VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**<br>EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS<br>ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQ<br>EKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQI<br>ITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFK<br>NHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQ<br>ELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCL<br>KGCCSCGSCCKFDEDDSEPVLKGV**ALA**YT |
| 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 128)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC |

[0274]

[0275]

| | |
|---|---|
| | ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA |

| | |
|---|---|
| | AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCC<br>CTATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGAC<br>GCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACAT<br>TGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCC<br>TCACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCT<br>CAATACACTAGCGCACTGCTGGCCGGAACCATCACATCAG<br>GCTGGACCTTCGGGGCCGGACCAGCACTGCAGATTCCATTC<br>CCTATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCAC<br>ACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAAC<br>CAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAG<br>CTCAACCCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |
| 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질 | **(서열번호 129)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE |

[0276]

| | |
|---|---|
| | LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| **퓨린 절단 부위를 제거하고 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열** | **(서열번호 130)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA |

[0277]

| | |
|---|---|
| | GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC<br>GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCCCT<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGACCAGCACTGCAGATTCCATTCCCT<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT |

[0278]

| | |
|---|---|
| | ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질 | **(서열번호 131)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |

[0279]

| | |
|---|---|
| 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되고, D614G 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 132)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGGCGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGC<br>AGACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGA<br>TCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGC |

[0280]

| | |
|---|---|
| | CGAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATT<br>GGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTC<br>TCCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTATTG<br>CCTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTAC<br>TCCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCT<br>GTGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAG<br>CGTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAAT<br>GTTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAG<br>CTGAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACA<br>AGAACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTA<br>TAAGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCT<br>CACAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAG<br>CCCTATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAG<br>ACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGAC<br>ATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGG<br>CCTCACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCG<br>CTCAATACACTAGCGCACTGCTGGCCGGAACCATCACATCA<br>GGCTGGACCTTCGGGGCCGGACCAGCACTGCAGATTCCATT<br>CCCTATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCA<br>CACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAA<br>CCAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCA<br>GCTCAACCCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTC<br>AACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGC<br>TGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGAC<br>ATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGAT<br>TGACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACAT<br>ACGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGC<br>ATCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTG<br>CTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCT<br>ACCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTT<br>GTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAA<br>CTTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCC<br>ACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACAC<br>TGGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCAT<br>CACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCG<br>TGATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAG<br>CCAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTT<br>TAAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATC<br>TCCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGA<br>TTGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCT<br>CTGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATAT<br>CAAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGAC<br>TGATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATG<br>ACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGG<br>CTCTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGC<br>TGAAGGGCGTGAAGCTGCATTATACCTGA |

[0281]

| | |
|---|---|
| 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되고, D614G 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 133)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQ<br>TRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSV<br>ASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMT<br>KTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQD<br>KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IED<br>LLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPP<br>LLTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYR<br>FNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되고, D614G 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 134)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG |

[0282]

[0283]

| | |
|---|---|
| | CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGGCGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGC<br>AGACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGA<br>TCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGC<br>CGAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATT<br>GGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTC<br>TCCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGC<br>CTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACT<br>CCAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTG<br>TGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGC<br>GTTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATG<br>TTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCT<br>GAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAG<br>AACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATA<br>AGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCA<br>CAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCC<br>CTATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGAC<br>GCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACAT<br>TGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCC<br>TCACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCT<br>CAATACACTAGCGCACTGCTGGCCGGAACCATCACATCAG<br>GCTGGACCTTCGGGGCCGGACCAGCACTGCAGATTCCATTC |

| | |
|---|---|
| | CCTATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCAC<br>ACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAAC<br>CAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAG<br>CTCAACCCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCA<br>ACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCT<br>GTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACA<br>TTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATT<br>GACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATA<br>CGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCA<br>TCCGCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGC<br>TGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTA<br>CCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTG<br>TTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAAC<br>TTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCA<br>CTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACT<br>GGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATC<br>ACCACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGT<br>GATCGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGC<br>CAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTT<br>AAGAACCACACAAGCCCAGATGTGGATCTCGGGGACATCT<br>CCGGAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGAT<br>TGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTC<br>TGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATC<br>AAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACT<br>GATTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGA<br>CCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCT<br>CTTGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTG<br>AAGGGCGTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되고, D614G 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | **(서열번호 135)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQ<br>TRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**PGSAS**SVA<br>SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTK<br>TSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL |

[0284]

[0285]

| | |
|---|---|
| | LTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVKLHYT |
| **퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되며, 연장된 신호 서열을 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열** | **(서열번호 136)**<br>ATGTTCCTGCTGACAACAAAAAGAACCATGTTTGTGTTCCT<br>GGTGCTGCTGCCTCTGGTGTCCTCACAGTGTGTCAACCTGA<br>CAACAAGAACTCAGCTGCCACCAGCCTACACCAACTCCTTC<br>ACCAGAGGCGTGTATTACCCAGACAAGGTGTTTAGAAGCA<br>GCGTGCTGCACTCTACCCAGGACCTCTTTCTGCCCTTTTTCA<br>GCAACGTGACATGGTTTCACGCAATTCACGTGTCCGGCACT<br>AATGGCACAAAGCGGTTCGACAATCCAGTCCTGCCTTTCAA<br>CGATGGCGTCTACTTTGCATCTACTGAGAAATCCAATATCA<br>TTAGGGGATGGATCTTCGGCACAACCCTGGATTCTAAGACC<br>CAGAGCCTGCTGATCGTCAACAACGCCACAAACGTGGTCA<br>TTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCCTTTTCTGG<br>GCGTGTATTATCATAAGAACAATAAGAGCTGGATGGAGTC<br>CGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACCTTTG<br>AGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGGAAA<br>ACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTCAAA<br>AACATCGACGGCTATTTCAAGATCTATAGCAAGCATACCCC<br>AATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGCGCAC<br>TGGAGCCACTGGTTGACCTGCCTATCGGCATTAATATCACA<br>AGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATCTGAC<br>CCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCCGCTG<br>CCTACTATGTGGGCTATCTGCAGCCACGGACATTCCTGCTG<br>AAATACAATGAGAACGGGACAATCACAGATGCTGTTGATT<br>GCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCTCAAG<br>AGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCAAACTT<br>CAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCCTAATA<br>TCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACGCCACC<br>AGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAGGATCT<br>CTAACTGCGTCGCCGACTATTCCGTGCTGTATAACAGCGCC<br>TCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCGACAAA<br>ACTGAACGATCTCTGCTTTACAAATGTCTACGCCGACTCTT<br>TTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCACCAGG<br>ACAGACAGGCAAGATTGCTGACTACAACTATAAGCTGCCT<br>GACGACTTCACAGGATGTGTGATCGCATGGAACTCAAACA<br>ATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTGTAT<br>CGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAGGG<br>ACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTTGC<br>AATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAGTCT<br>TACGGGTTTCAGCCTACTAATGGAGTTGGGTACCAGCCATA |

| | |
|---|---|
| | CAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTCCAG<br>CTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGTGAA<br>GAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCGGCA<br>CCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCCATTT<br>CAACAGTTTGGACGGGACATTGCCGACACCACCGATGCCG<br>TTCGGGATCCACAGACCCTGGAAATTCTGGACATTACACCG<br>TGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAACCA<br>ATACAAGCAACCAGGTTGCCGTCCTGTATCAGGATGTCAAT<br>TGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAGCTGA<br>CTCCCACATGGCGGGTGTATAGCACCGGATCCAACGTGTTT<br>CAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACGTGA<br>ATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGGCATT<br>TGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCTCCGC<br>CTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCCCTATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC<br>GGGGCCGGACCAGCACTGCAGATTCCATTCCCTATGCAGAT<br>GGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCCCCT<br>CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGAT<br>TACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA |

[0286]

| | |
|---|---|
| | CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하도록 돌연변이되며, 연장된 신호 서열을 함유하는 SARS-CoV-2 S 단백질 | **(서열번호 137)**<br>MFLLTTKRTMFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTR<br>GVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTK<br>RFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVN<br>NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSS<br>ANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYS<br>KHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLT<br>PGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDC<br>ALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLC<br>PFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFK<br>CYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADY<br>NYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLK<br>PFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGY<br>QPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLT<br>GTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS<br>FGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTW<br>RVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQT<br>QTNSP**GSAS**SVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTIS<br>VTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLN<br>RALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPD<br>PSKPSKRS**P**IEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICA<br>QKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQ<br>IPF**P**MQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSL<br>SST**P**SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDIL<br>SRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLA<br>ATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVT<br>YVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRN<br>FYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEEL<br>DKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE<br>SLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMT<br>SCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |

[0287]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 138)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCATA<br>GAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGACT<br>GCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCACG<br>GACATTCCTGCTGAAATACAATGAGAACGGGACAATCACA<br>GATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACAAA<br>GTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTATC<br>AGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCGTG<br>CGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAAGT<br>GTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAACA<br>GGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGCTG<br>TATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGAGT<br>GAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGTCT<br>ACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGCA<br>GATCGCACCAGGACAGACAGGCAAGATTGCTGACTACAAC<br>TATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCATG<br>GAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTAT<br>AATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGCC<br>CTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGGC<br>TCCACCCCTTGCAATGGCGTCGAAGGCTTTAATTGTTATTTT<br>CCCCTGCAGTCTTACGGGTTTCAGCCTACTAATGGAGTTGG<br>GTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCTCC<br>TGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCACT<br>AACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAACGG<br>GCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGAAG<br>TTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGACAC<br>CACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCTGG<br>ACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATCAC<br>ACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGTATC<br>AGGATGTCAATTGCACAGAAGTGCCAGTTGCTATCCACGCA<br>GACCAGCTGACTCCCACATGGCGGGTGTATAGCACCGGAT<br>CCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGGGGCC |

[0288]

| | |
|---|---|
| | GAGCACGTGAATAACAGCTACGAGTGCGACATCCCCATTG<br>GCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAACTCT<br>CCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCC<br>TATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTC<br>CAATAATTCCATCGCAATCCCTACTAACTTCACTATTTCTGT<br>GACCACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCG<br>TTGATTGTACCATGTATATTTGTGGCGACTCTACCGAATGTT<br>CTAACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTG<br>AACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGA<br>ACACACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAA<br>GACCCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCAC<br>AGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCCCT<br>ATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGC<br>CGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTG<br>CTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTC<br>ACAGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCA<br>ATACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCT<br>GGACCTTCGGGGCCGGACCAGCACTGCAGATTCCATTCCCT<br>ATGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACA<br>GAACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAG<br>TTTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTC<br>AACCCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAAC<br>CAGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTC<br>CTCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>GAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACTGACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGGCCCTGGCTTATACCTGA |

[0289]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이된 SARS-CoV-2 S 단백질 | (서열번호 139)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD<br>LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT<br>AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKC<br>TLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATR<br>FASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN<br>DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG<br>CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQ<br>AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE<br>LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK<br>FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTN<br>TSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT<br>RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQ<br>IDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG<br>QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT<br>APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT<br>FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD<br>VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE<br>QYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC<br>GSCCKFDEDDSEPVLKGVALAYT |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이되며, 연장된 신호 서열을 함유하는, SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 140)<br>ATGTTCCTGCTGACAACAAAAAGAACCATGTTTGTGTTCCT<br>GGTGCTGCTGCCTCTGGTGTCCTCACAGTGTGTCAACCTGA<br>CAACAAGAACTCAGCTGCCACCAGCCTACACCAACTCCTTC<br>ACCAGAGGCGTGTATTACCCAGACAAGGTGTTTAGAAGCA<br>GCGTGCTGCACTCTACCCAGGACCTCTTTCTGCCCTTTTTCA<br>GCAACGTGACATGGTTTCACGCAATTCACGTGTCCGGCACT<br>AATGGCACAAAGCGGTTCGACAATCCAGTCCTGCCTTTCAA<br>CGATGGCGTCTACTTTGCATCTACTGAGAAATCCAATATCA<br>TTAGGGGATGGATCTTCGGCACAACCCTGGATTCTAAGACC<br>CAGAGCCTGCTGATCGTCAACAACGCCACAAACGTGGTCA<br>TTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCCTTTTCTGG<br>GCGTGTATTATCATAAGAACAATAAGAGCTGGATGGAGTC<br>CGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACCTTTG<br>AGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGGAAA<br>ACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTCAAA<br>AACATCGACGGCTATTTCAAGATCTATAGCAAGCATACCCC<br>AATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGCGCAC |

[0290]

| | |
|---|---|
| | TGGAGCCACTGGTTGACCTGCCTATCGGCATTAATATCACA<br>AGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATCTGAC<br>CCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCCGCTG<br>CCTACTATGTGGGCTATCTGCAGCCACGGACATTCCTGCTG<br>AAATACAATGAGAACGGGACAATCACAGATGCTGTTGATT<br>GCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCTCAAG<br>AGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCAAACTT<br>CAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCCTAATA<br>TCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACGCCACC<br>AGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAGGATCT<br>CTAACTGCGTCGCCGACTATTCCGTGCTGTATAACAGCGCC<br>TCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCGACAAA<br>ACTGAACGATCTCTGCTTTACAAATGTCTACGCCGACTCTT<br>TTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCACCAGG<br>ACAGACAGGCAAGATTGCTGACTACAACTATAAGCTGCCT<br>GACGACTTCACAGGATGTGTGATCGCATGGAACTCAAACA<br>ATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTGTAT<br>CGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAGGG<br>ACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTTGC<br>AATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAGTCT<br>TACGGGTTTCAGCCTACTAATGGAGTTGGGTACCAGCCATA<br>CAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTCCAG<br>CTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGTGAA<br>GAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCGGCA<br>CCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCCATTT<br>CAACAGTTTGGACGGGACATTGCCGACACCACCGATGCCG<br>TTCGGGATCCACAGACCCTGGAAATTCTGGACATTACACCG<br>TGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAACCA<br>ATACAAGCAACCAGGTTGCCGTCCTGTATCAGGATGTCAAT<br>TGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAGCTGA<br>CTCCCACATGGCGGGTGTATAGCACCGGATCCAACGTGTTT<br>CAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACGTGA<br>ATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGGCATT<br>TGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCTCCGC<br>CTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCATGAG<br>CCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAATTCCA<br>TCGCAATCCCTACTAACTTCACTATTTCTGTGACCACCGAG<br>ATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATTGTAC<br>CATGTATATTTGTGGCGACTCTACCGAATGTTCTAACCTGC<br>TGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAGAGCC<br>CTGACTGGGATCGCTGTGGAGCAGGACAAGAACACACAGG<br>AGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCCCTCCT<br>ATTAAGGATTTCGGCGGATTCAATTTCTCACAGATTCTGCC<br>AGACCCCAGTAAGCCTTCCAAGAGGAGCCCTATCGAGGAT<br>CTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCTTTAT<br>TAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCCAGA<br>GACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGTGCT<br>GCCACCTCTGCTGACCGACGAGATGATCGCTCAATACACTA<br>GCGCACTGCTGGCCGGAACCATCACATCAGGCTGGACCTTC |

[0291]

| | |
|---|---|
| | GGGGCCGGACCAGCACTGCAGATTCCATTCCCTATGCAGAT<br>GGCCTATAGATTCAACGGCATTGGCGTCACACAGAACGTG<br>CTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTAATTC<br>CGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACCCCCT<br>CTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGAATGC<br>TCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCTAACT<br>TTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAGCCGC<br>CTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCCTGAT<br>TACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACCCAGC<br>AGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAAATCT<br>GGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCAGTCC<br>AAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGATGA<br>GCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTGCAC<br>GTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAACTGC<br>TCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCACGGG<br>AGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTGACC<br>CAGAGGAACTTCTATGAACCCCAGATCATCACCACTGACA<br>ATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCATC<br>GTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTGGA<br>CTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCACA<br>CAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATTAA<br>CGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCCTA<br>AATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGATCT<br>GCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGCCC<br>TGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCCAT<br>CGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTGTT<br>GTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTGT<br>AAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGCG<br>TGGCCCTGGCTTATACCTGA |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899, 942, 986 및 987 을 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이되며, 연장된 신호 서열을 함유하는, SARS-CoV-2 S 단백질 | **(서열번호 141)**<br>MFLLTTKRTMFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTR<br>GVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTK<br>RFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVN<br>NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSS<br>ANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYS<br>KHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLT<br>PGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDC<br>ALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLC<br>PFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFK<br>CYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADY<br>NYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLK<br>PFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGY<br>QPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLT<br>GTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS<br>FGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTW<br>RVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQT<br>QTNSP**GSAS**SVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTIS<br>VTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLN<br>RALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPD |

[0292]

| | |
|---|---|
| | PSKPSKRS**P**IEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICA<br>QKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAG**P**ALQ<br>IPF**P**MQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSL<br>SST**P**SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDIL<br>SRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLA<br>ATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVT<br>YVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRN<br>FYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEEL<br>DKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE<br>SLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMT<br>SCCSCLKGCCSCGSCCKFDEDDSEPVLKGVALAYT |
| H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 150)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTTCC<br>GGCACTAATGGCACAAAGCGGTTCGACAATCCAGTCCTGC<br>CTTTCAACGATGGCGTCTACTTTGCATCTACTGAGAAATCC<br>AATATCATTAGGGGATGGATCTTCGGCACAACCCTGGATTC<br>TAAGACCCAGAGCCTGCTGATCGTCAACAACGCCACAAAC<br>GTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCC<br>TTTTCTGGGCGTTTACCATAAGAACAATAAGAGCTGGATGG<br>AGTCCGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACC<br>TTTGAGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGG<br>AAAACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTC<br>AAAAACATCGACGGCTATTTCAAGATCTATAGCAAGCATA<br>CCCCAATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGC<br>GCACTGGAGCCACTGGTTGACCTGCCTATCGGCATTAATAT<br>CACAAGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGCAAGATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAT**GGAGTTGGGTACCAGCC |

[0293]

[0294]

| | |
|---|---|
| | ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATT**GAT**GACACCACCGAT<br>GCCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTAC<br>ACCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGA<br>ACCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGA**G<br>TCAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCT**CAT**AGAA<br>GGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGCCTATACC<br>ATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATA<br>ATTCCATCGCAATCCCT**ATA**AACTTCACTATTTCTGTGACC<br>ACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGA<br>TTGTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTA<br>ACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAAC<br>AGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACA<br>CACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGAC<br>CCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGA<br>TTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATC<br>GAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCG<br>GCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCT<br>GCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCAC<br>AGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAAT<br>ACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTG<br>GACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCA<br>TGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACAG<br>AACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGT<br>TTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCA<br>ACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACC<br>AGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCC<br>TCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>G**GCA**CGCCTGGATAAGGTGGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACT**CAC**AATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG |

| | |
|---|---|
| | GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |
| H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하도록 돌연변이된 SARS-CoV-2 S 단백질 | **(서열번호151)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAISGTNGTKRFDNPVLPFND<br>GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE<br>FQFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVSQPF<br>LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ<br>GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDI**D**DTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**H**RRARSVAS<br>QSIIAYTMSLGAENSVAYSNNSIAI**P**INFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL<br>TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDIL**A**RLDKVEAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITT**H**NTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT* |

[0295]

| | |
|---|---|
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 152)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTTCC<br>GGCACTAATGGCACAAAGCGGTTCGACAATCCAGTCCTGC<br>CTTTCAACGATGGCGTCTACTTTGCATCTACTGAGAAATCC<br>AATATCATTAGGGGATGGATCTTCGGCACAACCCTGGATTC<br>TAAGACCCAGAGCCTGCTGATCGTCAACAACGCCACAAAC<br>GTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCC<br>TTTTCTGGGCGTTTACCATAAGAACAATAAGAGCTGGATGG<br>AGTCCGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACC<br>TTTGAGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGG<br>AAAACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTC<br>AAAAACATCGACGGCTATTTCAAGATCTATAGCAAGCATA<br>CCCCAATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGC<br>GCACTGGAGCCACTGGTTGACCTGCCTATCGGCATTAATAT<br>CACAAGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGCAAGATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAT**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATT**GAT**GACACCACCGAT<br>GCCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTAC<br>ACCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGA<br>ACCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGA**G<br>TCAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG |

[0296]

| | |
|---|---|
| | TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCT**CAT**AGAA<br>GGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGCCTATACC<br>ATGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATA<br>ATTCCATCGCAATCCCT**ATA**AACTTCACTATTTCTGTGACC<br>ACCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGA<br>TTGTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTA<br>ACCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAAC<br>AGAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACA<br>CACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGAC<br>CCCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGA<br>TTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATC<br>GAGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCG<br>GCTTTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCT<br>GCCAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCAC<br>AGTGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAAT<br>ACACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTG<br>GACCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCA<br>TGCAGATGGCCTATAGATTCAACGGCATTGGCGTCACACAG<br>AACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGT<br>TTAATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCA<br>ACCGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACC<br>AGAATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCC<br>TCTAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT<br>G**GCA**CGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGAC<br>CGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGT<br>GACCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCC<br>GCAAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGG<br>GCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCA<br>CCTGATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTT<br>TTCTGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTT<br>ACAACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTT<br>CCCACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGT<br>TCGTGACCCAGAGGAACTTCTATGAACCCCAGATCATCACC<br>ACT**CAC**AATACCTTCGTGTCTGGAAATTGCGACGTCGTGAT<br>CGGCATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAG<br>AGCTGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAG<br>AACCACACAAGCCCAGATGTGGATCTCGGGGACATCTCCG<br>GAATTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGA<br>CCGCCTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGA<br>TTGATCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAA<br>ATGGCCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGA<br>TTGCCATCGTCATGGTGACCATCATGCTGTGTTGCATGACC<br>TCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCT<br>TGCTGTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGA<br>AGGGCGTGAAGCTGCATTATACCTGA |

[0297]

| | |
|---|---|
| 프롤린으로 돌연변이된 잔기 986 및 987 을 갖고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 153)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAISGTNGTKRFDNPVLPFND<br>GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE<br>FQFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVSQPF<br>LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ<br>GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDI**D**DTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**H**RRARSVAS<br>QSIIAYTMSLGAENSVAYSNNSIAI**P**INFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL<br>TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDIL**A**RLD**PP**EAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITT**H**NTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |
| 활성화에 필요한 퓨린 절단 부위를 제거하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 154)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTTCC<br>GGCACTAATGGCACAAAGCGGTTCGACAATCCAGTCCTGC<br>CTTTCAACGATGGCGTCTACTTTGCATCTACTGAGAAATCC<br>AATATCATTAGGGGATGGATCTTCGGCACAACCCTGGATTC<br>TAAGACCCAGAGCCTGCTGATCGTCAACAACGCCACAAAC<br>GTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCC<br>TTTTCTGGGCGTTTACCATAAGAACAATAAGAGCTGGATGG<br>AGTCCGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACC<br>TTTGAGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGG<br>AAAACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTC<br>AAAAACATCGACGGCTATTTCAAGATCTATAGCAAGCATA<br>CCCCAATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGC<br>GCACTGGAGCCACTGGTTGACCTGCCTATCGGCATTAATAT |

[0298]

| | |
|---|---|
| | CACAAGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGCAAGATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAT**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATT**GAT**GACACCACCGAT<br>GCCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTAC<br>ACCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGA<br>ACCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGA**G<br>TCAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCT**CAT**GGCT<br>CCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCA<br>TGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAAT<br>TCCATCGCAATCCCT**ATA**AACTTCACTATTTCTGTGACCAC<br>CGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATT<br>GTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAAC<br>CTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAG<br>AGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACACA<br>CAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCC<br>CTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGATT<br>CTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGA<br>GGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCT<br>TTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCC<br>AGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGT<br>GCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATACA<br>CTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGAC<br>CTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGC |

[0299]

| | |
|---|---|
| | AGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGAA<br>CGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTA<br>ATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACC<br>GCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGA<br>ATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCT<br>AACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTG**GC**<br>**A**CGCCTGGATAAGGTGGAGGCTGAAGTCCAGATTGACCGC<br>CTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGAC<br>CCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCA<br>AATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCC<br>AGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTG<br>ATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCT<br>GCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAA<br>CTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCA<br>CGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGT<br>GACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACT**C**<br>**AC**AATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGC<br>ATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCT<br>GGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACC<br>ACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAAT<br>TAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGC<br>CTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGA<br>TCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGG<br>CCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGC<br>CATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCT<br>GTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCT<br>GTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGG<br>CGTGAAGCTGCATTATACCTGA |
| **활성화에 필요한 퓨린 절단 부위를 제거하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질** | **(서열번호 155)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAISGTNGTKRFDNPVLPFND<br>GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE<br>FQFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVSQPF<br>LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ<br>GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDI**D**DTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**HG**SA**S**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIP**I**NFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL |

[0300]

| | |
|---|---|
| | TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDILARLDKVEAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTHNTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 156)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTTCC<br>GGCACTAATGGCACAAAGCGGTTCGACAATCCAGTCCTGC<br>CTTTCAACGATGGCGTCTACTTTGCATCTACTGAGAAATCC<br>AATATCATTAGGGGATGGATCTTCGGCACAACCCTGGATTC<br>TAAGACCCAGAGCCTGCTGATCGTCAACAACGCCACAAAC<br>GTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCC<br>TTTTCTGGGCGTTTACCATAAGAACAATAAGAGCTGGATGG<br>AGTCCGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACC<br>TTTGAGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGG<br>AAAACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTC<br>AAAAACATCGACGGCTATTTCAAGATCTATAGCAAGCATA<br>CCCCAATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGC<br>GCACTGGAGCCACTGGTTGACCTGCCTATCGGCATTAATAT<br>CACAAGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGCAAGATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAT**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC |

[0301]

[0302]

| | |
|---|---|
| | CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATT**GAT**GACACCACCGAT<br>GCCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTAC<br>ACCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGA<br>ACCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGA**G<br>TCAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCT**CAT**GGCT<br>CCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCA<br>TGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAAT<br>TCCATCGCAATCCCT**ATA**AACTTCACTATTTCTGTGACCAC<br>CGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATT<br>GTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAAC<br>CTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAG<br>AGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACACA<br>CAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCC<br>CTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGATT<br>CTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGA<br>GGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCT<br>TTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCC<br>AGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGT<br>GCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATACA<br>CTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGAC<br>CTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGC<br>AGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGAA<br>CGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTA<br>ATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACC<br>GCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGA<br>ATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCT<br>AACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCT**GGC**<br>**A**CGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGC<br>CTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGAC<br>CCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCA<br>AATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCC<br>AGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTG<br>ATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCT<br>GCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAA<br>CTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCA<br>CGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGT<br>GACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACT**C**<br>**AC**AATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGC<br>ATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCT<br>GGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACC<br>ACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAAT<br>TAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGC |

| | |
|---|---|
| | CTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGA<br>TCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGG<br>CCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGC<br>CATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCT<br>GTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCT<br>GTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGG<br>CGTGAAGCTGCATTATACCTGA |
| **퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질** | **(서열번호 157)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAISGTNGTKRFDNPVLPFND<br>GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE<br>FQFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVSQPF<br>LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ<br>GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDI**D**DTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**H**GSASSVAS<br>QSIIAYTMSLGAENSVAYSNNSIAIP**I**NFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL<br>TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDIL**A**RLD**PP**EAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITT**H**NTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |

[0303]

[0304]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899 및 942, 986 및 987 을 대체하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 158)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTTCC<br>GGCACTAATGGCACAAAGCGGTTCGACAATCCAGTCCTGC<br>CTTTCAACGATGGCGTCTACTTTGCATCTACTGAGAAATCC<br>AATATCATTAGGGGATGGATCTTCGGCACAACCCTGGATTC<br>TAAGACCCAGAGCCTGCTGATCGTCAACAACGCCACAAAC<br>GTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAACGATCC<br>TTTTCTGGGCGTTTACCATAAGAACAATAAGAGCTGGATGG<br>AGTCCGAGTTTAGAGTGTATAGCTCTGCAAATAATTGTACC<br>TTTGAGTACGTGAGCCAGCCCTTTCTGATGGACCTGGAGGG<br>AAAACAAGGAAACTTCAAAAACCTGCGGGAATTCGTTTTC<br>AAAAACATCGACGGCTATTTCAAGATCTATAGCAAGCATA<br>CCCCAATCAACCTCGTGAGGGACCTCCCCCAGGGCTTTAGC<br>GCACTGGAGCCACTGGTTGACCTGCCTATCGGCATTAATAT<br>CACAAGATTTCAGACCCTGCTGGCACTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGCAAGATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTCGAAGGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAT**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATT**GAT**GACACCACCGAT<br>GCCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTAC<br>ACCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGA<br>ACCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGC**G<br>TCAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG |

| | |
|---|---|
| | TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCT**CAT**GGCT<br>CCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCA<br>TGAGCCTCGGAGCTGAGAATAGCGTGGCCTACTCCAATAAT<br>TCCATCGCAATCCCT**ATA**AACTTCACTATTTCTGTGACCAC<br>CGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATT<br>GTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAAC<br>CTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAG<br>AGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACACA<br>CAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCC<br>CTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGATT<br>CTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCCCTATCG<br>AGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGC<br>TTTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGC<br>CAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAG<br>TGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATAC<br>ACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGA<br>CCTTCGGGGCCGGACCAGCACTGCAGATTCCATTCCCTATG<br>CAGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGA<br>ACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTT<br>AATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAAC<br>CCCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAG<br>AATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTC<br>TAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTG**G**<br>**CAC**GCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCG<br>CCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGA<br>CCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGC<br>AAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGC<br>CAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCT<br>GATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTC<br>TGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACA<br>ACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCC<br>ACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCG<br>TGACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACT<br>**CAC**AATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGG<br>CATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGC<br>TGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAAC<br>CACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAA<br>TTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGC<br>CTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGA<br>TCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGG<br>CCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGC<br>CATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCT<br>GTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCT<br>GTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGG<br>CGTGAAGCTGCATTATACCTGA |

[0305]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 817, 892, 899 및 942, 986 및 987 을 대체하도록 돌연변이되고, H69-, V70-, Y144-, N501Y, A570D, D614G, P681H, T716I, S982A 및 D1118H 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 159)<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAISGTNGTKRFDNPVLPFND<br>GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE<br>FQFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVSQPF<br>LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ<br>GFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDI**D**DTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS**HGSAS**SVAS<br>QSIIAYTMSLGAENSVAYSNNSIAI**P**I**N**FTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS**P**IEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL<br>TDEMIAQYTSALLAGTITSGWTFGAG**P**ALQIPF**P**MQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSST**P**SALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDIL**A**RLD**PP**EAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITT**H**NTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |
| D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 160)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTC**GCC**AATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGG**GGC**CTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG |

[0306]

[0307]

| | |
|---|---|
| | CATTAATATCACAAGATTTCAGACCCTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGC**AAC**ATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTC**AAG**GGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAC**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATTGCCGACACCACCGATG<br>CCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTACA<br>CCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAA<br>CCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGC**GT<br>CAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCTCCTAGAA<br>GGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGCCTATACC<br>ATGAGCCTCGGA**GTG**GAGAATAGCGTGGCCTACTCCAATA<br>ATTCCATCGCAATCCCTACTAACTTCACTATTTCTGTGACCA<br>CCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGAT<br>TGTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAA<br>CCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACA<br>GAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACAC<br>ACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACC<br>CCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGAT<br>TCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCG<br>AGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGC<br>TTTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGC<br>CAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAG<br>TGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATAC<br>ACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGA<br>CCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATG |

| | |
|---|---|
| | CAGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGA<br>ACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTT<br>AATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAAC<br>CGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAG<br>AATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTC<br>TAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGA<br>GCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATTGACCG<br>CCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGA<br>CCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGC<br>AAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGC<br>CAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCT<br>GATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTC<br>TGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACA<br>ACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCC<br>ACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCG<br>TGACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACT<br>GACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGG<br>CATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGC<br>TGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAAC<br>CACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAA<br>TTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGC<br>CTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGA<br>TCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGG<br>CCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGC<br>CATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCT<br>GTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCT<br>GTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGG<br>CGTGAAGCTGCATTATACCTGA |
| D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | **(서열번호 161)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRF**A**NPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVR**G**<br>LPQGFSALEPLVDLPIGINITRFQTLHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGN**I**ADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGV**K**GFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVAS<br>QSIIAYTMSLG**V**ENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL |

[0308]

| | |
|---|---|
| | LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987을 대체하도록 돌연변이되고, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 162)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACCTGACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTC**GCC**AATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGG**GGC**CTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGC**AAC**ATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTC**AAG**GGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAC**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC |

[0309]

| | |
|---|---|
| | CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATTGCCGACACCACCGATG<br>CCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTACA<br>CCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAA<br>CCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGC**GT<br>CAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCT<br>CCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCA<br>TGAGCCTCGGA**GTG**GAGAATAGCGTGGCCTACTCCAATAA<br>TTCCATCGCAATCCCTACTAACTTCACTATTTCTGTGACCAC<br>CGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATT<br>GTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAAC<br>CTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAG<br>AGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACACA<br>CAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCC<br>CTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGATT<br>CTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGA<br>GGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCT<br>TTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCC<br>AGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGT<br>GCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATACA<br>CTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGAC<br>CTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGC<br>AGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGAA<br>CGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTA<br>ATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACC<br>GCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGA<br>ATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCT<br>AACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAG<br>CCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCC<br>TGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACC<br>CAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAA<br>ATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCA<br>GTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGA<br>TGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTG<br>CACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAAC<br>TGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCAC<br>GGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTG<br>ACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACTGA<br>CAATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCA<br>TCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTG<br>GACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCA<br>CACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATT<br>AACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCC |

[0310]

| | |
|---|---|
| | TAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGAT<br>CTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGC<br>CCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCC<br>ATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTG<br>TTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTG<br>TAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGC<br>GTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | **(서열번호 163)**<br>MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRF**A**NPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVR**G**<br>LPQGFSALEPLVDLPIGINITRFQTLHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTG**N**IADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGV**K**GFNCYFPLQSYGFQPT**Y**GVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSP**GSAS**SVASQ<br>SIIAYTMSLG**V**ENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL<br>TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |

[0311]

| | |
|---|---|
| L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 164)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAAC**TTC**ACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTC**GCC**AATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGG**GGC**CTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG<br>CATTAATATCACAAGATTTCAGACCCTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGC**AAC**ATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTC**AAG**GGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAC**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATTGCCGACACCACCGATG<br>CCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTACA<br>CCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAA<br>CCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGC**GT<br>CAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG |

[0312]

| | |
|---|---|
| | TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCTCCTAGAA<br>GGGCCAGGTCCGTTGCTAGTCAGTCTATTATTGCCTATACC<br>ATGAGCCTCGGA**GTG**GAGAATAGCGTGGCCTACTCCAATA<br>ATTCCATCGCAATCCCTACTAACTTCACTATTTCTGTGACCA<br>CCGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGAT<br>TGTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAA<br>CCTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACA<br>GAGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACAC<br>ACAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACC<br>CCTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGAT<br>TCTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCG<br>AGGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGC<br>TTTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGC<br>CAGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAG<br>TGCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATAC<br>ACTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGA<br>CCTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATG<br>CAGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGA<br>ACGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTT<br>AATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAAC<br>CGCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAG<br>AATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTC<br>TAACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGA<br>GCCGCCTGGATAAGGTGGAGGCTGAAGTCCAGATTGACCG<br>CCTGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGA<br>CCCAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGC<br>AAATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGC<br>CAGTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCT<br>GATGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTC<br>TGCACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACA<br>ACTGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCC<br>ACGGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCG<br>TGACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACT<br>GACAATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGG<br>CATCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGC<br>TGGACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAAC<br>CACACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAA<br>TTAACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGC<br>CTAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGA<br>TCTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGG<br>CCCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGC<br>CATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCT<br>GTTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCT<br>GTAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGG<br>CGTGAAGCTGCATTATACCTGA |

[0313]

| | |
|---|---|
| L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 165)<br>MFVFLVLLPLVSSQCVNFTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFANPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRG<br>LPQGFSALEPLVDLPIGINITRFQTLHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGNIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVKGFNCYFPLQSYGFQPTYGVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVAS<br>QSIIAYTMSLGVENSVAYSNNSIAIPTNFTISVTTEILPVSMTKT<br>SVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK<br>NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPL<br>LTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF<br>NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ<br>DVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV<br>QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL<br>GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF<br>TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTD<br>NTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS<br>PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK<br>YEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC<br>SCGSCCKFDEDDSEPVLKGVKLHYT |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 166)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAACTTCACAACTAGGACTCAGCTGCCACCAGCCTA<br>CACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAGG<br>TGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTTT<br>CTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTCA<br>CGTGTCCGGCACTAATGGCACAAAGCGGTTCGCCAATCCA<br>GTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTGA<br>GAAATCCAATATCATTAGGGGATGGATCTTCGGCACAACCC<br>TGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACGCC<br>ACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGTAA<br>CGATCCTTTTCTGGGCGTGTATTATCATAAGAACAATAAGA<br>GCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAAAT<br>AATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATGGA<br>CCTGGAGGGAAAACAAGGAAACTTCAAAAACCTGCGGGAA<br>TTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTATAG<br>CAAGCATACCCCAATCAACCTCGTGAGGGGCCTCCCCCAG<br>GGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATCGG |

[0314]

[0315]

| | |
|---|---|
| | CATTAATATCACAAGATTTCAGACCCTGCATAGAAGCTATC<br>TGACCCCTGGAGACTCCTCTAGTGGGTGGACTGCCGGCGCC<br>GCTGCCTACTATGTGGGCTATCTGCAGCCACGGACATTCCT<br>GCTGAAATACAATGAGAACGGGACAATCACAGATGCTGTT<br>GATTGCGCACTCGACCCCCTGTCCGAGACAAAGTGCACTCT<br>CAAGAGCTTTACCGTCGAGAAGGGCATCTATCAGACCTCA<br>AACTTCAGGGTGCAGCCCACAGAATCTATCGTGCGCTTCCC<br>TAATATCACTAACCTGTGTCCTTTCGGTGAAGTGTTCAACG<br>CCACCAGGTTTGCTAGCGTGTATGCCTGGAACAGGAAGAG<br>GATCTCTAACTGCGTCGCCGACTATTCCGTGCTGTATAACA<br>GCGCCTCCTTCTCCACATTCAAATGCTATGGAGTGAGCCCG<br>ACAAAACTGAACGATCTCTGCTTTACAAATGTCTACGCCGA<br>CTCTTTTGTGATCAGAGGGGACGAGGTCCGGCAGATCGCAC<br>CAGGACAGACAGGC**AAC**ATTGCTGACTACAACTATAAGCT<br>GCCTGACGACTTCACAGGATGTGTGATCGCATGGAACTCAA<br>ACAATCTGGACTCCAAAGTCGGGGGCAACTATAATTACCTG<br>TATCGCCTGTTCCGGAAGTCCAACCTGAAGCCCTTCGAGAG<br>GGACATCAGTACAGAGATCTATCAGGCTGGCTCCACCCCTT<br>GCAATGGCGTC**AAG**GGCTTTAATTGTTATTTTCCCCTGCAG<br>TCTTACGGGTTTCAGCCTACT**TAC**GGAGTTGGGTACCAGCC<br>ATACAGAGTGGTCGTGCTCAGCTTCGAGCTCCTGCATGCTC<br>CAGCTACAGTTTGCGGGCCAAAGAAGTCCACTAACCTGGT<br>GAAGAATAAGTGCGTCAACTTCAACTTTAACGGGCTCACCG<br>GCACCGGCGTGCTGACTGAGAGCAACAAGAAGTTTCTGCC<br>ATTTCAACAGTTTGGACGGGACATTGCCGACACCACCGATG<br>CCGTTCGGGATCCACAGACCCTGGAAATTCTGGACATTACA<br>CCGTGCAGCTTCGGGGGCGTGAGCGTGATCACACCCGGAA<br>CCAATACAAGCAACCAGGTTGCCGTCCTGTATCAG**GGC**GT<br>CAATTGCACAGAAGTGCCAGTTGCTATCCACGCAGACCAG<br>CTGACTCCCACATGGCGGGTGTATAGCACCGGATCCAACGT<br>GTTTCAGACCCGCGCCGGATGTCTCATTGGGGCCGAGCACG<br>TGAATAACAGCTACGAGTGCGACATCCCCATTGGCGCCGG<br>CATTTGTGCGTCTTACCAGACTCAGACCAACTCTCCTGGCT<br>CCGCCTCTTCCGTTGCTAGTCAGTCTATTATTGCCTATACCA<br>TGAGCCTCGGA**GTG**GAGAATAGCGTGGCCTACTCCAATAA<br>TTCCATCGCAATCCCTACTAACTTCACTATTTCTGTGACCAC<br>CGAGATCCTGCCTGTGTCTATGACTAAGACTAGCGTTGATT<br>GTACCATGTATATTTGTGGCGACTCTACCGAATGTTCTAAC<br>CTGCTGCTTCAGTACGGCTCATTTTGCACACAGCTGAACAG<br>AGCCCTGACTGGGATCGCTGTGGAGCAGGACAAGAACACA<br>CAGGAGGTGTTTGCACAGGTGAAGCAGATCTATAAGACCC<br>CTCCTATTAAGGATTTCGGCGGATTCAATTTCTCACAGATT<br>CTGCCAGACCCCAGTAAGCCTTCCAAGAGGAGCTTCATCGA<br>GGATCTCCTGTTTAACAAGGTGACCCTGGCAGACGCCGGCT<br>TTATTAAGCAATATGGGGATTGCCTGGGCGACATTGCTGCC<br>AGAGACCTGATTTGCGCCCAGAAATTCAATGGCCTCACAGT<br>GCTGCCACCTCTGCTGACCGACGAGATGATCGCTCAATACA<br>CTAGCGCACTGCTGGCCGGAACCATCACATCAGGCTGGAC<br>CTTCGGGGCCGGAGCAGCACTGCAGATTCCATTCGCCATGC |

| | |
|---|---|
| | AGATGGCCTATAGATTCAACGGCATTGGCGTCACACAGAA<br>CGTGCTGTACGAAAACCAGAAGCTCATCGCTAACCAGTTTA<br>ATTCCGCAATTGGAAAGATCCAAGATTCACTCAGCTCAACC<br>GCCTCTGCACTCGGAAAGCTGCAGGACGTGGTCAACCAGA<br>ATGCTCAGGCCCTGAACACACTCGTCAAGCAGCTGTCCTCT<br>AACTTTGGCGCTATCAGCTCCGTTCTGAACGACATTCTGAG<br>CCGCCTGGATCCCCCAGAGGCTGAAGTCCAGATTGACCGCC<br>TGATTACCGGCCGGCTGCAGTCTCTGCAAACATACGTGACC<br>CAGCAGCTGATCAGAGCAGCCGAGATCCGGGCATCCGCAA<br>ATCTGGCAGCAACTAAGATGAGCGAATGCGTGCTGGGCCA<br>GTCCAAGCGGGTGGACTTTTGTGGCAAGGGCTACCACCTGA<br>TGAGCTTCCCCCAGAGCGCCCCACATGGCGTTGTTTTTCTG<br>CACGTGACCTATGTCCCTGCTCAGGAAAAGAACTTTACAAC<br>TGCTCCTGCTATCTGCCATGACGGCAAGGCCCACTTCCCAC<br>GGGAGGGAGTGTTTGTGTCCAATGGCACACACTGGTTCGTG<br>ACCCAGAGGAACTTCTATGAACCCCAGATCATCACCACTGA<br>CAATACCTTCGTGTCTGGAAATTGCGACGTCGTGATCGGCA<br>TCGTTAACAACACCGTGTACGACCCTCTCCAGCCAGAGCTG<br>GACTCCTTTAAGGAGGAACTGGATAAGTATTTTAAGAACCA<br>CACAAGCCCAGATGTGGATCTCGGGGACATCTCCGGAATT<br>AACGCCTCCGTGGTGAATATCCAGAAGGAGATTGACCGCC<br>TAAATGAAGTTGCCAAGAACCTCAATGAGTCTCTGATTGAT<br>CTGCAGGAACTGGGCAAGTATGAGCAGTATATCAAATGGC<br>CCTGGTACATTTGGCTGGGGTTTATCGCCGGACTGATTGCC<br>ATCGTCATGGTGACCATCATGCTGTGTTGCATGACCTCCTG<br>TTGTTCCTGTCTGAAGGGCTGCTGTAGTTGCGGCTCTTGCTG<br>TAAATTCGACGAAGATGATAGCGAGCCCGTGCTGAAGGGC<br>GTGAAGCTGCATTATACCTGA |
| 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 167)<br>MFVFLVLLPLVSSQCVNFTTRTQLPPAYTNSFTRGVYYPDKVF<br>RSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFANPVLPF<br>NDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV<br>CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVS<br>QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRG<br>LPQGFSALEPLVDLPIGINITRFQTLHRSYLTPGDSSSGWTAGA<br>AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK<br>SFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFAS<br>VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC<br>FTNVYADSFVIRGDEVRQIAPGQTGNIADYNYKLPDDFTGCVI<br>AWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG<br>STPCNGVKGFNCYFPLQSYGFQPTYGVGYQPYRVVVLSFELL<br>HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL<br>PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS<br>NQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTR<br>AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQ<br>SIIAYTMSLGVENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTS<br>VDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN<br>TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLL<br>FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL |

[0316]

| | |
|---|---|
| | TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN<br>GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD<br>VVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQI<br>DRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ<br>SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA<br>PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTF<br>VSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDV<br>DLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ<br>YIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCG<br>SCCKFDEDDSEPVLKGVKLHYT |
| L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 168)**<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAAC**TTT**ACA**AAC**AGGACTCAGCTGCCA**TCC**GCCT<br>ACACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAG<br>GTGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTT<br>TCTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTC<br>ACGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCC<br>AGTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTG<br>AGAAATCCAATATCATTAGGGGATGGATCTTCGGCACAAC<br>CCTGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACG<br>CCACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGT<br>AAC**TAC**CCTTTTCTGGGCGTGTATTATCATAAGAACAATAA<br>GAGCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAA<br>ATAATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATG<br>GACCTGGAGGGAAAACAAGGAAACTTCAAAAACCTG**AGC**<br>GAATTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTA<br>TAGCAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCC<br>AGGGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATC<br>GGCATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCA<br>TAGAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGA<br>CTGCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCA<br>CGGACATTCCTGCTGAAATACAATGAGAACGGGACAATCA<br>CAGATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACA<br>AAGTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTA<br>TCAGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCG<br>TGCGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAA<br>GTGTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAA<br>CAGGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGC<br>TGTATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGA<br>GTGAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGT<br>CTACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGC<br>AGATCGCACCAGGACAGACAGGC**ACC**ATTGCTGACTACAA<br>CTATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCAT<br>GGAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTA<br>TAATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGC<br>CCTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGG<br>CTCCACCCCTTGCAATGGCGTC**AAG**GGCTTTAATTGTTATT<br>TTCCCCTGCAGTCTTACGGGTTTCAGCCTACT**TAC**GGAGTT<br>GGGTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCT |

[0317]

| | |
|---|---|
| | CCTGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCA<br>CTAACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAAC<br>GGGCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGA<br>AGTTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGAC<br>ACCACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCT<br>GGACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATC<br>ACACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGT<br>ATCAG**GGC**GTCAATTGCACAGAAGTGCCAGTTGCTATCCA<br>CGCAGACCAGCTGACTCCCACATGGCGGGTGTATAGCACC<br>GGATCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGG<br>GGCCGAG**TAC**GTGAATAACAGCTACGAGTGCGACATCCCC<br>ATTGGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAA<br>CTCTCCTAGAAGGGCCAGGTCCGTTGCTAGTCAGTCTATTA<br>TTGCCTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCC<br>TACTCCAATAATTCCATCGCAATCCCTACTAACTTCACTATT<br>TCTGTGACCACCGAGATCCTGCCTGTGTCTATGACTAAGAC<br>TAGCGTTGATTGTACCATGTATATTTGTGGCGACTCTACCG<br>AATGTTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACA<br>CAGCTGAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGG<br>ACAAGAACACACAGGAGGTGTTTGCACAGGTGAAGCAGAT<br>CTATAAGACCCCTCCTATTAAGGATTTCGGCGGATTCAATT<br>TCTCACAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGG<br>AGCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGC<br>AGACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCG<br>ACATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAAT<br>GGCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGATGAT<br>CGCTCAATACACTAGCGCACTGCTGGCCGGAACCATCACAT<br>CAGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCC<br>ATTCGCCATGCAGATGGCCTATAGATTCAACGGCATTGGCG<br>TCACACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGC<br>TAACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCAC<br>TCAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTG<br>GTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGC<br>AGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAAC<br>GACATTCTGAGCCGCCTGGATAAGGTGGAGGCTGAAGTCC<br>AGATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAA<br>ACATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCC<br>GGGCATCCGCAAATCTGGCAGCA**ATC**AAGATGAGCGAATG<br>CGTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAG<br>GGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGG<br>CGTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAA<br>AGAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCAAG<br>GCCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCAC<br>ACACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGA<br>TCATCACCACTGACAATACCTTCGTGTCTGGAAATTGCGAC<br>GTCGTGATCGGCATCGTTAACAACACCGTGTACGACCCTCT<br>CCAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAG<br>TATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGGGG<br>ACATCTCCGGAATTAACGCCTCC**TTC**GTGAATATCCAGAAG |

[0318]

| | |
|---|---|
| | GAGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAATG<br>AGTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGCA<br>GTATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATCG<br>CCGGACTGATTGCCATCGTCATGGTGACCATCATGCTGTGT<br>TGCATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAG<br>TTGCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAGC<br>CCGTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |
| L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 169)<br>MFVFLVLLPLVSSQCVN**F**T**N**RTQLP**S**AYTNSFTRGVYYPDKV<br>FRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLP<br>FNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIK<br>VCEFQFCN**Y**PFLGVYYHKNNKSWMESEFRVYSSANNCTFEY<br>VSQPFLMDLEGKQGNFKNL**S**EFVFKNIDGYFKIYSKHTPINLV<br>RDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSG<br>WTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET<br>KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN<br>ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT<br>KLNDLCFTNVYADSFVIRGDEVRQIAPGQTG**T**IADYNYKLPD<br>DFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS<br>TEIYQAGSTPCNGV**K**GFNCYFPLQSYGFQPT**Y**GVGYQPYRVV<br>VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLT<br>ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSV<br>ITPGTNTSNQVAVLYQ**G**VNCTEVPVAIHADQLTPTWRVYSTG<br>SNVFQTRAGCLIGAE**Y**VNNSYECDIPIGAGICASYQTQTNSPRR<br>ARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILP<br>VSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIA<br>VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKR<br>SFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGL<br>TVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDK<br>VEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAA**I**KM<br>SECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPA<br>QEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEP<br>QIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYF<br>KNHTSPDVDLGDISGINAS**F**VNIQKEIDRLNEVAKNLNESLIDL<br>QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSC<br>LKGCCSCGSCCKFDEDDSEPVLKGVKLHYT |

[0319]

[0320]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이를 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 170)<br>ATGTTCGTCTTCCTCGTGCTGCTCCCACTCGTTTCTTCCCAG<br>TGTGTCAAC**TTT**ACA**AAC**AGGACTCAGCTGCCA**TCC**GCCT<br>ACACCAACTCCTTCACCAGAGGCGTGTATTACCCAGACAAG<br>GTGTTTAGAAGCAGCGTGCTGCACTCTACCCAGGACCTCTT<br>TCTGCCCTTTTTCAGCAACGTGACATGGTTTCACGCAATTC<br>ACGTGTCCGGCACTAATGGCACAAAGCGGTTCGACAATCC<br>AGTCCTGCCTTTCAACGATGGCGTCTACTTTGCATCTACTG<br>AGAAATCCAATATCATTAGGGGATGGATCTTCGGCACAAC<br>CCTGGATTCTAAGACCCAGAGCCTGCTGATCGTCAACAACG<br>CCACAAACGTGGTCATTAAGGTTTGCGAGTTTCAGTTCTGT<br>AAC**TAC**CCTTTTCTGGGCGTGTATTATCATAAGAACAATAA<br>GAGCTGGATGGAGTCCGAGTTTAGAGTGTATAGCTCTGCAA<br>ATAATTGTACCTTTGAGTACGTGAGCCAGCCCTTTCTGATG<br>GACCTGGAGGGAAAACAAGGAAACTTCAAAAACCTG**AGC**<br>GAATTCGTTTTCAAAAACATCGACGGCTATTTCAAGATCTA<br>TAGCAAGCATACCCCAATCAACCTCGTGAGGGACCTCCCCC<br>AGGGCTTTAGCGCACTGGAGCCACTGGTTGACCTGCCTATC<br>GGCATTAATATCACAAGATTTCAGACCCTGCTGGCACTGCA<br>TAGAAGCTATCTGACCCCTGGAGACTCCTCTAGTGGGTGGA<br>CTGCCGGCGCCGCTGCCTACTATGTGGGCTATCTGCAGCCA<br>CGGACATTCCTGCTGAAATACAATGAGAACGGGACAATCA<br>CAGATGCTGTTGATTGCGCACTCGACCCCCTGTCCGAGACA<br>AAGTGCACTCTCAAGAGCTTTACCGTCGAGAAGGGCATCTA<br>TCAGACCTCAAACTTCAGGGTGCAGCCCACAGAATCTATCG<br>TGCGCTTCCCTAATATCACTAACCTGTGTCCTTTCGGTGAA<br>GTGTTCAACGCCACCAGGTTTGCTAGCGTGTATGCCTGGAA<br>CAGGAAGAGGATCTCTAACTGCGTCGCCGACTATTCCGTGC<br>TGTATAACAGCGCCTCCTTCTCCACATTCAAATGCTATGGA<br>GTGAGCCCGACAAAACTGAACGATCTCTGCTTTACAAATGT<br>CTACGCCGACTCTTTTGTGATCAGAGGGGACGAGGTCCGGC<br>AGATCGCACCAGGACAGACAGGC**ACC**ATTGCTGACTACAA<br>CTATAAGCTGCCTGACGACTTCACAGGATGTGTGATCGCAT<br>GGAACTCAAACAATCTGGACTCCAAAGTCGGGGGCAACTA<br>TAATTACCTGTATCGCCTGTTCCGGAAGTCCAACCTGAAGC<br>CCTTCGAGAGGGACATCAGTACAGAGATCTATCAGGCTGG<br>CTCCACCCCTTGCAATGGCGTC**AAG**GGCTTTAATTGTTATT<br>TTCCCCTGCAGTCTTACGGGTTTCAGCCTACT**TAC**GGAGTT<br>GGGTACCAGCCATACAGAGTGGTCGTGCTCAGCTTCGAGCT<br>CCTGCATGCTCCAGCTACAGTTTGCGGGCCAAAGAAGTCCA<br>CTAACCTGGTGAAGAATAAGTGCGTCAACTTCAACTTTAAC<br>GGGCTCACCGGCACCGGCGTGCTGACTGAGAGCAACAAGA<br>AGTTTCTGCCATTTCAACAGTTTGGACGGGACATTGCCGAC<br>ACCACCGATGCCGTTCGGGATCCACAGACCCTGGAAATTCT<br>GGACATTACACCGTGCAGCTTCGGGGGCGTGAGCGTGATC<br>ACACCCGGAACCAATACAAGCAACCAGGTTGCCGTCCTGT<br>ATCAG**GGC**GTCAATTGCACAGAAGTGCCAGTTGCTATCCA<br>CGCAGACCAGCTGACTCCCACATGGCGGGTGTATAGCACC<br>GGATCCAACGTGTTTCAGACCCGCGCCGGATGTCTCATTGG |

| | |
|---|---|
| | GGCCGAG**TAC**GTGAATAACAGCTACGAGTGCGACATCCCC<br>ATTGGCGCCGGCATTTGTGCGTCTTACCAGACTCAGACCAA<br>CTCTCCTGGCTCCGCCTCTTCCGTTGCTAGTCAGTCTATTAT<br>TGCCTATACCATGAGCCTCGGAGCTGAGAATAGCGTGGCCT<br>ACTCCAATAATTCCATCGCAATCCCTACTAACTTCACTATTT<br>CTGTGACCACCGAGATCCTGCCTGTGTCTATGACTAAGACT<br>AGCGTTGATTGTACCATGTATATTTGTGGCGACTCTACCGA<br>ATGTTCTAACCTGCTGCTTCAGTACGGCTCATTTTGCACAC<br>AGCTGAACAGAGCCCTGACTGGGATCGCTGTGGAGCAGGA<br>CAAGAACACACAGGAGGTGTTTGCACAGGTGAAGCAGATC<br>TATAAGACCCCTCCTATTAAGGATTTCGGCGGATTCAATTT<br>CTCACAGATTCTGCCAGACCCCAGTAAGCCTTCCAAGAGGA<br>GCTTCATCGAGGATCTCCTGTTTAACAAGGTGACCCTGGCA<br>GACGCCGGCTTTATTAAGCAATATGGGGATTGCCTGGGCGA<br>CATTGCTGCCAGAGACCTGATTTGCGCCCAGAAATTCAATG<br>GCCTCACAGTGCTGCCACCTCTGCTGACCGACGAGATGATC<br>GCTCAATACACTAGCGCACTGCTGGCCGGAACCATCACATC<br>AGGCTGGACCTTCGGGGCCGGAGCAGCACTGCAGATTCCA<br>TTCGCCATGCAGATGGCCTATAGATTCAACGGCATTGGCGT<br>CACACAGAACGTGCTGTACGAAAACCAGAAGCTCATCGCT<br>AACCAGTTTAATTCCGCAATTGGAAAGATCCAAGATTCACT<br>CAGCTCAACCGCCTCTGCACTCGGAAAGCTGCAGGACGTG<br>GTCAACCAGAATGCTCAGGCCCTGAACACACTCGTCAAGC<br>AGCTGTCCTCTAACTTTGGCGCTATCAGCTCCGTTCTGAAC<br>GACATTCTGAGCCGCCTGGATCCCCCAGAGGCTGAAGTCCA<br>GATTGACCGCCTGATTACCGGCCGGCTGCAGTCTCTGCAAA<br>CATACGTGACCCAGCAGCTGATCAGAGCAGCCGAGATCCG<br>GGCATCCGCAAATCTGGCAGCA**ATC**AAGATGAGCGAATGC<br>GTGCTGGGCCAGTCCAAGCGGGTGGACTTTTGTGGCAAGG<br>GCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCACATGGC<br>GTTGTTTTTCTGCACGTGACCTATGTCCCTGCTCAGGAAAA<br>GAACTTTACAACTGCTCCTGCTATCTGCCATGACGGCAAGG<br>CCCACTTCCCACGGGAGGGAGTGTTTGTGTCCAATGGCACA<br>CACTGGTTCGTGACCCAGAGGAACTTCTATGAACCCCAGAT<br>CATCACCACTGACAATACCTTCGTGTCTGGAAATTGCGACG<br>TCGTGATCGGCATCGTTAACAACACCGTGTACGACCCTCTC<br>CAGCCAGAGCTGGACTCCTTTAAGGAGGAACTGGATAAGT<br>ATTTTAAGAACCACACAAGCCCAGATGTGGATCTCGGGGA<br>CATCTCCGGAATTAACGCCTCC**TTC**GTGAATATCCAGAAGG<br>AGATTGACCGCCTAAATGAAGTTGCCAAGAACCTCAATGA<br>GTCTCTGATTGATCTGCAGGAACTGGGCAAGTATGAGCAGT<br>ATATCAAATGGCCCTGGTACATTTGGCTGGGGTTTATCGCC<br>GGACTGATTGCCATCGTCATGGTGACCATCATGCTGTGTTG<br>CATGACCTCCTGTTGTTCCTGTCTGAAGGGCTGCTGTAGTT<br>GCGGCTCTTGCTGTAAATTCGACGAAGATGATAGCGAGCCC<br>GTGCTGAAGGGCGTGAAGCTGCATTATACCTGA |

[0321]

| | |
|---|---|
| 퓨린 절단 부위를 제거하고, 프롤린으로 잔기 986 및 987을 대체하도록 돌연변이되고, L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I 및 V1176F 돌연변이를 함유하는 SARS-CoV-2 S 단백질 | (서열번호 171)<br>MFVFLVLLPLVSSQCVNFTNRTQLPSAYTNSFTRGVYYPDKV<br>FRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLP<br>FNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIK<br>VCEFQFCNYPFLGVYYHKNNKSWMESEFRVYSSANNCTFEY<br>VSQPFLMDLEGKQGNFKNLSEFVFKNIDGYFKIYSKHTPINLV<br>RDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSG<br>WTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET<br>KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN<br>ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT<br>KLNDLCFTNVYADSFVIRGDEVRQIAPGQTGTIADYNYKLPD<br>DFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS<br>TEIYQAGSTPCNGVKGFNCYFPLQSYGFQPTYGVGYQPYRVV<br>VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLT<br>ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSV<br>ITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTG<br>SNVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTNSPGS<br>ASSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPV<br>SMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV<br>EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS<br>FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLT<br>VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ<br>MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASA<br>LGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPP<br>EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAAIKMSE<br>CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQE<br>KNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQII<br>TTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKN<br>HTSPDVDLGDISGINASFVNIQKEIDRLNEVAKNLNESLIDLQE<br>LGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLK<br>GCCSCGSCCKFDEDDSEPVLKGVKLHYT |

[0322]

[0323] *펩티드 융합*

[0324] 본 발명자는 고도로 항원성일 것으로 예상되는 SARS-CoV-2 S 단백질 중의 영역을 식별하였다. 이들은 잔기 815-833(FP), 820-846(D1) 1078-1111(D2) 및 잔기 815-846(F1/D1)을 포함한다. 서열번호 1의 아미노산 서열을 갖는 전장 SARS-CoV-2 단백질 중의 이들 항원성 단편에 대한 서열은 각각, SFIEDLLFNKVTLADAGF(서열번호 21), LLFNKVTLADAGFIKQYGDCLGDIAA(서열번호 22), PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYE(서열번호 23), 및 GGFNFSQILPDPSKPSKRSFIEDLLFNKVTLA(서열번호 24)이다. 항원성 영역은 고도로 항원성일 가능성이 있고 따라서 강한 면역원성 반응을 유도할 것으로 예상되는 다양한 융합 펩티드를 형성하기 위해 상이한 순서로 배열될 수 있다. 도메인은 링커 서열, 예를 들어, GGGGS에 의해 연결될 수 있다. 대안적으로, 이들의 아미노산 서열의 유사성을 고려하여, FP 및 D1 영역은 중첩되어 단일 면역원성 모티프를 생성할 수 있다: SFIEDLLFNKVTLADAGFIKQYGDCLGDIAA(FP/D1)(서열번호 99)(중첩된 서열에는 밑줄이 쳐짐).

[0325] 예시적인 펩티드 융합체는 다음의 도메인을 가질 수 있다:

[0326] **D1- 링커- FP - 링커 - D2 - 링커 - D1(융합 펩티드 A)**

[0327] **FP/D1- 링커- FP/D1- 링커- FP/D1(융합 펩티드 B)**

[0328] 따라서, 본 발명은 SARS-CoV-2 S 단백질의 항원성 영역을 포함하는 융합 펩티드를 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 일 구현예에서, 최적화된 뉴클레오티드 서열은 융합 펩티드 A를 포함하는 아미노산 서열을 암호화한다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 25의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 26의 서열을 갖는다. 또 다른 구현예에서, 최적화된 뉴클레오티드 서열은 융합 펩티드 B를 포함하는 아미노산 서열을 암호화한다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 27의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 28의 서열을 갖는다.

[0329] 특정 구현예에서, 융합 펩티드는 서열번호 7과 같은 N-말단 신호 서열에 작동 가능하게 연결될 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 N-말단 신호 서열과 작동 가능하게 연결된 융합 펩티드 A를 포함하는 아미노산 서열을 암호화할 수 있다. 최적화된 뉴클레오티드 서열은 서열번호 51의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 52의 서열을 갖는다. 대안적으로, 최적화된 뉴클

레오티드 서열은 N-말단 신호 서열과 작동 가능하게 연결된 융합 펩티드 B를 포함하는 아미노산 서열을 암호화할 수 있다. 최적화된 뉴클레오티드 서열은 서열번호 53의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 54의 서열을 갖는다.

[0330] 또한, 융합 펩티드는 일반적으로 N-말단 신호 서열에 더하여, C-말단 Fc 도메인과 작동 가능하게 연결될 수 있다. 예를 들어, 최적화된 뉴클레오티드 서열은 C 말단 Fc 도메인(예를 들어, 서열번호 18) 및 N-말단 신호 서열(예를 들어, 서열번호 7)과 작동 가능하게 연결된 융합 펩티드 A를 포함하는 아미노산 서열을 암호화할 수 있다. 최적화된 뉴클레오티드 서열은 서열번호 55의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 56의 서열을 갖는다. 대안적으로, 최적화된 뉴클레오티드 서열은 C 말단 Fc 도메인(예를 들어, 서열번호 18) 및 N-말단 신호 서열(예를 들어, 서열번호 7)과 작동 가능하게 연결된 융합 펩티드 B를 포함하는 아미노산 서열을 암호화할 수 있다. 최적화된 뉴클레오티드 서열은 서열번호 57의 아미노산 서열을 암호화할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 58의 서열을 갖는다.

[0331] 일부 구현예에서, 융합 펩티드는 생성된 융합 단백질의 순환 반감기를 개선하기 위해 변경된 C-말단 Fc 도메인과 작동 가능하게 연결될 수 있다. 특정 구현예에서, 순환 반감기가 개선된 Fc 도메인은 서열번호 100, 서열번호 101, 서열번호 102 또는 서열번호 103의 아미노산 서열을 갖는다. 따라서, 본 발명은 또한, N-말단 신호 펩티드 및 서열번호 100. 서열번호 101, 서열번호 102 또는 서열번호 103의 아미노산 서열을 갖는 C-말단 Fc 도메인과 작동 가능하게 연결된 융합 펩티드 A 또는 융합 펩티드 B를 암호화하는 최적화된 뉴클레오티드 서열을 제공한다. 신호 펩티드는 서열번호 7의 아미노산 서열을 가질 수 있다.

[0332] *융합 펩티드를 암호화하는 예시적인 최적화된 뉴클레오티드 서열*

[0333] 본 발명에 따른 최적화된 뉴클레오티드 서열은 융합 펩티드의 형태인 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 암호화할 수 있다. 일 구현예에서, 본 발명은 융합 폴리펩티드의 형태인 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 융합 폴리펩티드의 형태인 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서열은 인간 세포에서의 효율적인 발현에 대해 최적화된 융합 펩티드의 형태의 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 암호화하는 뉴클레오티드 서열을 포함한다. 본 발명에 따른 최적화된 뉴클레오티드 서열 및 이에 상응하는 아미노산 서열을 생성하는 프로세스로 생성된 융합 펩티드의 형태의 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 암호화하는 예시적인 최적화된 뉴클레오티드 서열이 표 2에 제시된다. 굵은체로 표시되는 잔기는 자연 발생 SARS-CoV-2 S 단백질과 비교하여 돌연변이된 아미노산을 나타내고, 밑줄이 쳐진 잔기는 신호 펩티드를 나타내며, 이탤릭체로 표시되는 잔기는 Fc 영역의 존재를 나타낸다.

[표 2] 예시적인 융합 펩티드

| | |
|---|---|
| 융합 펩티드 A 를 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 25)<br>ATGCTGCTGTTTAACAAAGTGACTCTGGCAGACGCAG<br>GCTTTATCAAGCAGTACGGAGACTGTCTCGGGGACAT<br>TGCAGCCGGCGGCGGAGGCTCATCTTTCATTGAGGAC<br>CTGCTGTTCAACAAGGTCACTCTGGCAGATGCCGGAT<br>TCGGAGGAGGGGGATCTCCAGCTATCTGCCATGACGG<br>AAAGGCTCATTTTCCTCGGGAGGGTGTGTTTGTGTCCA<br>ACGGAACCCATTGGTTCGTCACACAGCGCAACTTCTA<br>TGAAGGAGGGGGGGGCTCCAGCTTCATCGAGGACCTG<br>CTCTTTAACAAAGTGACCCTGGCCGATGCTGGATTTG<br>GGGGAGGGGGATCCCTGCTGTTCAACAAAGTTACACT<br>GGCCGACGCAGGCTTCATCAAACAGTACGGCGATTGT<br>TTAGGGGACATCGCCGCTGGCGGCGGAGGATCACCTA<br>AGTCCTGCGACAAAACCCATACATGTCCACCATGCCC<br>AGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTCCTCT<br>TCCCTCCTAAGCCCAAGGATACCCTCATGATCTCTCGC<br>ACACCAGAAGTGACCTGCGTGGTCGTGGATGTCTCTC<br>ACGAGGATCCTGAAGTGAAGTTTAACTGGTATGTCGA<br>CGGAGTGGAAGTGCACAACGCCAAGACAAAGCCAAG<br>AGAAGAACAATACAATTCTACTTATAGGGTGGTGTCT<br>GTGCTGACAGTGCTGCACCAGGATTGGCTGAATGGAA<br>AAGAATATAAGTGTAAGGTCTCTAACAAGGCCCTGCC<br>CGCTCCAATTGAGAAGACAATTTCCAAGGCCAAGGGG<br>CAGCCTCGGGAACCTCAGGTGTACACACTGCCCCCAT<br>CCAGGGATGAACTGACTAAAAATCAGGTGTCTCTGAC<br>ATGCCTGGTGAAAGGGTTTTATCCAAGTGACATTGCT<br>GTGGAGTGGGAGTCTAATGGGCAGCCTGAAAATAACT<br>ACAAGACCACACCACCAGTGCTCGATAGCGACGGGTC<br>TTTCTTTCTGTATTCTAAACTGACCGTGGATAAATCTC<br>GGTGGCAGCAGGGAAACGTGTTTTCTTGCTCAGTGAT<br>GCACGAAGCTCTGCACAATCACTATACACAGAAATCC<br>CTGTCCCTGTCTCCAGGCAAATAA |
| 융합 펩티드 A | (서열번호 26)<br>MLLFNKVTLADAGFIKQYGDCLGDIAAGGGGSSFIEDLL<br>FNKVTLADAGFGGGGSPAICHDGKAHFPREGVFVSNGT<br>HWFVTQRNFYEGGGGSSFIEDLLFNKVTLADAGFGGGG<br>SLLFNKVTLADAGFIKQYGDCLGDIAA |
| 융합 펩티드 B 를 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 27)<br>ATGTCCTTCATTGAGGACCTGCTGTTTAATAAGGTGAC<br>CCTGGCCGACGCTGGGTTCATCAAACAGTATGGAGAT<br>TGTCTGGGAGATATTGCAGCAGGCGGGGGCGGCAGC<br>AGCTTTATTGAGGACCTCCTGTTCAACAAGGTGACCC<br>TTGCCGACGCAGGGTTTATTAAGCAGTATGGCGACTG<br>TCTGGGAGACATTGCAGCCGGCGGCGGCGGGTCTTCT |

[0334]

| | |
|---|---|
| | TTTATCGAGGACCTGCTGTTCAACAAGGTGACACTGG<br>CCGACGCAGGCTTTATTAAGCAGTACGGGGACTGCCT<br>GGGAGACATTGCCGCCTGA |
| 융합 펩티드 B | **(서열번호 28)**<br>MSFIEDLLFNKVTLADAGFIKQYGDCLGDIAAGGGGSSFI<br>EDLLFNKVTLADAGFIKQYGDCLGDIAAGGGGSSFIEDL<br>LFNKVTLADAGFIKQYGDCLGDIAA |
| 신호 펩티드를 갖는 융합 펩티드 A 를 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 52)**<br>ATGTTCGTGTTCCTGGTGCTGCTGCCACTGGTTTCCTC<br>CCAGTGTCTGCTGTTTAACAAGGTTACACTGGCAGAC<br>GCCGGCTTCATCAAGCAGTATGGGGACTGTCTGGGCG<br>ATATCGCCGCTGGCGGCGGAGGATCTAGCTTCATTGA<br>GGACCTGCTGTTCAACAAAGTGACTCTGGCTGACGCC<br>GGATTTGGCGGAGGAGGGTCTCCTGCCATTTGTCATG<br>ACGGGAAGGCTCATTTCCCTAGGGAGGGGGTTTTTGT<br>CTCCAATGGAACTCACTGGTTCGTGACCCAAAGAAAC<br>TTCTATGAGGGAGGTGGCGGATCCTCTTTTATCGAGG<br>ACCTGCTGTTTAACAAGGTCACTCTGGCCGATGCAGG<br>CTTCGGAGGAGGAGGGTCTCTGCTGTTCAACAAAGTT<br>ACTCTGGCAGATGCTGGGTTCATTAAGCAGTACGGCG<br>ACTGTCTGGGCGATATTGCCGCCTGA |
| 신호 펩티드를 갖는 융합 펩티드 A | **(서열번호 51)**<br>MFVFLVLLPLVSSQCLLFNKVTLADAGFIKQYGDCLGDI<br>AAGGGGSSFIEDLLFNKVTLADAGFGGGGSPAICHDGKA<br>HFPREGVFVSNGTHWFVTQRNFYEGGGGSSFIEDLLFNK<br>VTLADAGFGGGGSLLFNKVTLADAGFIKQYGDCLGDIA<br>A |
| 신호 펩티드를 갖는 융합 펩티드 B 를 암호화하는 최적화된 뉴클레오티드 서열 | **(서열번호 54)**<br>ATGTTCGTGTTCCTGGTCCTGCTACCCCTGGTGTCCTC<br>TCAGTGCTCCTTCATTGAGGACCTGCTGTTTAATAAGG<br>TGACCCTGGCCGACGCTGGGTTCATCAAACAGTATGG<br>AGATTGTCTGGGAGATATTGCAGCAGGCGGGGGCGGC<br>AGCAGCTTTATTGAGGACCTCCTGTTCAACAAGGTGA<br>CCCTTGCCGACGCAGGGTTTATTAAGCAGTATGGCGA<br>CTGTCTGGGAGACATTGCAGCCGGCGGCGGCGGGTCT<br>TCTTTTATCGAGGACCTGCTGTTCAACAAGGTGACACT<br>GGCCGACGCAGGCTTTATTAAGCAGTACGGGGACTGC<br>CTGGGAGACATTGCCGCCTGA |
| 신호 펩티드를 갖는 융합 펩티드 B | **(서열번호 53)**<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGD<br>CLGDIAAGGGGSSFIEDLLFNKVTLADAGFIKQYGDCLG<br>DIAAGGGGSSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>A |

[0335]

| | |
|---|---|
| 신호 펩티드 및 Fc 영역을 갖는 융합 펩티드 A 를 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 56)<br>ATGTTTGTGTTCCTCGTTCTGCTGCCTCTGGTGAGCTC<br>CCAGTGTCTGCTGTTTAACAAAGTGACTCTGGCAGAC<br>GCAGGCTTTATCAAGCAGTACGGAGACTGTCTCGGGG<br>ACATTGCAGCCGGCGGCGGAGGCTCATCTTTCATTGA<br>GGACCTGCTGTTCAACAAGGTCACTCTGGCAGATGCC<br>GGATTCGGAGGAGGGGGATCTCCAGCTATCTGCCATG<br>ACGGAAAGGCTCATTTTCCTCGGGAGGGTGTGTTTGT<br>GTCCAACGGAACCCATTGGTTCGTCACACAGCGCAAC<br>TTCTATGAAGGAGGGGGGGGCTCCAGCTTCATCGAGG<br>ACCTGCTCTTTAACAAAGTGACCCTGGCCGATGCTGG<br>ATTTGGGGGAGGGGGATCCCTGCTGTTCAACAAAGTT<br>ACACTGGCCGACGCAGGCTTCATCAAACAGTACGGCG<br>ATTGTTTAGGGGACATCGCCGCTGGCGGCGGAGGATC<br>ACCTAAGTCCTGCGACAAAACCCATACATGTCCACCA<br>TGCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTT<br>CCTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCT<br>CTCGCACACCAGAAGTGACCTGCGTGGTCGTGGATGT<br>CTCTCACGAGGATCCTGAAGTGAAGTTTAACTGGTAT<br>GTCGACGGAGTGGAAGTGCACAACGCCAAGACAAAG<br>CCAAGAGAAGAACAATACAATTCTACTTATAGGGTGG<br>TGTCTGTGCTGACAGTGCTGCACCAGGATTGGCTGAA<br>TGGAAAAGAATATAAGTGTAAGGTCTCTAACAAGGCC<br>CTGCCCGCTCCAATTGAGAAGACAATTTCCAAGGCCA<br>AGGGGCAGCCTCGGGAACCTCAGGTGTACACACTGCC<br>CCCATCCAGGGATGAACTGACTAAAAATCAGGTGTCT<br>CTGACATGCCTGGTGAAAGGGTTTTATCCAAGTGACA<br>TTGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAA<br>TAACTACAAGACCACACCACCAGTGCTCGATAGCGAC<br>GGGTCTTTCTTTCTGTATTCTAAACTGACCGTGGATAA<br>ATCTCGGTGGCAGCAGGGAAACGTGTTTTCTTGCTCA<br>GTGATGCACGAAGCTCTGCACAATCACTATACACAGA<br>AATCCCTGTCCCTGTCTCCAGGCAAATAA |
| 신호 펩티드 및 Fc 영역을 갖는 융합 펩티드 A | (서열번호 55)<br>MFVFLVLLPLVSSQCLLFNKVTLADAGFIKQYGDCLGDI<br>AAGGGGSSFIEDLLFNKVTLADAGFGGGGSPAICHDGKA<br>HFPREGVFVSNGTHWFVTQRNFYEGGGGSSFIEDLLFNK<br>VTLADAGFGGGGSLLFNKVTLADAGFIKQYGDCLGDIA<br>AGGGGS*PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDT*<br>*LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTK*<br>*PREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAP*<br>*IEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFY*<br>*PSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKS*<br>*RWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK* |

[0336]

| | |
|---|---|
| 신호 펩티드 및 Fc 영역을 갖는 융합 펩티드 B 를 암호화하는 최적화된 뉴클레오티드 서열 | (서열번호 58)<br>ATGTTCGTGTTCCTGGTCCTGCTGCCTCTGGTGTCCTC<br>TCAGTGCAGCTTCATCGAGGACCTGCTCTTTAACAAG<br>GTGACTCTCGCAGATGCTGGCTTCATCAAGCAGTACG<br>GAGACTGCCTTGGAGACATCGCTGCAGGCGGAGGGG<br>GCAGCAGTTTCATCGAGGACCTGCTGTTTAACAAGGT<br>GACCCTGGCCGACGCCGGGTTCATTAAGCAATACGGC<br>GATTGTCTGGGAGACATCGCAGCTGGGGGAGGGGGG<br>AGCTCTTTTATTGAGGACCTGCTGTTCAACAAGGTGA<br>CTCTGGCCGACGCAGGGTTCATCAAACAGTATGGGGA<br>CTGTCTGGGAGATATCGCAGCCGGGGGAGGAGGCTCC<br>CCTAAGTCCTGCGACAAAACCCATACATGTCCACCAT<br>GCCCAGCTCCTGAACTGCTCGGCGGGCCTAGTGTTTTC<br>CTCTTCCCTCCTAAGCCCAAGGATACCCTCATGATCTC<br>TCGCACACCAGAAGTGACCTGCGTGGTCGTGGATGTC<br>TCTCACGAGGATCCTGAAGTGAAGTTTAACTGGTATG<br>TCGACGGAGTGGAAGTGCACAACGCCAAGACAAAGC<br>CAAGAGAAGAACAATACAATTCTACTTATAGGGTGGT<br>GTCTGTGCTGACAGTGCTGCACCAGGATTGGCTGAAT<br>GGAAAAGAATATAAGTGTAAGGTCTCTAACAAGGCCC<br>TGCCCGCTCCAATTGAGAAGACAATTTCCAAGGCCAA<br>GGGGCAGCCTCGGGAACCTCAGGTGTACACACTGCCC<br>CCATCCAGGGATGAACTGACTAAAAATCAGGTGTCTC<br>TGACATGCCTGGTGAAAGGGTTTTATCCAAGTGACAT<br>TGCTGTGGAGTGGGAGTCTAATGGGCAGCCTGAAAAT<br>AACTACAAGACCACACCACCAGTGCTCGATAGCGACG<br>GGTCTTTCTTTCTGTATTCTAAACTGACCGTGGATAAA<br>TCTCGGTGGCAGCAGGGAAACGTGTTTTCTTGCTCAG<br>TGATGCACGAAGCTCTGCACAATCACTATACACAGAA<br>ATCCCTGTCCCTGTCTCCAGGCAAATAA |
| 신호 펩티드 및 Fc 영역을 갖는 융합 펩티드 B | (서열번호 57)<br>MFVFLVLLPLVSSQCSFIEDLLFNKVTLADAGFIKQYGD<br>CLGDIAAGGGGSSFIEDLLFNKVTLADAGFIKQYGDCLG<br>DIAAGGGGSSFIEDLLFNKVTLADAGFIKQYGDCLGDIA<br>AGGGGS*PKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDT*<br>*LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTK*<br>*PREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAP*<br>*IEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFY*<br>*PSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKS*<br>*RWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK* |

[0337]

[0338] *다른 필수 구조 단백질*

[0339] 관련 β-코로나바이러스에서의 단백질과의 상동성에 기초하여, SARS-CoV-2의 M, N 및 E 단백질은 바이러스 입자의 구조를 형성하는 데 중요한 역할을 하는 것으로 간주된다. M 단백질은 비리온에서 가장 풍부한 구조적 단백질인 것으로 여겨진다. 이는 3개의 막관통 도메인을 갖는 222개 아미노산의 길이이다. M 단백질은 바이러스 입자의 형상을 제공하는 것으로 제안되었다. M 단백질은 비리온에서 이량체로서 존재하는 것으로 제안되며, 여기에서 M 단백질은 막 뒤틀림(membrane curvature)을 촉진하고 핵 캡시드에 결합할 수 있게 하는 2개의 상이한 형태를 채택할 수 있다.

[0340] 419개의 아미노산 길이의 N 단백질은 핵 캡시드를 형성할 가능성이 있다. 이는 상이한 메커니즘을 사용하여 시험관 내에서 RNA에 결합할 수 있는 2개의 별도의 도메인으로 구성된다. N 단백질은 비드-온-어-스트링(beads-on-a-string) 유형 형태의 바이러스 게놈에 결합하고, 또한 바이러스 복제화 복합체의 주요 성분인 nsp3, 및 M 단백질에 결합할 수 있다.

[0341] E 단백질은 77개의 아미노산 길이이고, 바이러스 입자 내에는 소량만 존재하는 것으로 여겨진다. E 단백질의 제안된 기능 중 하나는 바이러스의 조립 및 방출을 용이하게 하는 것이다. SARS-CoV-2의 M, N 및 E 단백질에 대한 아미노산 서열이 아래의 표 3에 제시된다.

[0342] 메모리 CD8+ T 세포는 ORF1ab, S, N, M, 및 ORF3a를 포함하는 많은 SARS-CoV-2 단백질에 대해 광범위한 반응성을 갖고, 대부분의 에피토프는 ORF1ab에 위치하고, 가장 높은 밀도의 에피토프는 N 단백질에 위치한다(Ferretti 등 (2020) https://doi.org/10.1101/2020.07.24.20161653). ORF1ab는 NC_045512.2 SARS-CoV-2 게놈의 잔기 266...13555에 의해 암호화된다. 따라서, SARS-CoV-2의 ORF1ab 및 N 단백질은 T 세포 반응을 유도하는 데 유용할 수 있다.

[표 3] SARS-CoV-2 M, E 및 N 단백질

| | |
|---|---|
| SARS-CoV-2 M 단백질의 뉴클레오티드 서열<br><br>NC_004718.3 SARS-CoV-2 게놈<br><br>범위 26398..27063 | (서열번호 59)<br>ATGGCAGACAACGGTACTATTACCGTTGAGGAGCTTA<br>AACAACTCCTGGAACAATGGAACCTAGTAATAGGTTT<br>CCTATTCCTAGCCTGGATTATGTTACTACAATTTGCCT<br>ATTCTAATCGGAACAGGTTTTTGTACATAATAAAGCTT<br>GTTTTCCTCTGGCTCTTGTGGCCAGTAACACTTGCTTG<br>TTTTGTGCTTGCTGCTGTCTACAGAATTAATTGGGTGA<br>CTGGCGGGATTGCGATTGCAATGGCTTGTATTGTAGG<br>CTTGATGTGGCTTAGCTACTTCGTTGCTTCCTTCAGGC<br>TGTTTGCTCGTACCCGCTCAATGTGGTCATTCAACCCA<br>GAAACAAACATTCTTCTCAATGTGCCTCTCCGGGGGA<br>CAATTGTGACCAGACCGCTCATGGAAAGTGAACTTGT<br>CATTGGTGCTGTGATCATTCGTGGTCACTTGCGAATGG<br>CCGGACACTCCCTAGGGCGCTGTGACATTAAGGACCT<br>GCCAAAAGAGATCACTGTGGCTACATCACGAACGCTT<br>TCTTATTACAAATTAGGAGCGTCGCAGCGTGTAGGCA<br>CTGATTCAGGTTTTGCTGCATACAACCGCTACCGTATT<br>GGAAACTATAAATTAAATACAGACCACGCCGGTAGCA<br>ACGACAATATTGCTTTGCTAGTACAGTAA |
| SARS-CoV-2 M 단백질 서열<br><br>수탁번호 QII57163 | (서열번호 60)<br>MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAY<br>ANRNRFLYIIKLIFLWLLWPVTLACFVLAAVYRINWITG<br>GIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETN<br>ILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGR<br>CDIKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYS<br>RYRIGNYKLNTDHSSSSDNIALLVQ |
| SARS-CoV-2 E 단백질의 뉴클레오티드 서열<br><br>NC_004718.3 SARS-CoV-2 게놈<br><br>범위 26117..26347 | (서열번호 61)<br>ATGTACTCATTCGTTTCGGAAGAAACAGGTACGTTAA<br>TAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTAT<br>TCTTGCTAGTCACACTAGCCATCCTTACTGCGCTTCGA<br>TTGTGTGCGTACTGCTGCAATATTGTTAACGTGAGTTT<br>AGTAAAACCAACGGTTTACGTCTACTCGCGTGTTAAA<br>AATCTGAACTCTTCTGAAGGAGTTCCTGATCTTCTGGT<br>CTAA |
| SARS-CoV-2 E 단백질 서열<br><br>수탁번호 P59637.1 | (서열번호 62)<br>MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLC<br>AYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV |

[0343]

| | |
|---|---|
| **SARS-CoV-2 N 단백질의 뉴클레오티드 서열**<br>NC_045512.2 SARS-CoV-2 게놈<br><br>범위 28274..29533 | **(서열번호 63)**<br>ATGTCTGATAATGGACCCCAAAATCAGCGAAATGCAC<br>CCCGCATTACGTTTGGTGGACCCTCAGATTCAACTGG<br>CAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATC<br>AAAACAACGTCGGCCCCAAGGTTTACCCAATAATACT<br>GCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGG<br>AAGACCTTAAATTCCCTCGAGGACAAGGCGTTCCAAT<br>TAACACCAATAGCAGTCCAGATGACCAAATTGGCTAC<br>TACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACG<br>GTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTA<br>CTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCTAT<br>GGTGCTAACAAAGACGGCATCATATGGGTTGCAACTG<br>AGGGAGCCTTGAATACACCAAAAGATCACATTGGCAC<br>CCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAA<br>CTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACG<br>CAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCG<br>TTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCA<br>ACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAA<br>TGGCTGGCAATGGCGGTGATGCTGCTCTTGCTTTGCTG<br>CTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGT<br>CTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCA<br>CTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCG<br>GCAAAAACGTACTGCCACTAAAGCATACAATGTAACA<br>CAAGCTTTCGGCAGACGTGGTCCAGAACAAACCCAAG<br>GAAATTTTGGGGACCAGGAACTAATCAGACAAGGAA<br>CTGATTACAAACATTGGCCGCAAATTGCACAATTTGC<br>CCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTG<br>GCATGGAAGTCACACCTTCGGGAACGTGGTTGACCTA<br>CACAGGTGCCATCAAATTGGATGACAAAGATCCAAAT<br>TTCAAAGATCAAGTCATTTTGCTGAATAAGCATATTG<br>ACGCATACAAAACATTCCCACCAACAGAGCCTAAAAA<br>GGACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTT<br>ACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCT<br>TCTTCCTGCTGCAGATTTGGATGATTTCTCCAAACAAT<br>TGCAACAATCCATGAGCAGTGCTGACTCAACTCAGGC<br>CTAA |
| **SARS-CoV-2 N 단백질 서열**<br><br>수탁번호 QIS29990.1 | **(서열번호 64)**<br>MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSK<br>QRRPQGLPNNTASWFTALTQHGKEDLKFPRGQGVPINT<br>NSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLG<br>TGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPAN<br>NAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNS<br>SRNLTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLE<br>SKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAY<br>NVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQ<br>FAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPN<br>FKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQ<br>RQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA |

[0344]

[0345] 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 일 구현예에서, 본 발명은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서열은 효율적인 발현 인간 세포에 대해 최적화된 SARS-CoV-2 작은 외피 단백질 또는 이의 항원성 단편을 암호화하는 뉴클레오티드 서열을 포함한다.

[0346] 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 일 구현예에서, 본 발명은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서열은 효율적인 발현 인간 세포에 대해 최적화된 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 뉴클레오티드 서열을 포함한다. 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 일 구현예에서, 본 발명은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서

열은 효율적인 발현 인간 세포에 대해 최적화된 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 뉴클레오티드 서열을 포함한다.

[0347] 본 발명에 따른 최적화된 뉴클레오티드 서열은 SARS-CoV-2 ORF1ab 단백질 또는 이의 항원성 단편을 암호화할 수 있다. 일 구현예에서, 본 발명은 SARS-CoV-2 ORF1ab 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공한다. 일부 구현예에서, 핵산은 SARS-CoV-2 ORF1ab 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA이다. 일부 구현예에서, 적절한 mRNA 서열은 효율적인 발현 인간 세포에 대해 최적화된 SARS-CoV-2 ORF1ab 단백질 또는 이의 항원성 단편을 암호화하는 뉴클레오티드 서열을 포함한다.

[0348] 일부 구현예에서, SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산과 조합된다. 일부 구현예에서, SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산과 조합된다. 일부 구현예에서, SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산과 조합된다. 다른 구현예에서, SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산은 제2, 제3 및/또는 제4 핵산과 조합되며, 여기에서 전술한 제2 핵산은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 전술한 제3 핵산은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 전술한 제4 핵산은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0349] **mRNA 서열**

[0350] 일부 구현예에서, SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA는 5' 및 3' UTR 서열을 또한 함유한다. 예시적인 5' 및 3' UTR 서열은 아래와 같다:

[0351] 예시적인 5' UTR 서열

[0352] GGACAGAUCGCCUGGAGACGCCAUCCACGCUGUUUUGACCUCCAUAGAAGACACCGGGACCGAUCCAGCCUCCGCGGCCGGGAACGGUGCAUUGGAACGCGGAUUCCCCGUGCCAAGAGUGACUCACCGUCCUUGACACG (**서열번호 144**)

[0353] 예시적인 3' UTR 서열

[0354] CGGGUGGCAUCCCUGUGACCCCUCCCCAGUGCCUCUCCUGGCCCUGGAAGUUGCCACUCCAGUGCCCACCAGCCUUGUCCUAAUAAAAUUAAGUUGCAUCAAGCU (**서열번호 145**)

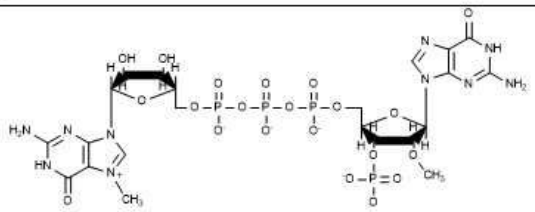
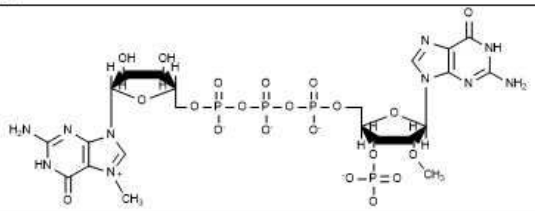
[0355] 또는

[0356] GGGUGGCAUCCCUGUGACCCCUCCCCAGUGCCUCUCCUGGCCCUGGAAGUUGCCACUCCAGUGCCCACCAGCCUUGUCCUAAUAAAAUUAAGUUGCAUCAAAGCU (**서열번호 146**)

[0357] 예시적인 mRNA 작제물

[0358] 특정 구현예에서, SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA는 다음의 구조 요소를 포함한다.

[표 4] 예시적인 mRNA 작제물의 구조 요소

| 구조 요소 | 설명 | 서열 좌표 |
|---|---|---|
| *mRNA 작제물 1* | | |
| 캡 구조 |  | 1 |
| 5′ UTR | *GGAC…CACG* | 1-140 **(서열번호 144)** |
| SARS-CoV-2 S 단백질 [1] | AUG…UGA | 141-3962 **(서열번호 148)** (이는 서열번호 44 의 뉴클레오티드 서열에 상응함) |
| 3′ UTR | *CGGG…AGCU* | 3963-4067 **(서열번호 145)** |
| 폴리 A 꼬리 | $(A)_x$, x=100-500[3] | NA |
| *mRNA 작제물 2* | | |
| 캡 구조 |  | 1 |
| 5′ UTR | *GGAC…CACG* | 1-140 **(서열번호 144)** |
| SARS-CoV-2 S 단백질 [2] | AUG…UGA | 141-3962 **(서열번호 173)** (이는 서열번호 166 의 뉴클레오티드 서열에 상응함) |
| 3′ UTR | *CGGG…AGCU* | 3963-4067 **(서열번호 145)** |
| 폴리 A 꼬리 | $(A)_x$, x=100-500[3] | NA |

[0359]

NA = 해당 없음
UTR = 비번역 영역
[1] 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열
[2] 퓨린 절단 부위를 제거하고 프롤린으로 잔기 986 및 987 을 대체하도록 돌연변이되고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 함유하는 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열
[3] 예상 범위

[0360]

[0361] 특정 일 구현예에서, 본 발명에 따른 mRNA는 다음의 핵산 서열을 갖는다:

```
   1 GGACAGAUCG CCUGGAGACG CCAUCCACGC UGUUUUGACC UCCAUAGAAG
  51 ACACCGGGAC CGAUCCAGCC UCCGCGGCCG GGAACGGUGC AUUGGAACGC
 101 GGAUUCCCCG UGCCAAGAGU GACUCACCGU CCUUGACACG AUGUUCGUCU
 151 UCCUCGUGCU GCUCCCACUC GUUUCUUCCC AGUGUGUCAA CCUGACAACU
 201 AGGACUCAGC UGCCACCAGC CUACACCAAC UCCUUCACCA GAGGCGUGUA
 251 UUACCCAGAC AAGGUGUUUA GAAGCAGCGU GCUGCACUCU ACCCAGGACC
 301 UCUUUCUGCC CUUUUUCAGC AACGUGACAU GGUUUCACGC AAUUCACGUG
 351 UCCGGCACUA AUGGCACAAA GCGGUUCGAC AAUCCAGUCC UGCCUUUCAA
 401 CGAUGGCGUC UACUUUGCAU CUACUGAGAA AUCCAAUAUC AUUAGGGGAU
 451 GGAUCUUCGG CACAACCCUG GAUUCUAAGA CCCAGAGCCU GCUGAUCGUC
 501 AACAACGCCA CAAACGUGGU CAUUAAGGUU UGCGAGUUUC AGUUCUGUAA
 551 CGAUCCUUUU CUGGGCGUGU AUUAUCAUAA GAACAAUAAG AGCUGGAUGG
 601 AGUCCGAGUU UAGAGUGUAU AGCUCUGCAA AUAAUUGUAC CUUUGAGUAC
 651 GUGAGCCAGC CCUUUCUGAU GGACCUGGAG GGAAAACAAG GAAACUUCAA
 701 AAACCUGCGG GAAUUCGUUU UCAAAAACAU CGACGGCUAU UUCAAGAUCU
 751 AUAGCAAGCA UACCCCAAUC AACCUCGUGA GGGACCUCCC CCAGGGCUUU
 801 AGCGCACUGG AGCCACUGGU UGACCUGCCU AUCGGCAUUA AUAUCACAAG
 851 AUUUCAGACC CUGCUGGCAC UGCAUAGAAG CUAUCUGACC CCUGGAGACU
```

[0362]

```
 901 CCUCUAGUGG GUGGACUGCC GGCGCCGCUG CCUACUAUGU GGGCUAUCUG
 951 CAGCCACGGA CAUUCCUGCU GAAAUACAAU GAGAACGGGA CAAUCACAGA
1001 UGCUGUUGAU UGCGCACUCG ACCCCCUGUC CGAGACAAAG UGCACUCUCA
1051 AGAGCUUUAC CGUCGAGAAG GGCAUCUAUC AGACCUCAAA CUUCAGGGUG
1101 CAGCCCACAG AAUCUAUCGU GCGCUUCCCU AAUAUCACUA ACCUGUGUCC
1151 UUUCGGUGAA GUGUUCAACG CCACCAGGUU UGCUAGCGUG UAUGCCUGGA
1201 ACAGGAAGAG GAUCUCUAAC UGCGUCGCCG ACUAUUCCGU GCUGUAUAAC
1251 AGCGCCUCCU UCUCCACAUU CAAAUGCUAU GGAGUGAGCC CGACAAAACU
1301 GAACGAUCUC UGCUUUACAA AUGUCUACGC CGACUCUUUU GUGAUCAGAG
1351 GGGACGAGGU CCGGCAGAUC GCACCAGGAC AGACAGGCAA GAUUGCUGAC
1401 UACAACUAUA AGCUGCCUGA CGACUUCACA GGAUGUGUGA UCGCAUGGAA
1451 CUCAAACAAU CUGGACUCCA AAGUCGGGGG CAACUAUAAU UACCUGUAUC
1501 GCCUGUUCCG GAAGUCCAAC CUGAAGCCCU UCGAGAGGGA CAUCAGUACA
1551 GAGAUCUAUC AGGCUGGCUC CACCCCUUGC AAUGGCGUCG AAGGCUUUAA
1601 UUGUUAUUUU CCCCUGCAGU CUUACGGGUU UCAGCCUACU AAUGGAGUUG
1651 GGUACCAGCC AUACAGAGUG GUCGUGCUCA GCUUCGAGCU CCUGCAUGCU
1701 CCAGCUACAG UUUGCGGGCC AAAGAAGUCC ACUAACCUGG UGAAGAAUAA
1751 GUGCGUCAAC UUCAACUUUA ACGGGCUCAC CGGCACCGGC GUGCUGACUG
1801 AGAGCAACAA GAAGUUUCUG CCAUUUCAAC AGUUUGGACG GGACAUUGCC
1851 GACACCACCG AUGCCGUUCG GGAUCCACAG ACCCUGGAAA UUCUGGACAU
1901 UACACCGUGC AGCUUCGGGG GCGUGAGCGU GAUCACACCC GGAACCAAUA
1951 CAAGCAACCA GGUUGCCGUC CUGUAUCAGG AUGUCAAUUG CACAGAAGUG
2001 CCAGUUGCUA UCCACGCAGA CCAGCUGACU CCCACAUGGC GGGUGUAUAG
2051 CACCGGAUCC AACGUGUUUC AGACCCGCGC CGGAUGUCUC AUUGGGGCCG
2101 AGCACGUGAA UAACAGCUAC GAGUGCGACA UCCCCAUUGG CGCCGGCAUU
2151 UGUGCGUCUU ACCAGACUCA GACCAACUCU CCUGGCUCCG CCUCUUCCGU
```

[0363]

2201 UGCUAGUCAG UCUAUUAUUG CCUAUACCAU GAGCCUCGGA GCUGAGAAUA

2251 GCGUGGCCUA CUCCAAUAAU UCCAUCGCAA UCCCUACUAA CUUCACUAUU

2301 UCUGUGACCA CCGAGAUCCU GCCUGUGUCU AUGACUAAGA CUAGCGUUGA

2351 UUGUACCAUG UAUAUUUGUG GCGACUCUAC CGAAUGUUCU AACCUGCUGC

2401 UUCAGUACGG CUCAUUUUGC ACACAGCUGA ACAGAGCCCU GACUGGGAUC

2451 GCUGUGGAGC AGGACAAGAA CACACAGGAG GUGUUUGCAC AGGUGAAGCA

2501 GAUCUAUAAG ACCCCUCCUA UUAAGGAUUU CGGCGGAUUC AAUUUCUCAC

2551 AGAUUCUGCC AGACCCCAGU AAGCCUUCCA AGAGGAGCUU CAUCGAGGAU

2601 CUCCUGUUUA ACAAGGUGAC CCUGGCAGAC GCCGGCUUUA UUAAGCAAUA

2651 UGGGGAUUGC CUGGGCGACA UUGCUGCCAG AGACCUGAUU UGCGCCCAGA

2701 AAUUCAAUGG CCUCACAGUG CUGCCACCUC UGCUGACCGA CGAGAUGAUC

2751 GCUCAAUACA CUAGCGCACU GCUGGCCGGA ACCAUCACAU CAGGCUGGAC

2801 CUUCGGGGCC GGAGCAGCAC UGCAGAUUCC AUUCGCCAUG CAGAUGGCCU

2851 AUAGAUUCAA CGGCAUUGGC GUCACACAGA ACGUGCUGUA CGAAAACCAG

2901 AAGCUCAUCG CUAACCAGUU UAAUUCCGCA AUUGGAAAGA UCCAAGAUUC

2951 ACUCAGCUCA ACCGCCUCUG CACUCGGAAA GCUGCAGGAC GUGGUCAACC

3001 AGAAUGCUCA GGCCCUGAAC ACACUCGUCA AGCAGCUGUC CUCUAACUUU

3051 GGCGCUAUCA GCUCCGUUCU GAACGACAUU CUGAGCCGCC UGGAUCCCCC

3101 AGAGGCUGAA GUCCAGAUUG ACCGCCUGAU UACCGGCCGG CUGCAGUCUC

3151 UGCAAACAUA CGUGACCCAG CAGCUGAUCA GAGCAGCCGA GAUCCGGGCA

3201 UCCGCAAAUC UGGCAGCAAC UAAGAUGAGC GAAUGCGUGC UGGGCCAGUC

3251 CAAGCGGGUG GACUUUUGUG GCAAGGGCUA CCACCUGAUG AGCUUCCCCC

3301 AGAGCGCCCC ACAUGGCGUU GUUUUUCUGC ACGUGACCUA UGUCCCUGCU

3351 CAGGAAAAGA ACUUUACAAC UGCUCCUGCU AUCUGCCAUG ACGGCAAGGC

3401 CCACUUCCCA CGGGAGGGAG UGUUUGUGUC CAAUGGCACA CACUGGUUCG

[0364] 3451 UGACCCAGAG GAACUUCUAU GAACCCCAGA UCAUCACCAC UGACAAUACC

3501 UUCGUGUCUG GAAAUUGCGA CGUCGUGAUC GGCAUCGUUA ACAACACCGU

3551 GUACGACCCU CUCCAGCCAG AGCUGGACUC CUUUAAGGAG GAACUGGAUA

3601 AGUAUUUUAA GAACCACACA AGCCCAGAUG UGGAUCUCGG GGACAUCUCC

3651 GGAAUUAACG CCUCCGUGGU GAAUAUCCAG AAGGAGAUUG ACCGCCUAAA

3701 UGAAGUUGCC AAGAACCUCA AUGAGUCUCU GAUUGAUCUG CAGGAACUGG

3751 GCAAGUAUGA GCAGUAUAUC AAAUGGCCCU GGUACAUUUG GCUGGGGUUU

3801 AUCGCCGGAC UGAUUGCCAU CGUCAUGGUG ACCAUCAUGC UGUGUUGCAU

3851 GACCUCCUGU UGUUCCUGUC UGAAGGGCUG CUGUAGUUGC GGCUCUUGCU

3901 GUAAAUUCGA CGAAGAUGAU AGCGAGCCCG UGCUGAAGGG CGUGAAGCUG

3951 CAUUAUACC**U GA**CGGGUGGC AUCCCUGUGA CCCCUCCCCA GUGCCUCUCC

4001 UGGCCCUGGA AGUUGCCACU CCAGUGCCCA CCAGCCUUGU CCUAAUAAAA

4051 UUAAGUUGCA UCAAGCU (**서열번호 147**)

[0365]

[0366] + 폴리A 꼬리

[0367] **굵은 글씨체**로 표시된 핵산은 시작 및 정지 코돈을 나타낸다.

[0368] 또 다른 특정 구현예에서, 본 발명에 따른 mRNA는 다음의 핵산 서열을 갖는다:

1 GGACAGAUCG CCUGGAGACG CCAUCCACGC UGUUUUGACC UCCAUAGAAG

51 ACACCGGGAC CGAUCCAGCC UCCGCGGCCG GGAACGGUGC AUUGGAACGC

101 GGAUUCCCCG UGCCAAGAGU GACUCACCGU CCUUGACACG **AUG**UUCGUCU

151 UCCUCGUGCU GCUCCCACUC GUUUCUUCCC AGUGUGUCAA CUUCACAACU

201 AGGACUCAGC UGCCACCAGC CUACACCAAC UCCUUCACCA GAGGCGUGUA

251 UUACCCAGAC AAGGUGUUUA GAAGCAGCGU GCUGCACUCU ACCCAGGACC

301 UCUUUCUGCC CUUUUUCAGC AACGUGACAU GGUUUCACGC AAUUCACGUG

351 UCCGGCACUA AUGGCACAAA GCGGUUCGCC AAUCCAGUCC UGCCUUUCAA

[0369]

401 CGAUGGCGUC UACUUUGCAU CUACUGAGAA AUCCAAUAUC AUUAGGGGAU

451 GGAUCUUCGG CACAACCCUG GAUUCUAAGA CCCAGAGCCU GCUGAUCGUC

501 AACAACGCCA CAAACGUGGU CAUUAAGGUU UGCGAGUUUC AGUUCUGUAA

551 CGAUCCUUUU CUGGGCGUGU AUUAUCAUAA GAACAAUAAG AGCUGGAUGG

601 AGUCCGAGUU UAGAGUGUAU AGCUCUGCAA AUAAUUGUAC CUUUGAGUAC

651 GUGAGCCAGC CCUUUCUGAU GGACCUGGAG GGAAAACAAG GAAACUUCAA

701 AAACCUGCGG GAAUUCGUUU UCAAAAACAU CGACGGCUAU UUCAAGAUCU

751 AUAGCAAGCA UACCCCAAUC AACCUCGUGA GGGGCCUCCC CCAGGGCUUU

801 AGCGCACUGG AGCCACUGGU UGACCUGCCU AUCGGCAUUA AUAUCACAAG

851 AUUUCAGACC CUGCAUAGAA GCUAUCUGAC CCCUGGAGAC UCCUCUAGUG

901 GGUGGACUGC CGGCGCCGCU GCCUACUAUG UGGGCUAUCU GCAGCCACGG

951 ACAUUCCUGC UGAAAUACAA UGAGAACGGG ACAAUCACAG AUGCUGUUGA

1001 UUGCGCACUC GACCCCCUGU CCGAGACAAA GUGCACUCUC AAGAGCUUUA

1051 CCGUCGAGAA GGGCAUCUAU CAGACCUCAA ACUUCAGGGU GCAGCCCACA

1101 GAAUCUAUCG UGCGCUUCCC UAAUAUCACU AACCUGUGUC CUUUCGGUGA

1151 AGUGUUCAAC GCCACCAGGU UUGCUAGCGU GUAUGCCUGG AACAGGAAGA

1201 GGAUCUCUAA CUGCGUCGCC GACUAUUCCG UGCUGUAUAA CAGCGCCUCC

1251 UUCUCCACAU UCAAAUGCUA UGGAGUGAGC CCGACAAAAC UGAACGAUCU

1301 CUGCUUUACA AAUGUCUACG CCGACUCUUU UGUGAUCAGA GGGGACGAGG

1351 UCCGGCAGAU CGCACCAGGA CAGACAGGCA ACAUUGCUGA CUACAACUAU

1401 AAGCUGCCUG ACGACUUCAC AGGAUGUGUG AUCGCAUGGA ACUCAAACAA

1451 UCUGGACUCC AAAGUCGGGG GCAACUAUAA UUACCUGUAU CGCCUGUUCC

1501 GGAAGUCCAA CCUGAAGCCC UUCGAGAGGG ACAUCAGUAC AGAGAUCUAU

1551 CAGGCUGGCU CCACCCCUUG CAAUGGCGUC AAGGGCUUUA AUUGUUAUUU

1601 UCCCCUGCAG UCUUACGGGU UUCAGCCUAC UUACGGAGUU GGGUACCAGC

[0370] 1651 CAUACAGAGU GGUCGUGCUC AGCUUCGAGC UCCUGCAUGC UCCAGCUACA

1701 GUUUGCGGGC CAAAGAAGUC CACUAACCUG GUGAAGAAUA AGUGCGUCAA

1751 CUUCAACUUU AACGGGCUCA CCGGCACCGG CGUGCUGACU GAGAGCAACA

1801 AGAAGUUUCU GCCAUUUCAA CAGUUUGGAC GGGACAUUGC CGACACCACC

1851 GAUGCCGUUC GGGAUCCACA GACCCUGGAA AUUCUGGACA UUACACCGUG

1901 CAGCUUCGGG GGCGUGAGCG UGAUCACACC CGGAACCAAU ACAAGCAACC

1951 AGGUUGCCGU CCUGUAUCAG GGCGUCAAUU GCACAGAAGU GCCAGUUGCU

2001 AUCCACGCAG ACCAGCUGAC UCCCACAUGG CGGGUGUAUA GCACCGGAUC

2051 CAACGUGUUU CAGACCCGCG CCGGAUGUCU CAUUGGGGCC GAGCACGUGA

2101 AUAACAGCUA CGAGUGCGAC AUCCCCAUUG GCGCCGGCAU UUGUGCGUCU

2151 UACCAGACUC AGACCAACUC UCCUGGCUCC GCCUCUUCCG UUGCUAGUCA

2201 GUCUAUUAUU GCCUAUACCA UGAGCCUCGG AGUGGAGAAU AGCGUGGCCU

2251 ACUCCAAUAA UUCCAUCGCA AUCCCUACUA ACUUCACUAU UUCUGUGACC

2301 ACCGAGAUCC UGCCUGUGUC UAUGACUAAG ACUAGCGUUG AUUGUACCAU

2351 GUAUAUUUGU GGCGACUCUA CCGAAUGUUC UAACCUGCUG CUUCAGUACG

2401 GCUCAUUUUG CACACAGCUG AACAGAGCCC UGACUGGGAU CGCUGUGGAG

2451 CAGGACAAGA ACACACAGGA GGUGUUUGCA CAGGUGAAGC AGAUCUAUAA

2501 GACCCCUCCU AUUAAGGAUU UCGGCGGAUU CAAUUUCUCA CAGAUUCUGC

2551 CAGACCCCAG UAAGCCUUCC AAGAGGAGCU UCAUCGAGGA UCUCCUGUUU

2601 AACAAGGUGA CCCUGGCAGA CGCCGGCUUU AUUAAGCAAU AUGGGGAUUG

2651 CCUGGGCGAC AUUGCUGCCA GAGACCUGAU UUGCGCCCAG AAAUUCAAUG

2701 GCCUCACAGU GCUGCCACCU CUGCUGACCG ACGAGAUGAU CGCUCAAUAC

2751 ACUAGCGCAC UGCUGGCCGG AACCAUCACA UCAGGCUGGA CCUUCGGGGC

2801 CGGAGCAGCA CUGCAGAUUC CAUUCGCCAU GCAGAUGGCC UAUAGAUUCA

2851 ACGGCAUUGG CGUCACACAG AACGUGCUGU ACGAAAACCA GAAGCUCAUC

[0371] 2901 GCUAACCAGU UUAAUUCCGC AAUUGGAAAG AUCCAAGAUU CACUCAGCUC

2951 AACCGCCUCU GCACUCGGAA AGCUGCAGGA CGUGGUCAAC CAGAAUGCUC

3001 AGGCCCUGAA CACACUCGUC AAGCAGCUGU CCUCUAACUU UGGCGCUAUC

3051 AGCUCCGUUC UGAACGACAU UCUGAGCCGC CUGGAUCCCC CAGAGGCUGA

3101 AGUCCAGAUU GACCGCCUGA UUACCGGCCG GCUGCAGUCU CUGCAAACAU

3151 ACGUGACCCA GCAGCUGAUC AGAGCAGCCG AGAUCCGGGC AUCCGCAAAU

3201 CUGGCAGCAA CUAAGAUGAG CGAAUGCGUG CUGGGCCAGU CCAAGCGGGU

3251 GGACUUUUGU GGCAAGGGCU ACCACCUGAU GAGCUUCCCC CAGAGCGCCC

3301 CACAUGGCGU UGUUUUUCUG CACGUGACCU AUGUCCCUGC UCAGGAAAAG

3351 AACUUUACAA CUGCUCCUGC UAUCUGCCAU GACGGCAAGG CCCACUUCCC

3401 ACGGGAGGGA GUGUUUGUGU CCAAUGGCAC ACACUGGUUC GUGACCCAGA

3451 GGAACUUCUA UGAACCCCAG AUCAUCACCA CUGACAAUAC CUUCGUGUCU

3501 GGAAAUUGCG ACGUCGUGAU CGGCAUCGUU AACAACACCG UGUACGACCC

3551 UCUCCAGCCA GAGCUGGACU CCUUUAAGGA GGAACUGGAU AAGUAUUUUA

3601 AGAACCACAC AAGCCCAGAU GUGGAUCUCG GGGACAUCUC CGGAAUUAAC

3651 GCCUCCGUGG UGAAUAUCCA GAAGGAGAUU GACCGCCUAA AUGAAGUUGC

3701 CAAGAACCUC AAUGAGUCUC UGAUUGAUCU GCAGGAACUG GGCAAGUAUG

3751 AGCAGUAUAU CAAAUGGCCC UGGUACAUUU GGCUGGGGUU UAUCGCCGGA

3801 CUGAUUGCCA UCGUCAUGGU GACCAUCAUG CUGUGUUGCA UGACCUCCUG

3851 UUGUUCCUGU CUGAAGGGCU GCUGUAGUUG CGGCUCUUGC UGUAAAUUCG

3901 ACGAAGAUGA UAGCGAGCCC GUGCUGAAGG GCGUGAAGCU GCAUUAUACC

3951 **UGA**CGGGUGG CAUCCCUGUG ACCCCUCCCC AGUGCCUCUC CUGGCCCUGG

4001 AAGUUGCCAC UCCAGUGCCC ACCAGCCUUG UCCUAAUAAA AUUAAGUUGC

4051 AUCAAGCU (**서열번호 172**)

[0372]

[0373] + 폴리A 꼬리

[0374] **굵은 글씨체**로 표시된 핵산은 시작 및 정지 코돈을 나타낸다.

[0375] **mRNA 합성**

[0376] *시험관 내 전사*

[0377] 본 발명에 따른 mRNA는 알려진 다양한 방법 중 어느 하나에 따라 합성될 수 있다. 여러 가지 방법이 미국 특허 출원 제US 2018/0258423호(이 내용 전체는 본원에 참조로서 포함됨)에 기술되어 있으며, 이는 본 발명을 실시하는 데 사용될 수 있다. 예를 들어, 본 발명에 따른 mRNA는 시험관 내 전사(IVT)를 통해 합성될 수 있다. 요약하면, IVT는 일반적으로 프로모터를 함유하는 선형 또는 원형 DNA 템플릿, 리보뉴클레오티드 트리포스페이트의 풀, DTT 및 마그네슘 이온을 포함할 수 있는 완충액 시스템, 및 적합한 RNA 중합효소(예를 들어, T3, T7, 또는 SP6 RNA 중합효소), DNAse I, 피로포스파타아제, 및/또는 RNAse 억제제로 수행된다. 정확한 조건은 특정 응용예에 따라 달라질 것이다.

[0378] 일부 구현예에서, 본 발명에 따른 mRNA의 제조를 위해, DNA 템플릿을 시험관 내에서 전사한다. 적절한 DNA 템플릿은 일반적으로, 시험관 내 전사를 위한 프로모터(예를 들어, T3, T7 또는 SP6 프로모터)를 가지며, 원하는

mRNA에 대한 원하는 뉴클레오티드 서열 및 종결 신호가 이어진다.

[0379] *뉴클레오티드*

[0380] 일부 구현예에서, mRNA는 자연 발생 뉴클레오시드(또는 비변형 뉴클레오시드; 즉, 아데노신, 구아노신, 시티딘, 및 우리딘)를 포함하거나 이로 이루어진다. 일부 구현예에서, mRNA는 하나 이상의 뉴클레오시드, 예컨대 뉴클레오시드 유사체(예를 들어, 아데노신 유사체, 구아노신 유사체, 시티딘 유사체, 또는 우리딘 유사체)를 포함한다. 하나 이상의 뉴클레오시드 유사체의 존재는, 해당 mRNA가, 동일한 서열을 가지지만 자연 발생 뉴클레오시드만을 함유하는 대조군 mRNA이 비해 보다 안정적이고/이거나 덜 면역원성이 되게 할 수 있다. 본 발명의 특정 구현예에서, SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA는 자연 발생 뉴클레오시드와 합성된다. 임의의 특정 이론에 구속되고자 함이 없이, 본 발명자는 자연 발생 뉴클레오시드로 제조된 mRNA의 사용이 본 발명의 면역원성 조성물을 제공하는 데 유리하다고 믿는다.

[0381] 일부 구현예에서, mRNA는 비변형 뉴클레오티드 및 변형된 뉴클레오티드 둘 모두를 포함한다. 일부 구현예에서, 하나 이상의 변형된 뉴클레오시드는 뉴클레오시드 유사체이다. 일부 구현예에서, 하나 이상의 변형된 뉴클레오시드는 변형된 당 및 변형된 핵염기로부터 선택되는 적어도 하나의 변형을 포함한다. 일부 구현예에서, mRNA는 하나 이상의 변형된 뉴클레오시드 간 연결을 포함한다.

[0382] 일부 구현예에서, 하나 이상의 변형된 뉴클레오시드는 뉴클레오시드 유사체, 예를 들어 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤-피리미딘, 3-메틸 아데노신, 5-메틸시티딘, C-5 프로피닐-시티딘, C-5 프로피닐-우리딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌, 슈도우리딘(예를 들어, N-1-메틸-슈도우리딘), 2-티오우리딘, 및 2-티오시티딘 중 하나이다. 5-메틸-시티딘, 슈도우리딘, 및 2-티오-우리딘 및 이들의 mRNA로의 혼입에 대한 논의는 예를 들어, 미국 특허 제8,278,036호 또는 WO 2011/012316를 참조한다. 일부 구현예에서, mRNA는 25%의 U 잔기가 2-티오-우리딘이고 25%의 C 잔기는 5-메틸시티딘인 RNA일 수 있다. 이러한 변형된 RNA의 사용에 대한 교시는 미국 특허 공개 US 2012/0195936 및 국제 공개 WO 2011/012316에 개시되어 있으며, 이들 모두는 그 전체가 참조로서 본원에 통합된다.

[0383] *합성-후 프로세싱*

[0384] 일반적으로, 5' 캡 및/또는 3' 꼬리는 mRNA 합성 후에 첨가될 수 있다. 캡의 존재는 대부분의 진핵세포에서 발견되는 뉴클레아제에 대한 내성을 제공하는 데 있어서 중요하다. "꼬리"의 존재는 엑소뉴클레아제 분해로부터 mRNA를 보호하는 역할을 한다. 대안적으로, 5' 캡 및/또는 3' 꼬리 서열은 시험관 내 전사 반응에 사용되는 DNA 템플릿 서열에 포함된다.

[0385] 5' 캡은 전형적으로 다음과 같이 추가된다: 우선, RNA 말단 인산가수분해효소가 5' 뉴클레오티드로부터 말단 인산기 중 하나를 제거하고, 2개의 말단 인산기를 남긴다; 그런 다음, 구아노신 삼인산(GTP)이 구아닐릴 전이효소를 통해 말단 인산에 첨가되고 5'5'5 삼인산 결합을 생성한다; 그런 다음 구아닌의 7-질소는 메틸기 전이효소에 의해 메틸화된다. 캡 구조의 예는, m7G(5')ppp (5'(A,G(5')ppp(5')A and G(5')ppp(5')G를 포함하나, 이에 한정되지 않는다. 추가적인 캡 구조는 공개된 미국 출원 US 2016/0032356 및 공개된 미국 출원 US 2018/0125989에 기술되어 있으며, 이들은 참조로서 본원에 통합된다.

[0386] 일반적으로, 꼬리 구조는 폴리(A) 및/또는 폴리(C) 꼬리를 포함한다. mRNA의 3' 말단에 있는 폴리-A 꼬리 또는 폴리-C 꼬리는 각각 적어도 50개의 아데노신 또는 시토신 뉴클레오티드, 적어도 150개의 아데노신 또는 시토신 뉴클레오티드, 적어도 200개의 아데노신 또는 시토신 뉴클레오티드, 적어도 250개의 아데노신 또는 시토신 뉴클레오티드, 적어도 300개의 아데노신 또는 시토신 뉴클레오티드, 적어도 350개의 아데노신 또는 시토신 뉴클레오티드, 적어도 400개의 아데노신 또는 시토신 뉴클레오티드, 적어도 450개의 아데노신 또는 시토신 뉴클레오티드, 적어도 500개의 아데노신 또는 시토신 뉴클레오티드, 적어도 550개의 아데노신 또는 시토신 뉴클레오티드, 적어도 600개의 아데노신 또는 시토신 뉴클레오티드, 적어도 650개의 아데노신 또는 시토신 뉴클레오티드, 적어도 700개의 아데노신 또는 시토신 뉴클레오티드, 적어도 750개의 아데노신 또는 시토신 뉴클레오티드, 적어도 800개의 아데노신 또는 시토신 뉴클레오티드, 적어도 850개의 아데노신 또는 시토신 뉴클레오티드, 적어도 900개의 아데노신 또는 시토신 뉴클레오티드, 적어도 950개의 아데노신 또는 시토신 뉴클레오티드, 또는 적어도 1kb의 아데노신 또는 시토신 뉴클레오티드를 일반적으로 포함한다. 일부 구현예에서, 폴리A 꼬리 또는 폴리C 꼬리는 각각 약 10 내지 800개의 아데노신 또는 시토신 뉴클레오티드(예컨대, 약 10 내지 200

개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 300개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 400개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 500개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 550개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 50 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 100 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 150 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 200 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 250 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 300 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 350 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 400 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 450 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 500 내지 600개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 150개의 아데노신 또는 시토신 뉴클레오티드, 약 10 내지 100개의 아데노신 또는 시토신 뉴클레오티드, 약 20 내지 70개의 아데노신 또는 시토신 뉴클레오티드, 또는 약 20 내지 60개의 아데노신 또는 시토신 뉴클레오티드)일 수 있다. 일부 구현예에서, 꼬리 구조는 본원에서 설명된 다양한 길이를 갖는 폴리(A) 및 폴리(C) 꼬리의 조합을 포함한다. 일부 구현예에서, 꼬리 구조는 적어도 50%, 55%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, 또는 99% 아데노신 뉴클레오티드를 포함한다. 일부 구현예에서, 꼬리 구조는 적어도 50%, 55%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, 또는 99%의 시토신 뉴클레오티드를 포함한다.

[0387] *합성-후 정제*

[0388] 합성 후 mRNA를 정제하기 위해 다양한 방법이 사용될 수 있다. 일부 구현예에서, mRNA는 접선 유동 여과를 사용하여 정제된다. 적절한 정제 방법은, 그 모든 내용이 본원에 참조로서 통합되고 본 발명을 실시하는 데 사용될 수 있는, 공개된 미국 특허 출원 US 2016/0040154, 공개된 미국 특허 출원 US 2015/0376220, 공개된 미국 특허 출원 US 2018/0251755, published U.S. Application No. US 2018/0251754, 2018년 11월 8일자로 출원된 미국 가출원 제62/757,612호, 및 2019년 8월 26일자로 출원된 미국 가출원 제62/891,781호에 기술되어 있는 내용을 포함한다.

[0389] 일부 구현예에서, mRNA는 캡핑 및 테일링 전에 정제된다. 일부 구현예에서, mRNA는 캡핑 및 테일링 후에 정제된다. 일부 구현예에서, mRNA는 캡핑 및 테일링 전 및 후 모두에서 정제된다.

[0390] 일부 구현예에서, mRNA는 캡핑 및 테일링 전 또는 후, 또는 전 및 후 모두에 원심분리에 의해 정제된다.

[0391] 일부 구현예에서, mRNA는 캡핑 및 테일링 전 또는 후, 또는 전 및 후 모두에 여과에 의해 정제된다.

[0392] 일부 구현예에서, mRNA는 캡핑 및 테일링 전 또는 후, 또는 전 및 후 모두에 접선 유동 여과(TFF)에 의해 정제된다.

[0393] **지질 나노입자**

[0394] 본 발명에 따라, 본 발명의 최적화된 뉴클레오티드 서열을 포함하는 mRNA는 지질 나노입자로 전달될 수 있다. 일반적으로, 본 발명과 함께 사용하기에 적합한 지질 나노입자는 하나 이상의 양이온성 지질을 포함한다. 일부 구현예에서, 지질 나노입자는 하나 이상의 양이온성 지질, 하나 이상의 비-양이온성 지질, 하나 이상의 콜레스테롤계 지질, 및 하나 이상의 PEG-변형 지질을 포함한다. 일부 구현예에서, 지질 나노입자는 하나 이상의 양이온성 지질, 하나 이상의 비-양이온성 지질, 및 하나 이상의 PEG-변형 지질을 포함한다. 일부 구현예에서, 지질 나노입자는 4개 이하의 구별되는 지질 성분을 포함한다.

[0395] 본 발명과 함께 사용하기 위한 일반적인 지질 나노입자는 다음의 4가지 성분으로 구성된다: 양이온성 지질(예를 들어, 스테롤계 양이온성 지질, 비-양이온성 지질(예를 들어, DOPE 또는 DEPE), 콜레스테롤계 지질(예를 들어, 콜레스테롤), 및 PEG-변형 지질(예를 들어, DMG-PEG2K). 일부 구현예에서, 지질 나노입자는 3개 이하의 구별되는 지질 성분을 포함한다. 예시적인 지질 나노입자는 다음의 3가지 지질 성분으로 구성된다: 양이온성 지질(예를 들어, 스테롤계 양이온성 지질), 비-양이온성 지질(예를 들어, DOPE 또는 DEPE) 및 PEG-변형 지질(예를 들어, DMG-PEG2K).

[0396] *mRNA를 캡슐화하는 지질 나노입자의 형성*

[0397] 본 발명에 사용하기 위한 지질 나노입자는 현재 당업계에 공지된 다양한 기술을 사용하여 제조할 수 있다. 예를 들어, 다층 소낭(MLV)이 예컨대, 선택된 지질을 적합한 용기 또는 베슬(vessel)의 내벽 상에 침전시키고 적절한 용매에 지질을 용해시킨 후 용매를 증발시켜 베슬의 내부 상에 박막을 남기거나 분무 건조시키는 것과 같은 종래의 기법에 따라 제조될 수 있다. 그런 다음 소용돌이 운동으로 수성 상을 용기에 첨가하여 MLV를 형성할 수

있다. 그런 다음 단층 소낭(ULV)이 다층 소낭의 균질화, 초음파처리 또는 압출에 의해 형성될 수 있다. 추가적으로, 단층 소낭은 세제 제거 기법에 의해 형성될 수 있다.

[0398] 다양한 방법이 공개된 미국 특허 출원 US 2011/0244026, 공개된 미국 특허 출원 US 2016/0038432, 공개된 미국 특허 출원 US 2018/0153822호, 공개된 미국 특허 출원 US 2018/0125989호, 및 2019년 7월 23일에 출원한 미국 특허 가출원 제62/877,597호에 기술되어 있고, 이들 모두는 본 발명을 실시하는 데 사용될 수 있으며, 이들 문헌 모두는 본원에 참조로서 통합된다. 본원에서 사용되는 바와 같이, 프로세스 A는, 미국 특허 US 2016/0038432에 기술된 바와 같이, 먼저 지질을 지질 나노입자로 사전에 형성하지 않고, mRNA를 지질의 혼합물과 혼합함으로써 아를 캡슐화하는 종래의 방법을 지칭한다. 본원에서 사용되는 바와 같이, 프로세스 B는, US 2018/0153822에 기술된 바와 같이, 사전에 형성된 지질 나노입자를 mRNA와 혼합함으로써 mRNA를 캡슐화하는 프로세스를 지칭한다.

[0399] 간략하게는, mRNA가 로딩된 지질 나노입자를 제조하는 프로세스는 용액 중 하나 이상을 주변 온도보다 높은 온도로 가열하는 (즉, 열원으로부터 용액에 열을 가하는) 단계(또는 가열하여 주변 온도보다 높은 온도로 유지시키는 단계)를 포함하며, 여기에서 하나 이상의 용액은 사전에 형성된 지질 나노입자를 포함하는 용액, mRNA를 포함하는 용액, 및 지질 나노입자로 캡슐화된 mRNA를 포함하는 혼합 용액이다. 일부 구현예에서, 방법은 mRNA 용액 및 사전에 형성된 지질 나노입자 용액 중 하나 또는 둘 모두를 가열하는 단계를 혼합 단계 전에 포함한다. 일부 구현예에서, 방법은 사전에 형성된 지질 나노입자를 포함하는 용액, mRNA를 포함하는 용액, 및 지질 나노입자로 캡슐화된 mRNA를 포함하는 용액 중 하나 이상을 가열하는 단계를 혼합 단계 동안에 포함한다. 일부 구현예에서, 방법은 지질 나노입자로 캡슐화된 mRNA를 가열하는 단계를 혼합 단계 후에 포함한다. 일부 구현예에서, 용액 중 하나 이상을 가열하는 온도(또는 용액 중 하나 이상이 유지되는 온도)는 약 30℃, 37℃, 40℃, 45℃, 50℃, 55℃, 60℃, 65℃, 또는 70℃이상이다. 일부 구현예에서, 용액 중 하나 이상을 가열하는 온도는 약 25 내지 70℃ 약 30 내지 70℃ 약 35 내지 70℃ 약 40 내지 70℃ 약 45 내지 70℃ 약 50 내지 70℃ 또는 약 60 내지 70℃의 범위이다. 일부 구현예에서, 용액 중 하나 이상을 가열하는, 주변 온도보다 높은 온도는 약 65℃이다.

[0400] 본 발명에 적절한 mRNA 용액은 다양한 방법을 사용하여 제조할 수 있다. 일부 구현예에서, mRNA는 본원에 기술된 완충액에 직접 용해될 수 있다. 일부 구현예에서, mRNA 용액은, 캡슐화를 위해 mRNA 용액을 지질 용액과 혼합하기 전에 mRNA 스톡 용액과 완충액을 혼합함으로써 생성할 수 있다. 일부 구현예에서, mRNA 용액은, 캡슐화를 위해 mRNA 용액을 지질 용액과 혼합하기 직전에 mRNA 스톡 용액과 완충액을 혼합함으로써 생성할 수 있다. 예를 들어, 적절한 mRNA 스톡 용액은 물 중 mRNA를 약 0.2 mg/ml, 0.4 mg/ml, 0.5 mg/ml, 0.6 mg/ml, 0.8 mg/ml, 1.0 mg/ml, 1.2 mg/ml, 1.4 mg/ml, 1.5 mg/ml, 또는 1.6 mg/ml, 2.0 mg/ml, 2.5 mg/ml, 3.0 mg/ml, 3.5 mg/ml, 4.0 mg/ml, 4.5 mg/ml, 또는 5.0 mg/ml 이상의 농도로 함유할 수 있다.

[0401] 일부 구현예에서, mRNA 스톡 용액은 펌프를 사용하여 완충액과 혼합된다. 예시적인 펌프는 기어 펌프, 연동 펌프, 및 원심 펌프를 포함하지만, 이에 한정되지는 않는다.

[0402] 일반적으로, 완충액은 mRNA 스톡 용액의 속도보다 더 높은 속도로 혼합된다. 예를 들어, 완충액은 mRNA 스톡 용액의 속도보다 적어도 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 또는 20x 더 높은 속도로 혼합될 수 있다. 일부 구현예에서, 완충액은 약 100 내지 6000 ml/분(예를 들어, 약 100 내지 300 ml/분, 300 내지 600 ml/분, 600 내지 1200 ml/분, 1200 내지 2400 ml/분, 2400 내지 3600 ml/분, 3600 내지 4800 ml/분, 4800 내지 6000 ml/분, 또는 60 내지 420 ml/분) 범위의 유속으로 혼합된다. 일부 구현예에서, 완충액은 약 60 ml/분, 100 ml/분, 140 ml/분, 180 ml/분, 220 ml/분, 260 ml/분, 300 ml/분, 340 ml/분, 380 ml/분, 420 ml/분, 480 ml/분, 540 ml/분, 600 ml/분, 1200 ml/분, 2400 ml/분, 3600 ml/분, 4800 ml/분, 또는 6000 ml/분 이상의 유속으로 혼합된다.

[0403] 일부 구현예에서, mRNA 스톡 용액은 약 10 내지 600 ml/분(예를 들어, 약 5 내지 50 ml/분, 약 10 내지 30 ml/분, 약 30 내지 60 ml/분, 약 60 내지 120 ml/분, 약 120 내지 240 ml/분, 약 240 내지 360 ml/분, 약 360 내지 480 ml/분, 또는 약 480 내지 600 ml/분) 범위의 유속으로 혼합된다. 일부 구현예에서, mRNA 스톡 용액은 약 5 ml/분, 10 ml/분, 15 ml/분, 20 ml/분, 25 ml/분, 30 ml/분, 35 ml/분, 40 ml/분, 45 ml/분, 50 ml/분, 60 ml/분, 80 ml/분, 100 ml/분, 200 ml/분, 300 ml/분, 400 ml/분, 500 ml/분, 또는 600 ml/분 이상의 유속으로 혼합된다.

[0404] 본 발명에 따르면, 지질 용액은 mRNA의 캡슐화를 위한 지질 나노입자를 형성하기에 적합한 지질의 혼합물을 함유한다. 일부 구현예에서, 적절한 지질 용액은 에탄올계이다. 예를 들어, 적합한 지질 용액은 순수 에탄올(즉, 100% 에탄올)에 용해된 바람직한 지질의 혼합물을 함유할 수 있다. 또 다른 구현예에서, 적절한 지질 용액은 이

소프로필 알코올계이다. 또 다른 구현예에서, 적절한 지질 용액은 디메틸설폭시드계이다. 또 다른 구현예에서, 적절한 지질 용액은 에탄올, 이소프로필 알코올, 및 디메틸설폭시드를 포함하지만 이에 한정되지 않는 적절한 용매의 혼합물이다.

[0405] 적절한 지질 용액은 바람직한 지질의 혼합물을 다양한 농도로 함유할 수 있다. 예를 들어, 적절한 지질 용액은 바람직한 지질의 혼합물을 약 0.1 mg/ml, 0.5 mg/ml, 1.0 mg/ml, 2.0 mg/ml, 3.0 mg/ml, 4.0 mg/ml, 5.0 mg/ml, 6.0 mg/ml, 7.0 mg/ml, 8.0 mg/ml, 9.0 mg/ml, 10 mg/ml, 15 mg/ml, 20 mg/ml, 30 mg/ml, 40 mg/ml, 50 mg/ml, 또는 100 mg/ml 이상의 총 농도로 함유할 수 있다. 일부 구현예에서, 적절한 지질 용액은 바람직한 지질의 혼합물을 약 0.1 내지 100 mg/ml, 0.5 내지 90 mg/ml, 1.0 내지 80 mg/ml, 1.0 내지 70 mg/ml, 1.0 내지 60 mg/ml, 1.0 내지 50 mg/ml, 1.0 내지 40 mg/ml, 1.0 내지 30 mg/ml, 1.0 내지 20 mg/ml/ml, 1.0 내지 15 mg/ml/ml, 1.0 내지 10 mg/ml, 1.0 내지 9 mg/ml/ml, 1.0 내지 8 mg/ml, 1.0 내지 7 mg/ml, 1.0 내지 6ml, 또는 1.0 내지 5 mg/ml 범위의 총 농도로 함유할 수 있다. 일부 구현예에서, 적절한 지질 용액은 바람직한 지질의 혼합물을 최대 약 100 mg/ml, 90 mg/ml, 80 mg/ml, 70 mg/ml, 60 mg/ml, 50 mg/ml, 40 mg/ml, 30 mg/ml, 20 mg/ml, 또는 10 mg/ml의 총 농도로 함유할 수 있다.

[0406] 임의의 바람직한 지질이 mRNA를 캡슐화하기에 적합한 임의의 비율로 혼합될 수 있다. 일부 구현예에서, 적합한 지질 용액은 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 양친매성 블록 공중합체(예를 들어, 폴록사머), 및/또는 PEG-변형 지질을 포함하는 바람직한 지질의 혼합물을 함유한다. 일부 구현예에서, 적합한 지질 용액은 하나 이상의 양이온성 지질, 하나 이상의 비-양이온성 지질, 하나 이상의 콜레스테롤계 지질, 및/또는 하나 이상의 PEG-변형 지질을 포함하는 바람직한 지질의 혼합물을 함유한다.

[0407] 일부 구현예에서, 제공된 약학적 조성물은 지질 나노입자를 포함하되, mRNA는 지질 나노입자의 양 표면 상에 결합되어 동일한 지질 나노입자 내에 캡슐화된다. 예를 들어, 본 발명의 약학적 조성물을 제조하는 동안, 양이온성 지질 나노입자는 정전기적 상호작용을 통해 mRNA와 결합할 수 있다.

[0408] 일부 구현예에서, 본 발명의 화합물, 약학적 조성물 및 방법은 지질 나노입자 내에 캡슐화된 mRNA를 포함한다. 일부 구현예에서, mRNA는 동일한 지질 나노입자에 캡슐화된다. 일부 구현예에서, mRNA는 상이한 지질 나노입자에 캡슐화된다. 일부 구현예에서, mRNA는 지질 조성, 지질 성분의 몰비, 크기, 전하(제타 전위), 표적 리간드 및/또는 이들의 조합이 다른 하나 이상의 지질 나노입자 내에 캡슐화된다. 일부 구현예에서, 하나 이상의 지질 나노입자는 스테롤계 양이온 지질, 중성 지질, PEG-변형 지질 및/또는 이들의 조합의 상이한 조성을 가질 수 있다. 일부 구현예에서, 하나 이상의 지질 나노입자는 지질 나노입자를 생성하는 데 사용된 콜레스테롤계 지질, 양이온성 지질, 중성 지질, 및 PEG-변형 지질의 상이한 몰비를 가질 수 있다.

[0409] 원하는 mRNA를 지질 나노입자 내에 혼입하는 프로세스는 종종 "로딩(loading)"으로서 지칭된다. 예시적인 방법은 본원에 참조로서 통합된 Lazorthes 등의 문헌[FEBS Lett., 312: 255-258, 1992]에 기술되어 있다. 지질 나노입자-혼입 핵산은 지질 나노입자의 내부 공간, 지질 나노입자의 이중층 막 내에 완전히 또는 부분적으로 위치하거나, 지질 나노입자 막의 외부 표면과 연관될 수 있다. 지질 나노입자 내로 mRNA를 통합하는 것을 또한 본원에서 "캡슐화"라고 지칭하며, 여기서 핵산은 지질 나노입자의 내부 공간에 완전히 함유된다. 지질 나노입자 내로 mRNA를 통합하는 목적은 종종 mRNA를 분해하는 효소 또는 화학물질 및/또는 mRNA의 급속한 분비를 야기하는 시스템 또는 수용체를 함유할 수 있는 환경으로부터 mRNA를 보호하기 위한 것이다. 따라서, 일부 구현예에서, 적합한 지질 나노입자는 그 내부에 포함된 mRNA의 안정성을 강화시킬 수 있고/있거나 표적 세포 또는 조직으로 mRNA의 전달을 용이하게 할 수 있다.

[0410] 본 발명에 따른 적합한 지질 나노입자는 다양한 크기로 만들어질 수 있다. 일부 구현예에서, 제공된 지질 나노입자는 이전에 알려진 지질 나노입자보다 더 작게 만들어질 수 있다. 일부 구현예에서, 지질 나노입자의 크기 감소는 mRNA의 더 효율적인 전달과 연관된다. 적절한 지질 나노입자 크기를 선택하는 데에는 표적 세포 또는 조직의 부위가 고려될 수 있고, 제작되는 지질 나노입자의 용도가 어느 정도 고려될 수 있다.

[0411] 일부 구현예에서, 지질 나노입자의 적절한 크기는 mRNA의 전신 분포를 촉진시키도록 선택된다. 대안적으로 또는 추가적으로, 지질 나노입자의 치수가 특정 세포 또는 조직으로의 분포를 제한하거나 분명히 피하기에 충분한 직경이 되도록 지질 나노입자의 크기가 조정될 수 있다.

[0412] 당업계에 알려진 다양한 대안적 방법이 지질 나노입자 집단의 크기 조정을 위해 사용될 수 있다. 하나의 이와 같은 크기 조정 방법이 미국 특허 제4,737,323호에 기술되어 있고, 이는 본원에 참조로서 포함된다. 배스(bath) 또는 프로브(probe) 초음파처리기를 사용하여 지질 나노입자 현탁액을 초음파처리함으로써 직경 약 0.05

미크론 미만의 작은 ULV로 점차 크기를 줄이게 된다. 균질화는 큰 지질 나노입자를 보다 작은 지질 나노입자로 분화시키는 전단 에너지에 의존하는 또 다른 방법이다. 전형적인 균질화 과정은, 일반적으로 약 0.1 내지 0.5 미크론의 선택된 지질 나노입자 크기가 관찰될 때까지, MLV를 표준 에멀전 균질화기를 통해 재순환시킨다. 지질 나노입자의 크기는 본원에 참조로서 포함되는 Bloomfield의 문헌[Ann. Rev. Biophys. Bioeng., 10:421-450 (1981)]에 기술된 바와 같이 준탄성 광산란(QELS)에 의해 결정될 수 있다. 지질 나노입자의 평균 직경은 형성된 지질 나노입자의 초음파처리에 의해 줄어들 수 있다. 간헐적인 초음파 처리 사이클은 효율적인 지질 나노입자 합성을 유도하기 위해 QELS 평가와 교대로 할 수 있다.

[0413] *지질 나노입자 제형*

[0414] 일부 구현예에서, 약학적 조성물 중 정제된 지질 나노입자의 대부분, 즉 지질 나노입자의 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%는 약 150 nm(예를 들어, 약 145 nm, 약 140 nm, 약 135 nm, 약 130 nm, 약 125 nm, 약 120 nm, 약 115 nm, 약 110 nm, 약 105 nm, 약 100 nm, 약 95 nm, 약 90 nm, 약 85 nm, 또는 약 80 nm)의 크기를 갖는다. 일부 구현예에서, 정제된 지질 나노입자의 실질적으로 전부는 약 150 nm(예를 들어, 약 145 nm, 약 140 nm, 약 135 nm, 약 130 nm, 약 125 nm, 약 120 nm, 약 115 nm, 약 110 nm, 약 105 nm, 약 100 nm, 약 95 nm, 약 90 nm, 약 85 nm, 또는 약 80 nm)의 크기를 갖는다.

[0415] 일부 구현예에서, 지질 나노입자는 150 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 120 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 100 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 90 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 80 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 70 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 60 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 50 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 30 nm 미만의 평균 크기를 갖는다. 일부 구현예에서, 지질 나노입자는 20 nm 미만의 평균 크기를 갖는다.

[0416] 일부 구현예에서, 본 발명에 의해 제공된 약학적 조성물 중 지질 나노입자의 약 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 초과는 약 40 내지 90 nm(예를 들어, 약 45 내지 85 nm, 약 50 내지 80 nm, 약 55 내지 75 nm, 약 60 내지 70 nm) 범위의 크기를 갖는다. 일부 구현예에서, 지질 나노입자의 실질적으로 전부는 약 40 내지 90 nm(예를 들어, 약 45 내지 85 nm, 약 50 내지 80 nm, 약 55 내지 75 nm, 약 60 내지 70 nm) 범위의 크기를 갖는다. 약 50 내지 70 nm(예를 들어, 55 내지 65 nm)의 평균 크기를 갖는 지질 나노입자를 갖는 조성물은 분무화를 통한 폐 전달에 특히 적합하다.

[0417] 일부 구현예에서, 본 발명에 의해 제공되는 약학적 조성물 중의 지질 나노입자의 분산성 또는 분자 크기의 이질성(PDI)의 척도는 약 0.5 미만이다. 일부 구현예에서, 지질 나노입자는 약 0.5 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.4 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.3 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.28 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.25 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.23 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.20 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.18 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.16 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.14 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.12 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.10 미만의 PDI를 갖는다. 일부 구현예에서, 지질 나노입자는 약 0.08 미만의 PDI를 갖는다.

[0418] 일부 구현예에서, 본 발명에 의해 제공되는 약학적 조성물 중의 정제된 지질 나노입자의 약 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 초과는 각각의 개별 입자 내에 mRNA를 캡슐화한다. 일부 구현예에서, 약학적 조성물 중의 정제된 지질 나노입자의 실질적으로 전부는 각각의 개별 입자 내에 mRNA를 캡슐화한다. 일부 구현예에서, 지질 나노입자는 50% 내지 99%의 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 60%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 65%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 70%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 75%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 80%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 85%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 90%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 92%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 95%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 98%를 초과하는 캡슐화 효율을 갖는다. 일부 구현예에서, 지질 나노입자는 약

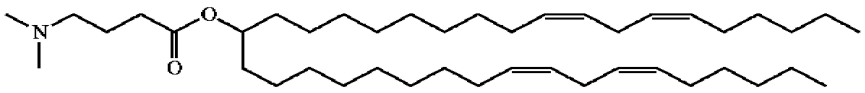
99%를 초과하는 캡슐화 효율을 갖는다. 일반적으로, 본 발명과 함께 사용하기 위한 지질 나노입자는 적어도 90% 내지 95%의 캡슐화 효율을 갖는다.

[0419] 일부 구현예에서, 지질 나노입자는 1 내지 10의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 1을 초과하는 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 1의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 2의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 3의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 4의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 5의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 6의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 7의 N/P 비율을 갖는다. 일부 구현예에서, 지질 나노입자는 약 8의 N/P 비율을 갖는다. 본 발명과 함께 사용하기 위한 일반적인 지질 나노입자는 약 4의 N/P 비율을 갖는다.

[0420] 일부 구현예에서, 본 발명에 따른 약학적 조성물은 적어도 약 0.5 mg, 1 mg, 5 mg, 10 mg, 100 mg, 500 mg, 또는 1000 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 0.1 mg 내지 1000 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 0.5 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 0.8 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 1 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 5 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 8 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 10 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 50 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 100 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 500 mg의 캡슐화된 mRNA를 함유한다. 일부 구현예에서, 약학적 조성물은 약 1000 mg의 캡슐화된 mRNA를 함유한다.

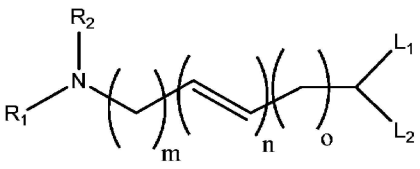
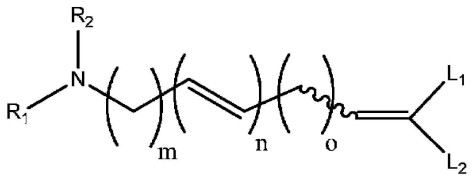
[0421] *양이온성 지질*

[0422] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 양이온성 지질은 본원에 참조로서 포함되는 국제 특허 공개 WO 2010/144740에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 (6Z,9Z,28Z,31Z)-헵타트리아콘타-6,9,28,31-테트라엔-19-일 4-(디메틸아미노) 부타노에이트 및 이의 약학적으로 허용 가능한 염을 포함한다:

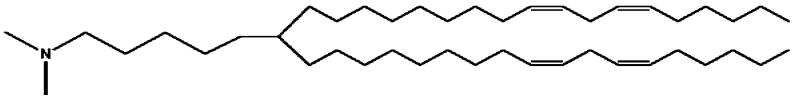
[0423] 

[0424]

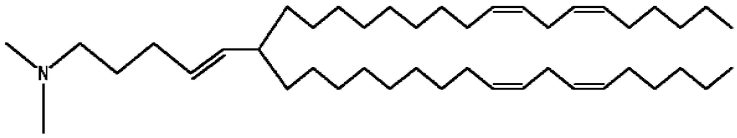
[0425] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2013/149140호에 기재된 이온화 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화학식 중 하나의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함하며:

[0426] , 

[0427] 식 중, $R_1$ 및 $R_2$는 각각 독립적으로 수소, 선택적으로 치환된 가변 포화 또는 불포화 $C_1$-$C_{20}$ 알킬, 및 선택적으로 치환된 가변 포화 또는 불포화 $C_6$-$C_{20}$ 아실로 이루어진 군에서 선택되고, $L_1$ 및 $L_2$는 각각 독립적으로 수소, 선택적으로 치환된 $C_1$-$C_{30}$ 알킬, 선택적으로 치환된 가변 불포화 $C_1$-$C_{30}$ 알케닐, 및 선택적으로 치환된 $C_1$-$C_{30}$ 알키닐로 이루어진 군에서 선택되며, m 및 o는 각각 독립적으로 0 및 임의의 양의 정수(예를 들어, m은 3)로 이루어진 군에서 선택되고, n은 0이거나 임의의 양의 정수(예를 들어, n은 1)이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 (15Z, 18Z)-N,N-디메틸-6-(9Z,12Z)-옥타데카-9,12-디엔-1-일) 테트라코사-15,18-디엔-1-아민("HGT5000") 및 이의 약학적으로 허용가능한 염을 포함한다:

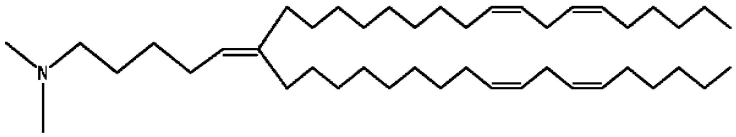
[0428]  (HGT-5000)

[0429] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 (15Z,18Z)-N,N-디메틸-6-((9Z,12Z)-옥타데카-9,12-디엔-1-일) 테트라코사-4,15,18-트리엔-1-아민("HGT5001") 및 이의 약학적으로 허용 가능한 염을 포함한다:



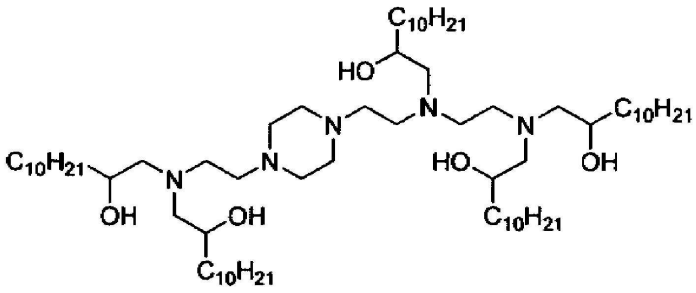
[0430] ' (HGT-5001)

[0431] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 (15Z,18Z)-N,N-디메틸-6-((9Z,12Z)-옥타데카-9,12-디엔-1-일) 테트라코사-5,15,18-트리엔-1-아민("HGT5002") 및 이의 약학적으로 허용 가능한 염을 포함한다:



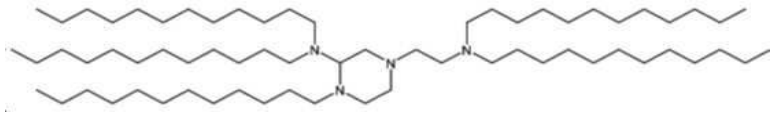
[0432] (HGT-5002)

[0433] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2010/053572에 아미노 알코올 리피도이드로 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



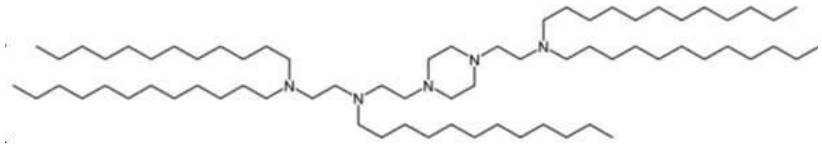
[0434]

[0435] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2016/118725에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0436]

[0437] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2016/118724에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:

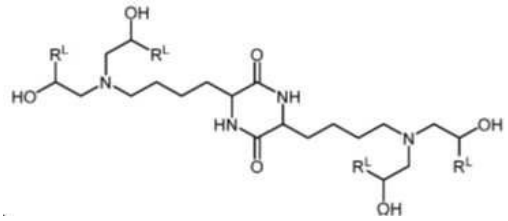


[0438]

[0439] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 이온은 14,25-디트리데실 15,18,21,24-테트라아자-옥타트리아콘탄의 화학식을 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다.

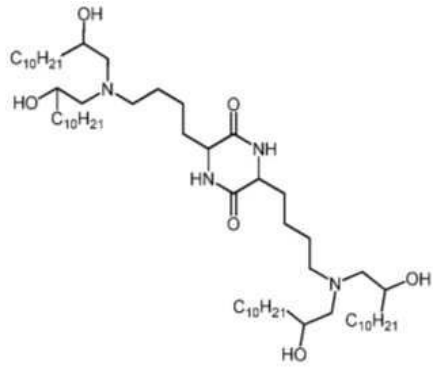
[0440] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 둘 모두 본원에 참조로서 포함되는 국제 특허 공개 WO 2013/063468 및 WO 2016/205691에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함

하며:



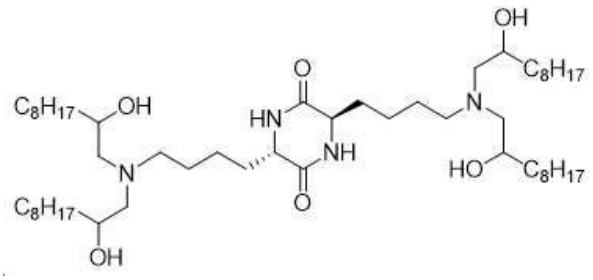
[0441]

[0442] 식 중, $R^L$의 각각의 경우는 독립적으로, 선택적으로 치환된 $C_6$-$C_{40}$ 알케닐이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



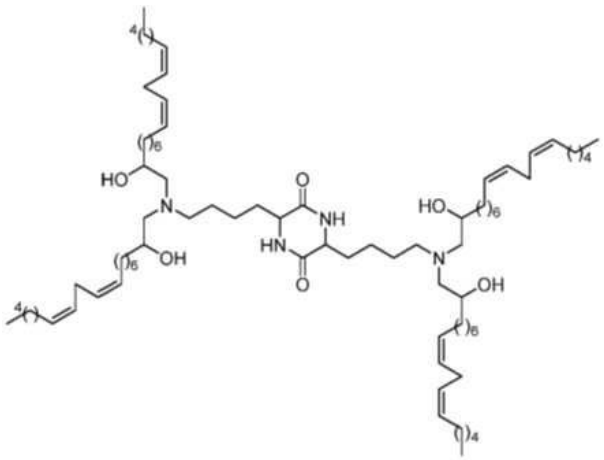
[0443] (cKK-E12).

[0444] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



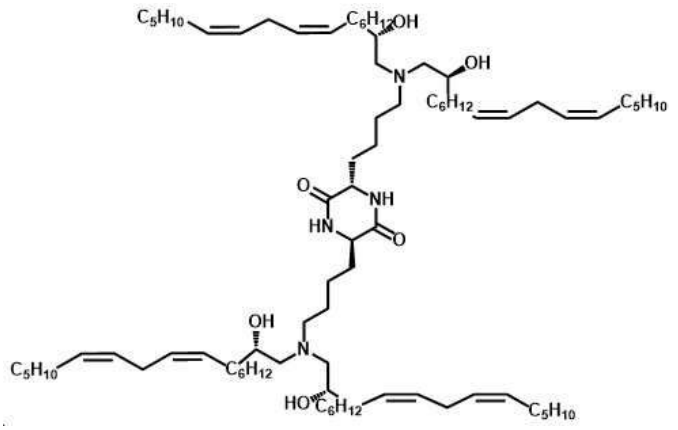
[0445] (cKK-E10)

[0446] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



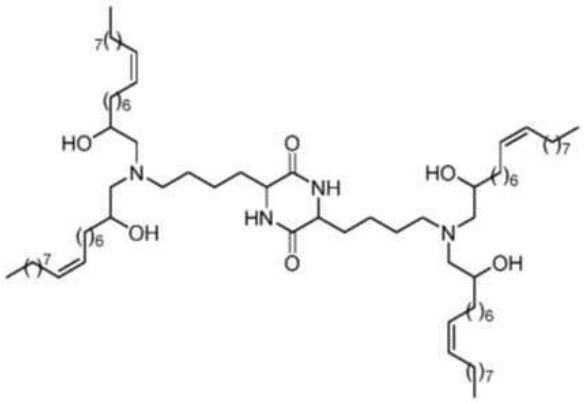
[0447]

[0448] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



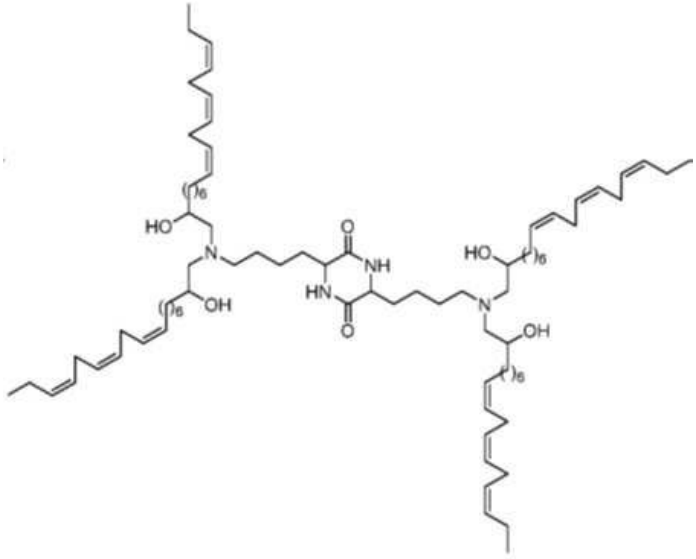
[0449] (OF-02)

[0450] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



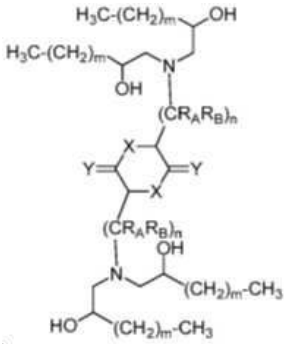
[0451]

[0452] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



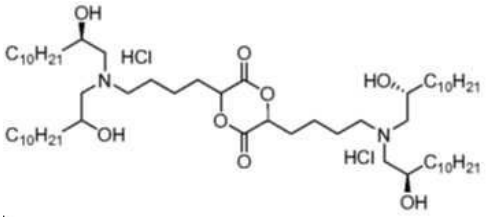
[0453]

[0454] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2015/184256에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함하며:



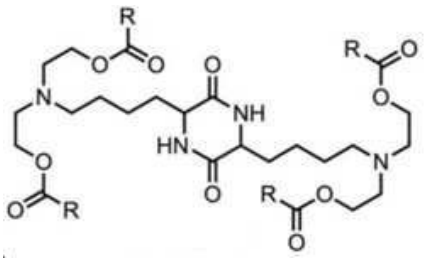
[0455]

[0456] 식 중, 각각의 X는 독립적으로 O 또는 S이고, 각각의 Y는 독립적으로 O 또는 S이고, 각각의 m은 독립적으로 0 내지 20이고, 각각의 n은 독립적으로 1 내지 6이고, 각각의 $R_A$는 독립적으로 수소, 선택적으로 치환된 C1-50 알킬, 선택적으로 치환된 C2-50 알케닐, 선택적으로 치환된 C2-50 알키닐, 선택적으로 치환된 C3-10 카르보시클릴, 선택적으로 치환된 3-14원 헤테로시클릴, 선택적으로 치환된 C6-14 아릴, 선택적으로 치환된 5-14원 헤테로아릴 또는 할로겐이고, 각각의 $R_B$는 독립적으로 수소, 선택적으로 치환된 C1-50 알킬, 선택적으로 치환된 C2-50 알케닐, 선택적으로 치환된 C2-50 알키닐, 선택적으로 치환된 C3-10 카보시클릴, 선택적으로 치환된 3-14원 헤테로시클릴, 선택적으로 치환된 C6-14 아릴, 선택적으로 치환된 5-14원 헤테로아릴 또는 할로겐이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "표적 23" 및 이의 약학적으로 허용가능한 염을 포함한다:

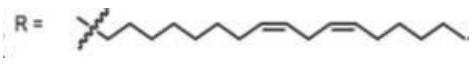


[0457] (표적 23)

[0458] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2016/004202에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:

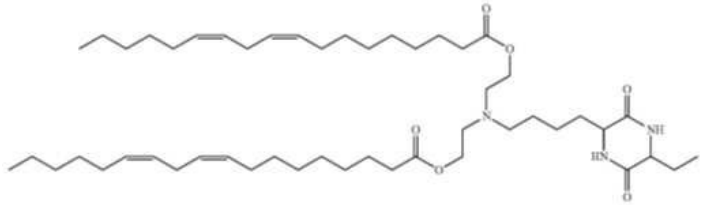


[0459]



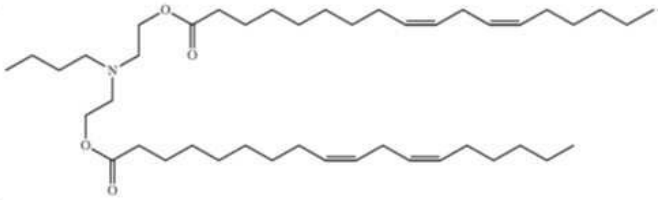
[0460] (OF-Deg-Lin)

[0461] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



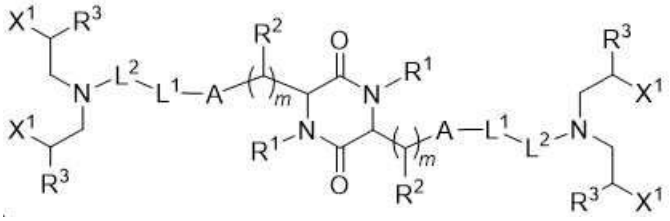
[0462]

[0463] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



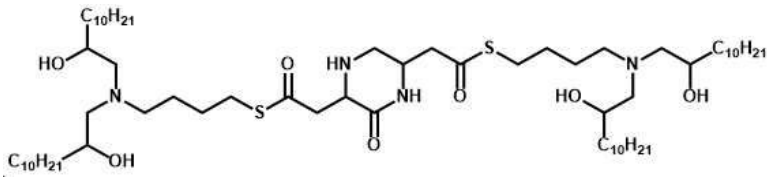
[0464]

[0465] 본 발명의 약학적 조성물 및 방법에 사용하기 적합한 다른 양이온성 지질은, 본원에 참조로서 통합되는, 2018년 11월 9일자로 출원된 미국 특허 가출원 제62/758,179호 및 2019년 7월 8일자로 출원된 미국 특허 가출원 제62/871,510호에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함하며:



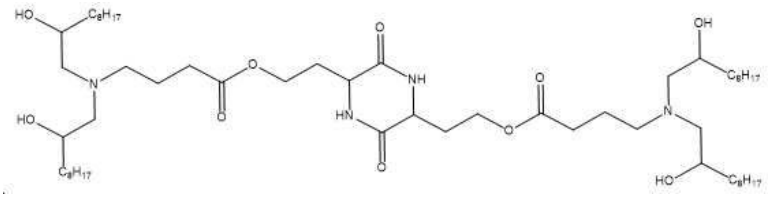
[0466]

[0467] 식 중 각각의 $R^1$ 및 $R^2$는 독립적으로 H 또는 $C_1$-$C_6$ 지방족이고; 각각의 $m$은 독립적으로 1 내지 4의 값을 갖는 정수이고; 각각의 A는 독립적으로 공유 결합 또는 아릴렌이고; 각각의 $L^1$은 독립적으로 에스테르 기, 티오에스테르 기, 이황화 기 또는 무수물 기이고; 각각의 $L^2$는 독립적으로 $C_2$-$C_{10}$ 지방족이고; 각각의 $X^1$은 독립적으로 H 또는 OH이며; 각각의 $R^3$은 독립적으로 $C_6$-$C_{20}$이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함한다:



[0468] (화합물 1)

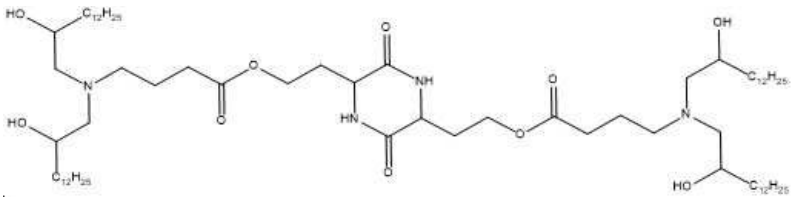
[0469] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함한다:



[0470]

[0471] (화합물 2; cHse-E-3-E10)

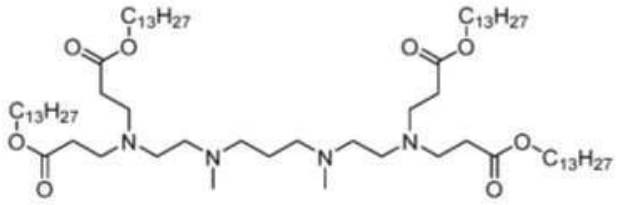
[0472] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함한다:



[0473] (화합물 3)

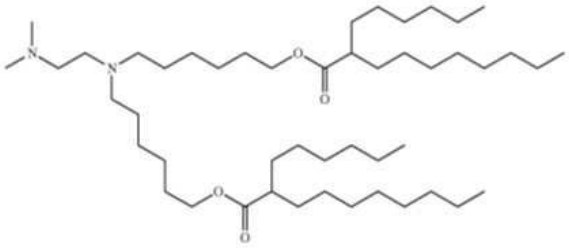
[0474] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 통합된 J. McClellan, M. C. King의 문헌[Cell 2010, 141, 210-217] 및 Whitehead 등의 문헌[Nature Communications

(2014) 5:4277]에 기술된 것과 같은 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법의 양이온성 지질은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



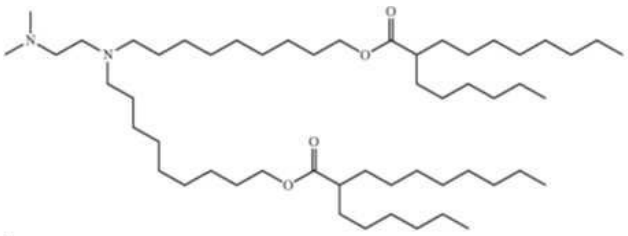
[0475]

[0476] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2015/199952에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



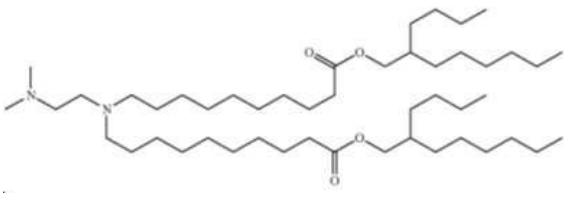
[0477]

[0478] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



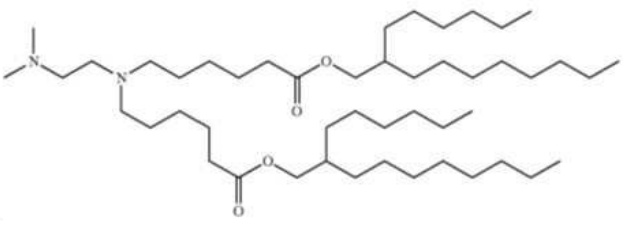
[0479]

[0480] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0481]

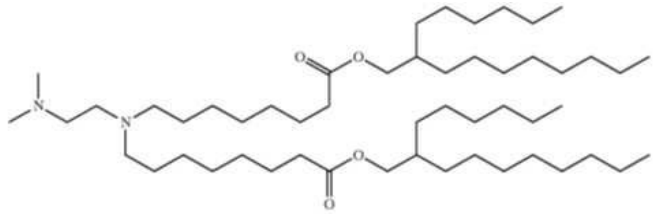
[0482] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0483]

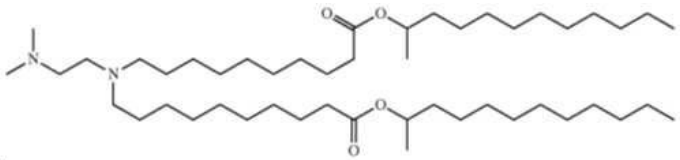
[0484] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학

적으로 허용 가능한 염을 포함한다:



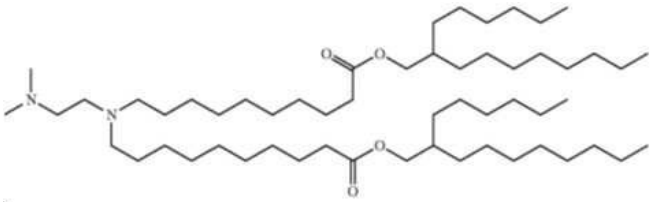
[0485]

[0486] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



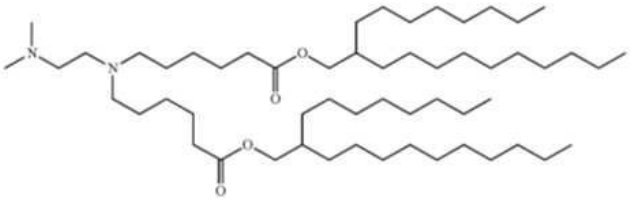
[0487]

[0488] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



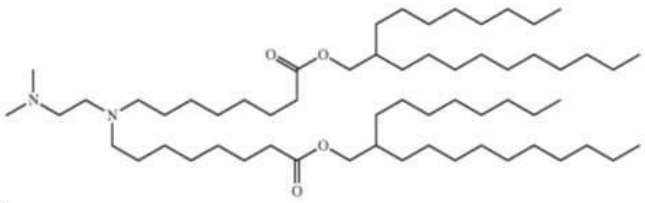
[0489]

[0490] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



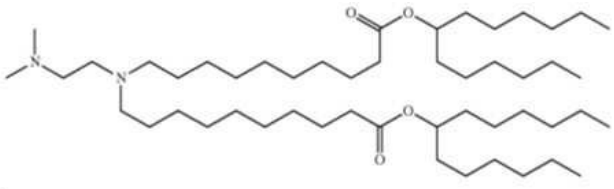
[0491]

[0492] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



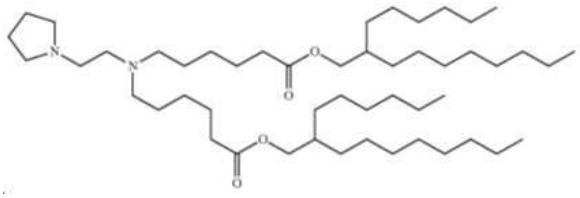
[0493]

[0494] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



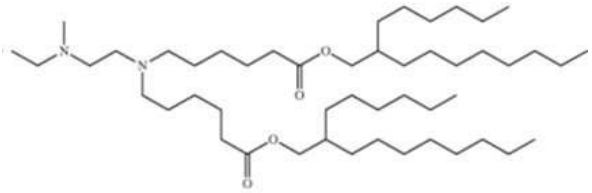
[0495]

[0496] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



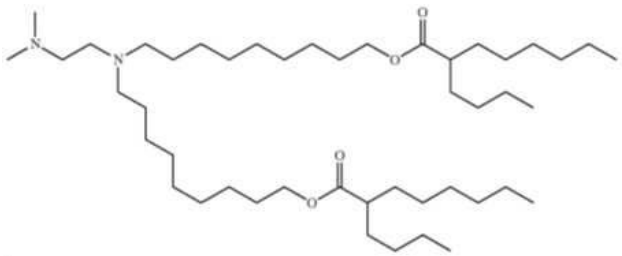
[0497]

[0498] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



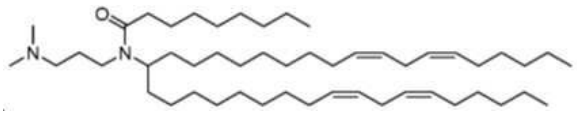
[0499]

[0500] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



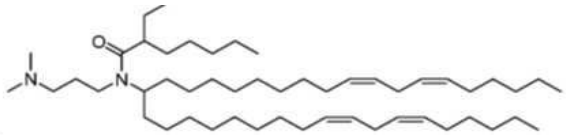
[0501]

[0502] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2017/004143에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



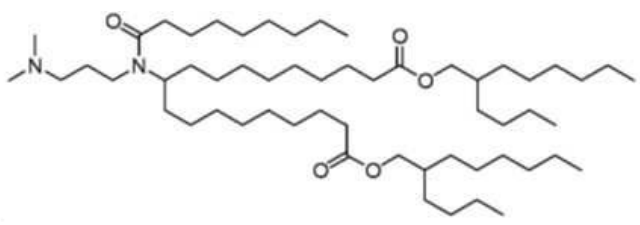
[0503]

[0504] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0505]

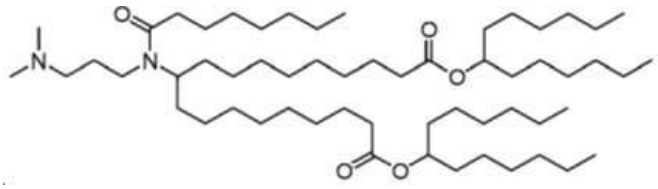
[0506] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0507]

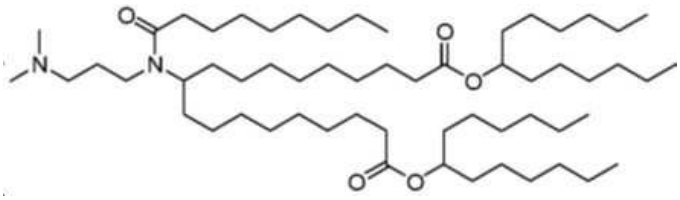
[0508] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학

적으로 허용 가능한 염을 포함한다:



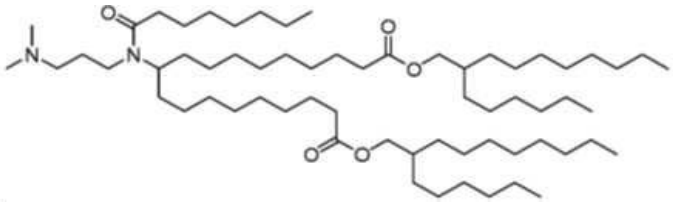
[0509]

[0510] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



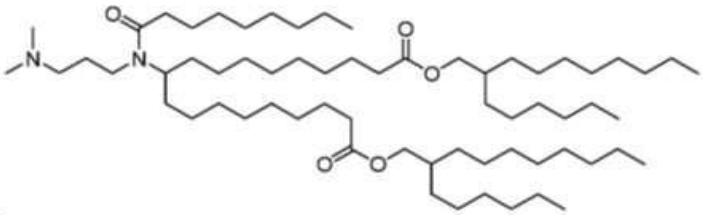
[0511]

[0512] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



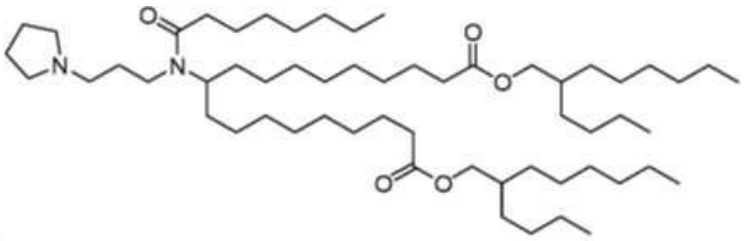
[0513]

[0514] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



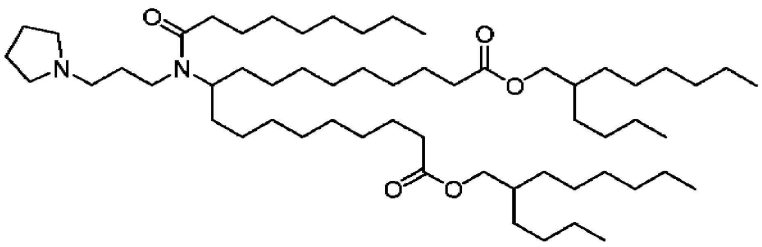
[0515]

[0516] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



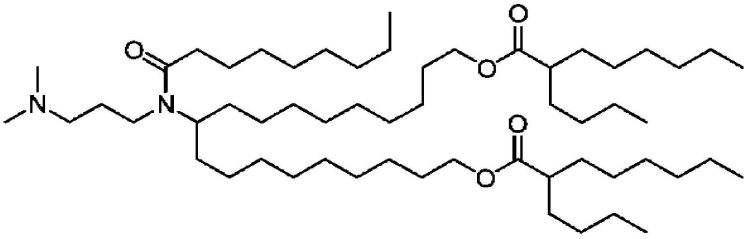
[0517]

[0518] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



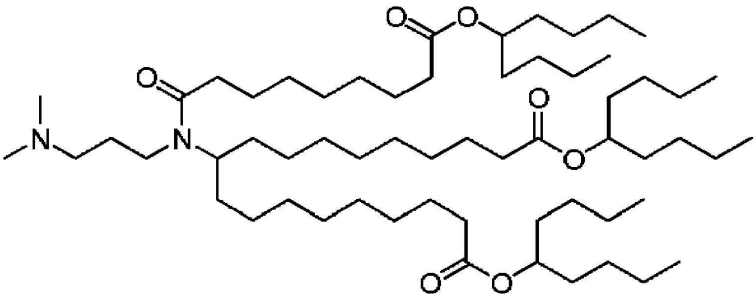
[0519]

[0520] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



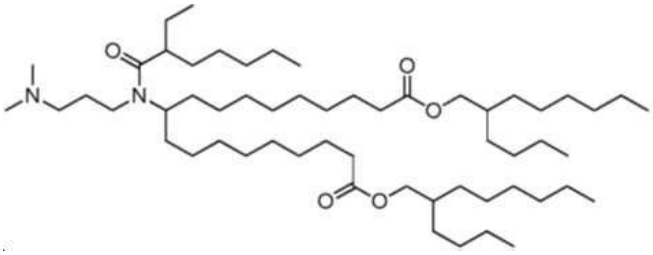
[0521]

[0522] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



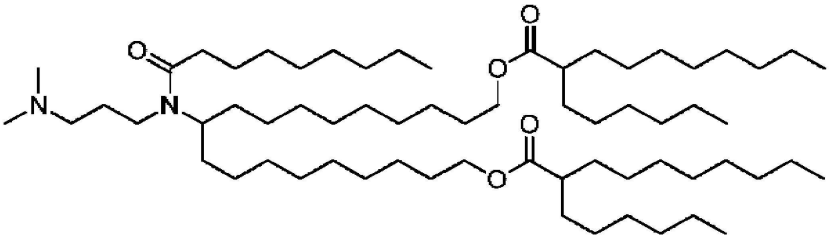
[0523]

[0524] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



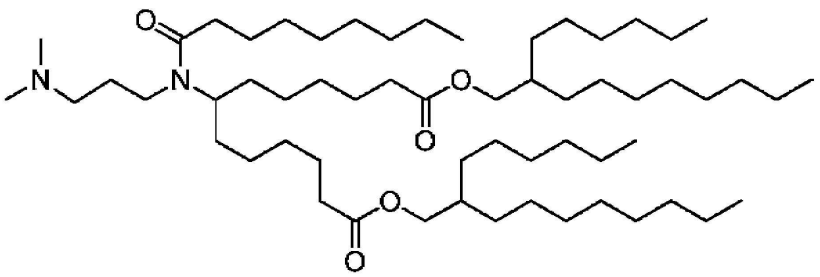
[0525]

[0526] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



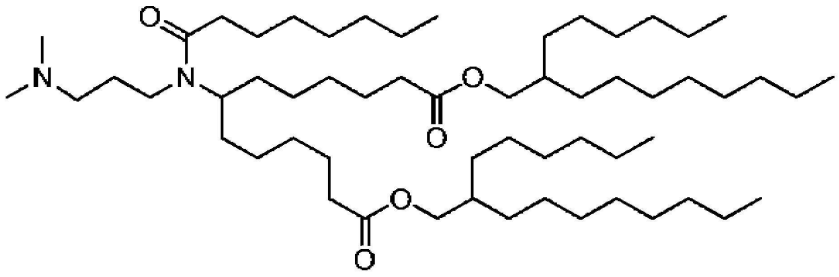
[0527]

[0528] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



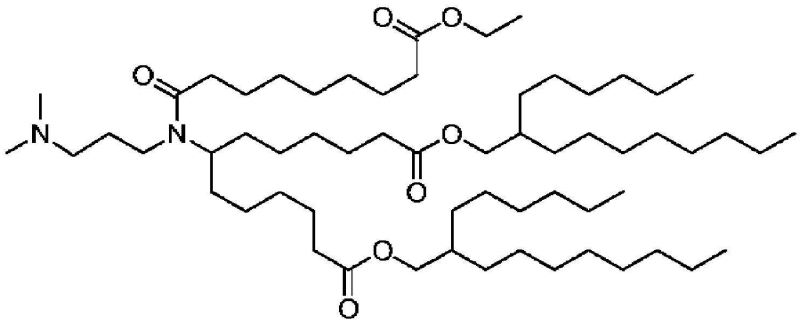
[0529]

[0530] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



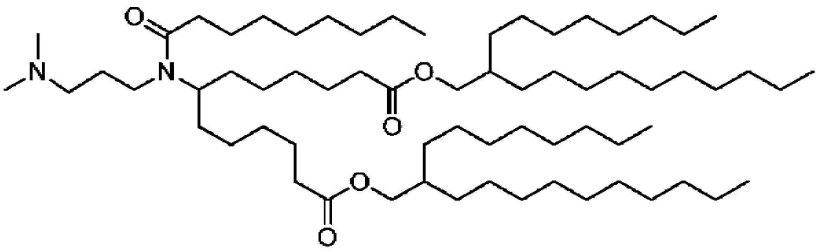
[0531]

[0532] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0533]

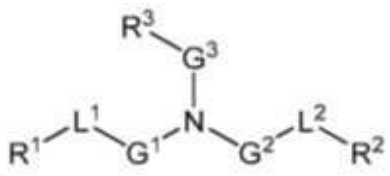
[0534] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0535]

[0536]

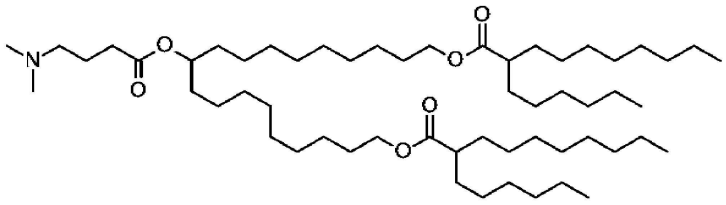
[0537] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2017/075531에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함하며:



[0538]

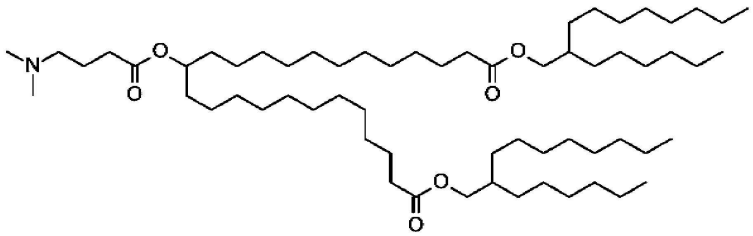
[0539] 식 중, $L^1$ 또는 $L^2$ 중 하나는 -O(C=O)-, -(C=O)O-, -C(=O)-, -O-, $-S(O)_x$, -S-S-, -C(=O)S-, -SC(=O)-, $-NR^aC(=O)-$, $-C(=O)NR^a-$, $NR^aC(=O)NR^a-$, $-OC(=O)NR^a-$, 또는 $-NR^aC(=O)O-$이고, $L^1$ 또는 $L^2$ 중 다른 하나는 -O(C=O)-, -(C=O)O-, -C(=O)-, -O-, $-S(O)_x$, -S-S-, -C(=O)S-, SC(=O)-, $-NR^aC(=O)-$, $-C(=O)NR^a-$, ,$NR^aC(=O)NR^a-$, $-OC(=O)NR^a-$ 또는 $-NR^aC(=O)O-$이거나 직접 결합이고, $G^1$ 및 $G^2$는 각각 독립적으로 치환되지 않은 $C_1$-$C_{12}$ 알킬렌 또는 $C_1$-$C_{12}$ 알케닐렌이고, $G^3$는 $C_1$-$C_{24}$ 알킬렌, $C_1$-$C_{24}$ 알케닐렌, $C_3$-$C_8$ 시클로알킬렌, $C_3$-$C_8$ 시클로알케닐렌이고, $R^a$는 H 또는 $C_1$-$C_{12}$ 알킬이고, $R^1$ 및 $R^2$는 각각 독립적으로 $C_6$-$C_{24}$ 알킬 또는 $C_6$-$C_{24}$ 알케닐이고, $R^3$은 H, $OR^5$, CN, $-C(=O)OR^4$, $-OC(=O)R^4$ 또는 $-NR^5C(=O)R^4$이고, $R^4$는 $C_1$-$C_{12}$ 알킬이고, $R^5$는 H 또는 $C_1$-$C_6$ 알킬이고, x는 0, 1 또는 2이다.

[0540] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2017/117528에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



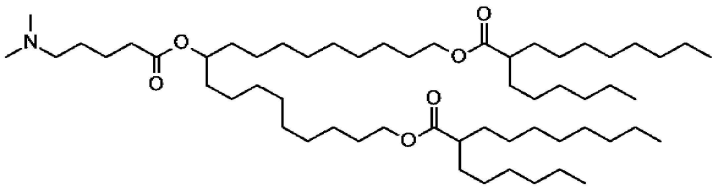
[0541]

[0542] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



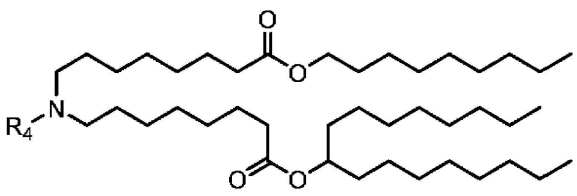
[0543]

[0544] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:

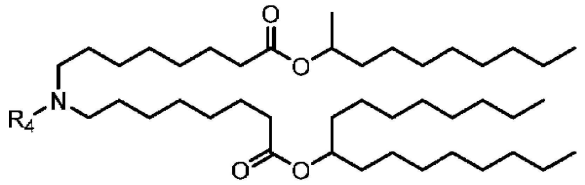


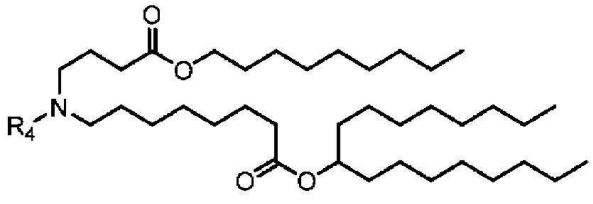
[0545]

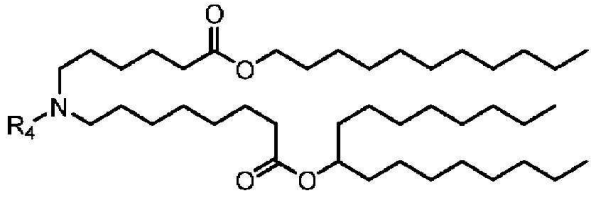
[0546] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2017/049245에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법의 양이온성 지질은 다음의 식 중 하나의 화합물 및 이의 약학적으로 허용 가능한 염을 포함한다:

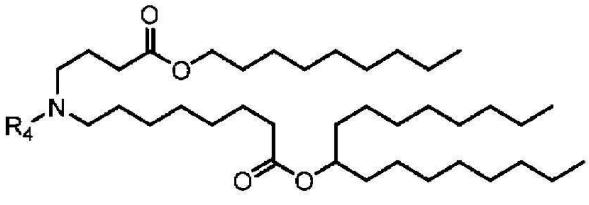


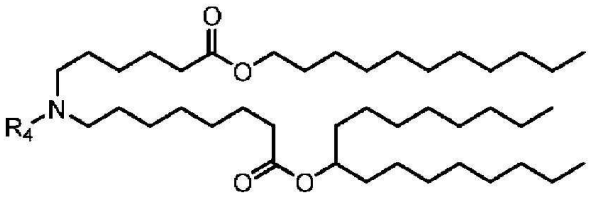
[0547] ,

[0548] ,

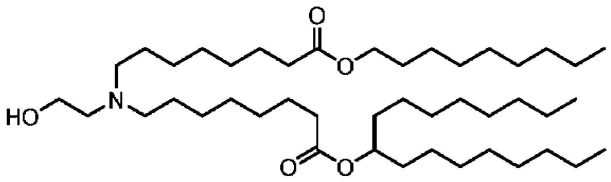
(IIc),

[0549] (IIe), 및

(IIc),

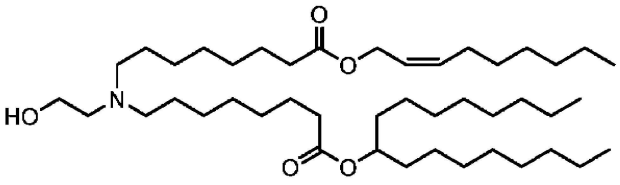
[0550] (IIe),

[0551] 이들 네 개의 화학식 중 어느 하나에 있어서, $R_4$는 $-(CH_2)_nQ$ 및 $-(CH_2)_nCHQR$에서 독립적으로 선택되고, Q는 -OR, -OH, $-O(CH_2)_nN(R)_2$, -OC(O)R, $-CX_3$, -CN, -N(R)C(O)R, -N(H)C(O)R, $-N(R)S(O)_2R$, $-N(H)S(O)_2R$, $-N(R)C(O)N(R)_2$, $-N(H)C(O)N(R)_2$, -N(H)C(O)N(H)(R), $-N(R)C(S)N(R)_2$, $-N(H)C(S)N(R)_2$, -N(H)C(S)N(H)(R) 및 헤테로고리로 이루어진 군에서 선택되며, n은 1, 2, 또는 3이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



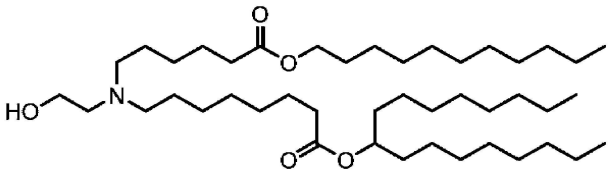
[0552]

[0553] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



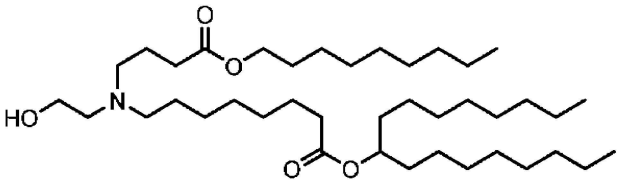
[0554]

[0555] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



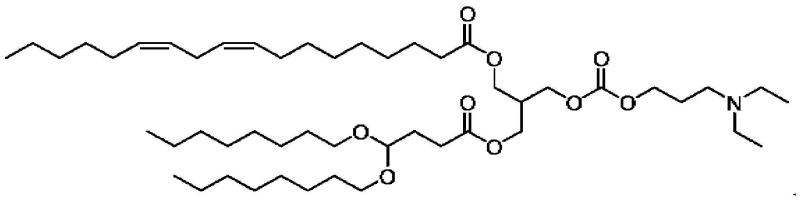
[0556]

[0557] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



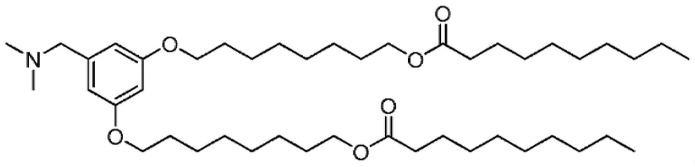
[0558]

[0559] 본 발명의 약학적 조성물 및 방법에서 사용하기 적합한 다른 양이온성 지질은 둘 모두 본원에 참조로서 포함되는, 국제 특허 공개 WO 2017/173054 및 WO 2015/095340에 기술된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



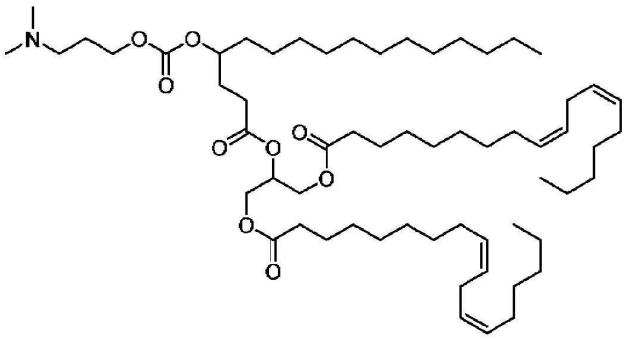
[0560]

[0561] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



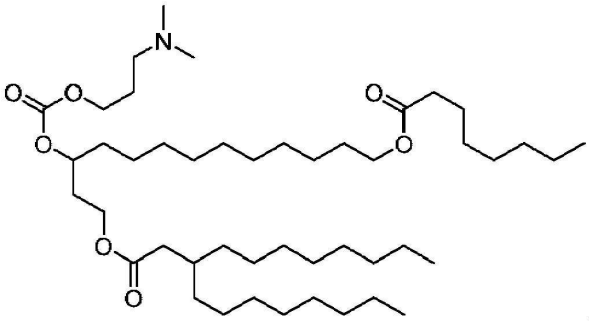
[0562]

[0563] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



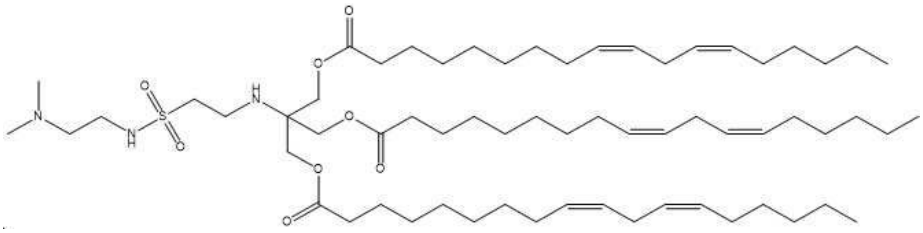
[0564]

[0565] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



[0566]

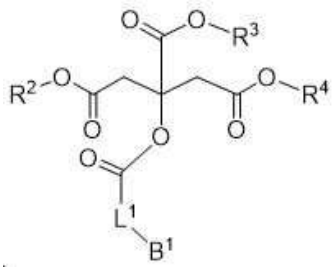
[0567] 본 발명의 약학적 조성물 및 방법에 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 통합되는, 2019년 6월 24일자로 출원된 미국 특허 가출원 제62/865,555호에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



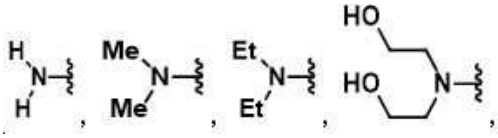
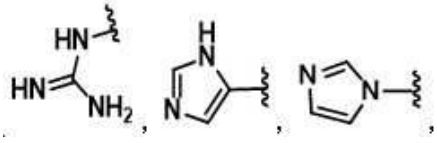
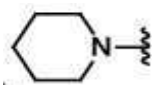
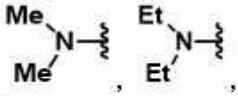
[0568]

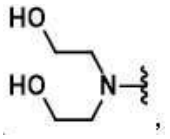
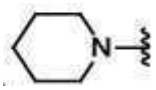
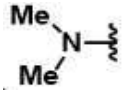
[0569] (GL-TES-SA-DME-E18-2)

[0570] 본 발명의 약학적 조성물 및 방법에 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 통합되는, 2019년 6월 21일자로 출원된 미국 특허 가출원 제62/864,818호에 기재된 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화학식에 따른 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함하며:

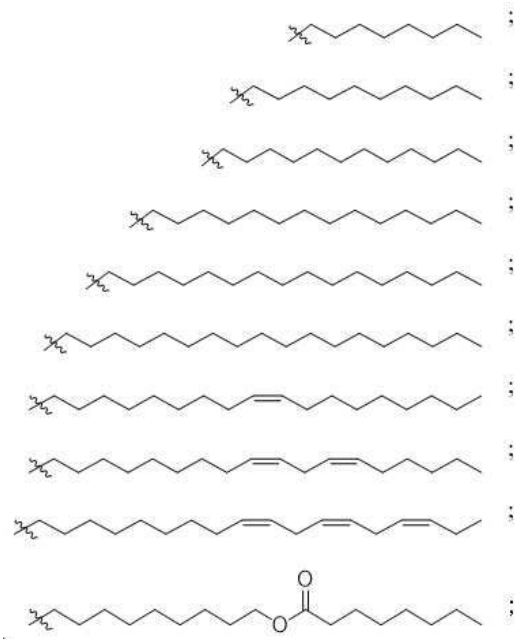


[0571] ,

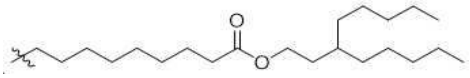
[0572] 식 중, 각각의 $R^2$, $R^3$, 및 $R^4$는 독립적으로 $C_6$-$C_{30}$ 알킬, $C_6$-$C_{30}$ 알케닐, 또는 $C_6$-$C_{30}$ 알키닐이고; $L^1$은 $C_1$-$C_{30}$ 알킬렌; $C_2$-$C_{30}$ 알케닐렌; 또는 $C_2$-$C_{30}$ 알키닐렌이고, $B^1$은 이온화 가능한 질소 함유 기이다. 구현예에서, $L^1$은 $C_1$-$C_{10}$ 알킬렌이다. 구현예에서, $L^1$은 치환되지 않은 $C_1$-$C_{10}$ 알킬렌이다. 구현예에서, $L^1$은 $(CH_2)_2$, $(CH_2)_3$, $(CH_2)_4$, 또는 $(CH_2)_5$이다. 구현예에서, $L^1$은 $(CH_2)$, $(CH_2)_6$, $(CH_2)_7$, $(CH_2)_8$, $(CH_2)_9$, 또는 $(CH_2)_{10}$이다. 구현예에서, $B^1$은 독립적으로 $NH_2$, 구아니딘, 아미딘, 모노- 또는 디알킬아민, 5-원 내지 6-원 질소 함유 헤테로시클로알킬, 또는 5-원 내지 6-원 질소 함유 헤테로아릴이다. 구현예에서, $B^1$은 , , 또는 이다. 구현예에서, $B^1$은 ,


, 또는 
이다. 구현예에서, $B^1$은 
이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$은 독립적으로, 치환되지 않은 선형 $C_6$-$C_{22}$ 알킬, 치환되지 않은 선형 $C_6$-$C_{22}$ 알케닐, 치환되지 않은 선형 $C_6$-$C_{22}$ 알키닐, 치환되지 않은 분지형 $C_6$-$C_{22}$ 알킬, 치환되지 않은 분지형 $C_6$-$C_{22}$ 알케닐, 또는 치환되지 않은 분지형 $C_6$-$C_{22}$ 알키닐이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$는 치환되지 않은 $C_6$-$C_{22}$ 알킬이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$는 $-C_6H_{13}$, $-C_7H_{15}$, $-C_8H_{17}$, $-C_9H_{19}$, $-C_{10}H_{21}$, $-C_{11}H_{23}$, $-C_{12}H_{25}$, $-C_{13}H_{27}$, $-C_{14}H_{29}$, $-C_{15}H_{31}$, $-C_{16}H_{33}$, $-C_{17}H_{35}$, $-C_{18}H_{37}$, $-C_{19}H_{39}$, $-C_{20}H_{41}$, $-C_{21}H_{43}$, $-C_{22}H_{45}$, $-C_{23}H_{47}$, $-C_{24}H_{49}$, 또는 $-C_{25}H_{51}$이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$는 독립적으로, $-O(CO)R^5$ 또는 $-C(O)OR^5$로 치환된 $C_6$-$C_{12}$ 알킬이고, 여기에서 $R^5$는 치환되지 않은 $C_6$-$C_{14}$ 알킬이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$는 치환되지 않은 $C_6$-$C_{22}$ 알케닐이다. 구현예에서, 각각의 $R^2$, $R^3$, 및 $R^4$는 $-(CH_2)_4CH=CH_2$, $-(CH_2)_5CH=CH_2$, $-(CH_2)_6CH=CH_2$, $-(CH_2)_7CH=CH_2$, $-(CH_2)_8CH=CH_2$, $-(CH_2)_9CH=CH_2$, $-(CH_2)_{10}CH=CH_2$, $-(CH_2)_{11}CH=CH_2$, $-(CH_2)_{12}CH=CH_2$, $-(CH_2)_{13}CH=CH_2$, $-(CH_2)_{14}CH=CH_2$, $-(CH_2)_{15}CH=CH_2$, $-(CH_2)_{16}CH=CH_2$, $-(CH_2)_{17}CH=CH_2$, $-(CH_2)_{18}CH=CH_2$, $-(CH_2)_7CH=CH(CH_2)_3CH_3$, $-(CH_2)_7CH=CH(CH_2)_5CH_3$, $-(CH_2)_4CH=CH(CH_2)_8CH_3$, $-(CH_2)_7CH=CH(CH_2)_7CH_3$, $-(CH_2)_6CH=CHCH_2CH=CH(CH_2)_4CH_3$, $-(CH_2)_7CH=CHCH_2CH=CH(CH_2)_4CH_3$, $-(CH_2)_7CH=CHCH_2CH=CHCH_2CH=CHCH_2CH_3$, $-(CH_2)_3CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CH(CH_2)_4CH_3$, $-(CH_2)_3CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH_3$, $-(CH_2)_{11}CH=CH(CH_2)_7CH_3$, 또는 $-(CH_2)_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH=CHCH_2CH_3$이다.

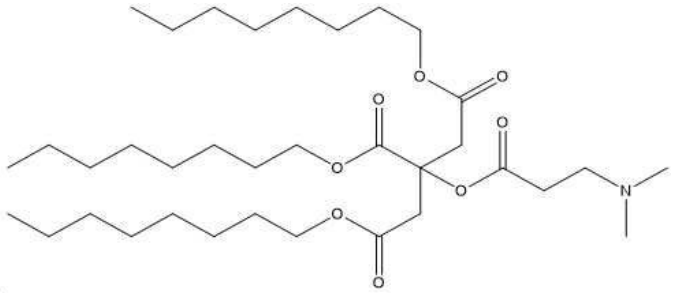
[0573] 구현예에서, 전술한 $C_6$-$C_{22}$ 알케닐은 모노알케닐, 디에닐, 또는 트리에닐이다. 구현예에서 각각의 $R^2$, $R^3$ 및 $R^4$는,


;

[0574] 
또는

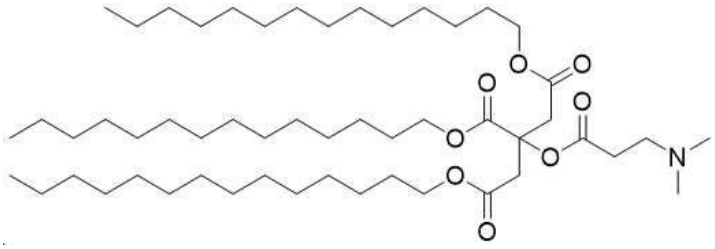
[0575] 
이다.

[0576] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



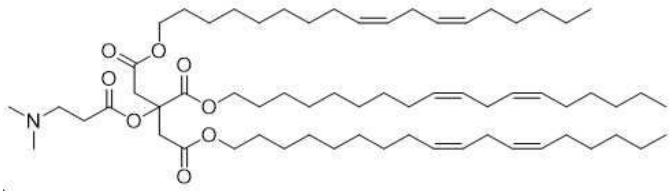
[0577] (TL1-01D-DMA)

[0578] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



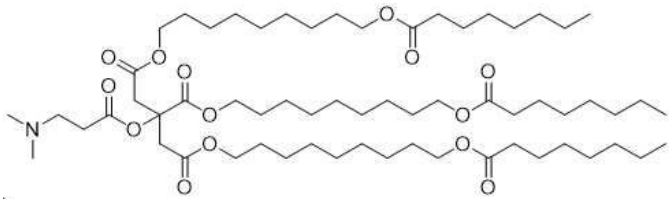
[0579] (TL1-04D-DMA)

[0580] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



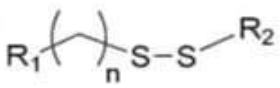
[0581] (TL1-08D-DMA)

[0582] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함한다:



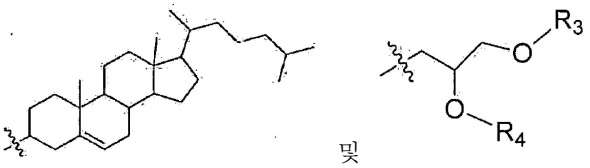
[0583] (TL1-10D-DMA)

[0584] 본 발명의 약학적 조성물 및 방법에 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2012/170889에 기재된 절단 가능한 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음 화학식의 양이온성 지질 또는 이의 약학적으로 허용 가능한 염을 포함하며:



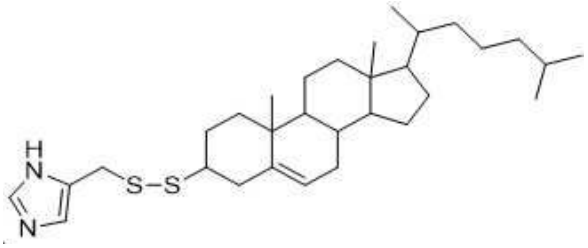
[0585] ,

[0586] 식 중, $R_1$은 이미다졸, 구아니디늄, 아미노, 이민, 엔아민, 선택적으로 치환된 알킬 아미노(예를 들어, 디메틸아미노와 같은 알킬 아미노) 및 피리딜로 이루어진 군으로부터 선택되고, $R_2$는 다음의 두 화학식 중 하나로 이루어진 군으로부터 선택되며:



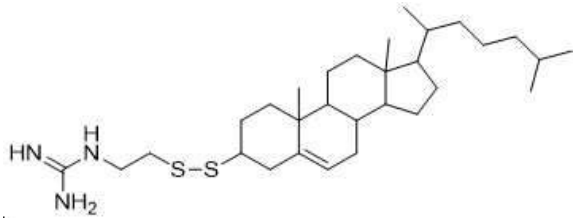
[0587] 및

[0588] 식 중 $R_3$ 및 $R_4$는 선택적으로 치환된 가변 포화 또는 불포화 $C_6$-$C_{20}$ 알킬, 및 선택적으로 치환된 가변 포화 또는 불포화 $C_6$-$C_{20}$ 아실로 이루어진 군으로부터 각각 독립적으로 선택되고; n은 0 또는 임의의 양의 정수(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 그 이상)이다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "HGT4001" 및 이의 약학적으로 허용 가능한 염을 포함한다:



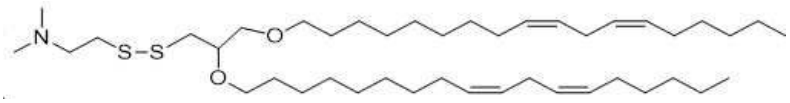

[0589] (HGT4001).

[0590] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "HGT4002" 및 이의 약학적으로 허용 가능한 염을 포함한다:



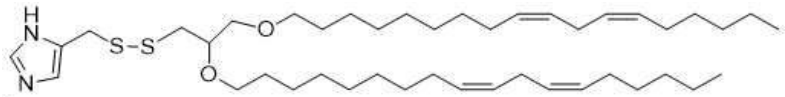

[0591] (HGT4002).

[0592] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "HGT4003" 및 이의 약학적으로 허용 가능한 염을 포함한다:



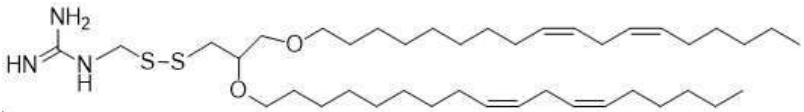

[0593] (HGT4003).

[0594] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "HGT4004" 및 이의 약학적으로 허용 가능한 염을 포함한다:



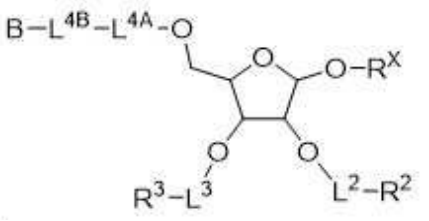

[0595] (HGT4004).

[0596] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 "HGT4005" 및 이의 약학적으로 허용 가능한 염을 포함한다:




[0597] (HGT4005).

[0598] 본 발명의 약학적 조성물 및 방법에 사용하기 적합한 다른 양이온성 지질은 본원에 참조로서 포함되는, 국제 특허 공개 WO 2019/222424에 기재된 절단 가능한 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 국제 특허 공개 WO 2019/222424에 기재된 일반 화학식 중 어느 하나 또는 구조식 (1a) 내지 (21a) 및 (1b) 내지 (21b) 및 (22)-(237) 중 어느 하나에 해당하는 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 식 (I')에 따른 구조를 갖는 양이온성 지질 및 이의 약학적으로 허용 가능한 염을 포함하며:

[0599]  (I'),

[0600] 식 중,

[0601] $R^X$는 독립적으로 -H, $-L^1-R^1$, 또는 $-L^{5A}-L^{5B}-B'$이고;

[0602] 각각의 $L^1$, $L^2$, 및 $L^3$은 독립적으로 공유 결합, -C(O)-, -C(O)O-, -C(O)S-, 또는 $-C(O)NR^L-$이고;

[0603] 각각의 $L^{4A}$ 및 $L^{5A}$는 독립적으로 -C(O)-, -C(O)O-, 또는 $-C(O)NR^L-$이고;

[0604] 각각의 $L^{4B}$ 및 $L^{5B}$는 독립적으로 $C_1$-$C_{20}$ 알킬렌, $C_2$-$C_{20}$ 알케닐렌, 또는 $C_2$-$C_{20}$ 알키닐렌이고;

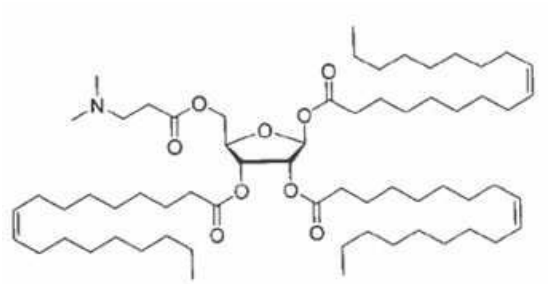
[0605] 각각의 B 및 B'은 $NR^4R^5$ 또는 5-원 내지 10-원 질소 함유 헤테로아릴이고;

[0606] 각각의 $R^1$, $R^2$, 및 $R^3$은 독립적으로 $C_6$-$C_{30}$ 알킬, $C_6$-$C_{30}$ 알케닐, 또는 $C_6$-$C_{30}$ 알키닐이고;

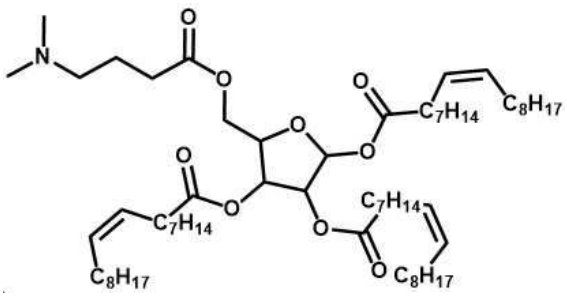
[0607] 각각의 $R^4$, 및 $R^5$는 독립적으로 $C_1$-$C_{10}$ 알킬, $C_2$-$C_{10}$ 알케닐, 또는 $C_2$-$C_{10}$ 알키닐이며;

[0608] 각각의 $R^L$은 독립적으로 수소, $C_1$-$C_{20}$ 알킬, $C_2$-$C_{20}$ 알케닐, 또는 $C_2$-$C_{20}$ 알키닐이다.

[0609] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 다음의 화합물 구조를 갖는 양이온성 지질, 즉 국제 출원 PCT/US2019/032522의 화합물 (139)를 포함한다:

[0610]  ("18:1 탄소 꼬리-리보스 지질").

[0611] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 다음의 화합물 구조를 갖는 RL3-DMA-07D 및 이의 약학적으로 허용 가능한 염을 포함한다:

[0612]  (RL3-DMA-07D)

[0613] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법은 양이온성 지질, 즉 N-[1-(2,3-디올레일옥시)프로필]-N,N,N-트리메틸암모늄 염화물("DOTMA")를 포함한다. (Feigner 등 (Proc. Nat'l Acad. Sci. 84, 7413 (1987); 미국 특허 제4,897,355호, 이는 본원에 참조로서 통합됨). 본 발명의 약학적 조성물 및 방법에 적합한 다른 양이온성 지질은 예를 들어, 5-카르복시스페르밀글리신디옥타데실아미드("DOGS"); 2,3-디올레일옥시-N-[2(스페르민-카르복스아미도에틸]-N,N-디메틸-1-프로판아미늄("DOSPA")(Behr 등, Proc. Nat.'l Acad. Sci. 86, 6982

(1989), 미국 특허 제5,171,678호; 미국 특허 제5,334,761호); 1,2-디올레오일-3-디메틸암모늄-프로판("DODAP"); 1,2-디올레오일-3-트리메틸암모늄-프로판("DOTAP")을 포함한다.

[0614] 본 발명의 약학적 조성물 및 방법에 적합한 양이온성 지질의 추가 예시는 또한 다음을 포함한다: 1,2-디스테아릴옥시-N,N-디메틸-3-아미노프로판("DSDMA"); 1,2-디올레일옥시-N,N-디메틸-3-아미노프로판("DODMA"); 1,2-디리놀레일옥시-N,N-디메틸-3-아미노프로판("DLinDMA"); 1,2-디리놀레닐옥시-N,N-디메틸-3-아미노프로판("DLenDMA"); N-디올레일-N,N-디메틸암모늄 염화물("DODAC"); N,N-디스테아릴-N,N-디메틸암모늄 브로마이드("DDAB"); N-(1,2-디미리스틸옥시프로프-3-일)-N,N-디메틸-N-히드록시에틸 암모늄 브로마이드("DMRIE"); 3-디메틸아미노-2-(콜레스트-5-엔-3-베타-옥시부탄-4-옥시)-1-(시스,시스-9,12-옥타데카디엔옥시)프로판("CLinDMA"); 2-[5'-(콜레스트-5-엔-3-베타-옥시)-3'-옥사펜톡시)-3-디메틸 1-1-(시스,시스-9', 1-2'-옥타데칸디엔옥시)프로판("CpLinDMA"); N,N-디메틸-3,4-디올레일옥시벤질아민("DMOBA"); 1,2-N,N'-디올레일카르바밀-3-디메틸아미노프로판("DOcarbDAP"); 2,3-디리놀레오일옥시-N,N-디메틸프로필아민("DLinDAP"); 1,2-N,N'-디리놀레일카르바밀-3-디메틸아미노프로판("DLincarbDAP"); 1,2-디리놀레오일카르바밀-3-디메틸아미노프로판("DLinCDAP"); 2,2-디리놀레일-4-디메틸아미노메틸-[1,3]-디옥솔란("DLin-K-DMA"); 2-((8-[(3P)-콜레스트-5-엔-3-일옥시]옥틸)옥시)-N,N-디메틸1-3-[(9Z, 12Z)-옥타데카-9,12-디엔-1-일옥시]프로판-1-아민("옥틸-CLinDMA"); (2R)-2-((8-[(3베타)-콜레스트-5-엔-3-일옥시]옥틸)옥시)-N, N-디메틸1-3-[(9Z, 12Z)-옥타데카-9,12-디엔-1-일옥시]프로판-1-아민("옥틸-CLinDMA (2R)"); (2S)-2-((8-[(3P)-콜레스트-5-엔-3-일옥시]옥틸)옥시)-N, fsl-dimethyh3-[(9Z, 12Z)-옥타데카-9, 12-디엔-1-일옥시]프로판-1-아민("Octyl-CLinDMA (2S)"); 2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란("DLin-K-XTC2-DMA"); 및 2-(2,2-디((9Z,12Z)-옥타데카-9,1 2-디엔-1-일)-1 ,3-디옥솔란-4-일)-N,N-디메틸에탄아민("DLin-KC2-DMA")(본원에 참조로서 포함되는 WO 2010/042877; Semple 등, Nature Biotech. 28: 172-176 (2010) 참조). (Heyes, J. 등, J Controlled Release 107: 276-287 (2005); Morrissey, DV. 등, Nat. Biotechnol. 23(8): 1003-1007 (2005); 국제 특허 공개 WO 2005/121348). 일부 구현예에서, 양이온성 지질 중 하나 이상은 이미다졸, 디알킬아미노, 또는 구아니디늄 모이어티 중 적어도 하나를 포함한다.

[0615] 일부 구현예에서, 본 발명의 약학적 조성물 및 방법에 적합한 하나 이상의 양이온성 지질은 2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란("XTC"); (3aR,5s,6aS)-N,N-디메틸-2,2-디((9Z,12Z)-옥타데카-9,12-디에틸)테트라히드로-3aH-시클로펜타[d] [1 ,3]디옥솔-5-아민("ALNY-100") 및/또는 4,7,13-트리스(3-옥소-3-(운데실아미노)프로필)-N1,N16-디운데실-4,7,10,13-테트라아자헥사데칸-1,16-디아미드("NC98-5")를 포함한다.

[0616] 일부 구현예에서, 본 발명의 약학적 조성물은, 약학적 조성물 중 총 지질 함량, 예컨대 지질 나노입자의 적어도 약 5 중량%, 10 중량%, 20 중량%, 30 중량%, 35 중량%, 40 중량%, 45 중량%, 50 중량%, 55 중량%, 60 중량%, 65 중량%, 또는 70 중량%를 구성하는 하나 이상의 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물은 약학적 조성물(예를 들어, 지질 나노입자) 중 총 지질 함량의 적어도 약 5 mol%, 10 mol%, 20 mol%, 30 mol%, 35 mol%, 40 mol%, 45 mol%, 50 mol%, 55 mol%, 60 mol%, 65 mol%, 또는 70 mol%를 구성하는 하나 이상의 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물은 약학적 조성물(예를 들어, 지질 나노입자) 중 총 지질 함량의 약 30 내지 70 중량%(예컨대, 약 30 내지 65 중량%, 약 30 내지 60 중량%, 약 30 내지 55 중량%, 약 30 내지 50 중량%, 약 30 내지 45 중량%, 약 30 내지 40 중량%, 약 35 내지 50 중량%, 약 35 내지 45 중량%, 또는 약 35 내지 40 중량%)를 구성하는 하나 이상의 양이온성 지질을 포함한다. 일부 구현예에서, 본 발명의 약학적 조성물은 약학적 조성물(예를 들어, 지질 나노입자) 중 총 지질 함량의 약 30 내지 70 mol%(예컨대, 약 30 내지 65 mol%, 약 30 내지 60 mol%, 약 30 내지 55 mol%, 약 30 내지 50 mol%, 약 30 내지 45 mol%, 약 30 내지 40 mol%, 약 35 내지 50 mol%, 약 35 내지 45 mol%, 또는 약 35 내지 40 mol%)를 구성하는 하나 이상의 양이온성 지질을 포함한다.

[0617] *비-양이온성 지질*

[0618] 일부 구현예에서, 지질 나노입자는 하나 이상의 비-양이온성 지질을 함유한다. 본원에서 사용되는 용어 "비-양이온성 지질"은 임의의 중성, 쌍성이온성(zwitterionic), 또는 음이온성 지질을 지칭한다. 본원에서 사용되는 용어 "음이온성 지질"은 선택된 pH, 예컨대 생리적 pH에서 순 음전하를 보유하는 다수의 지질 종 중 어느 하나를 지칭한다. 비-양이온성 지질은 디스테아로일포스파티딜콜린(DSPC), 디올레오일포스파티딜콜린(DOPC), 디팔미토일포스파티딜콜린(DPPC), 디올레오일포스파티딜글리세롤(DOPG), 디팔미토일포스파티딜글리세롤(DPPG), 디올레오일포스파티딜에탄올아민(DOPE), 팔미토일올레오일포스파티딜콜린(POPC), 팔미토일올레오일-포스파티딜에탄올아민(POPE), 디올레오일-포스파티딜에탄올아민 4-(N-말레이미도메틸)-시클로헥산-1-카르복실레이트(DOPE-mal), 디팔미토일 포스파티딜 에탄올아민(DPPE), 디미리스토일포스포에탄올아민(DMPE), 디스테아로일-포스파티딜-에탄올아민(DSPE), 1,2-디에루코일-sn-글리세로-3-포스포에탄올아민(DEPE), 포스파티딜세린, 스핑고지질, 세레브로

시드, 강글리오시드, 16-O-모노메틸 PE, 16-O-디메틸 PE, 18-1-트랜스 PE, 1-스테아로일-2-올레오일-포스파티디에탄올아민, 또는 이의 혼합물을 포함하지만 이에 한정되지는 않는다. 일부 구현예에서, 본 발명과 함께 사용하기에 적합한 지질 나노입자는 비-양이온성 지질 성분으로서 DOPE를 포함한다. 다른 구현예에서, 본 발명과 함께 사용하기에 적합한 지질 나노입자는 비-양이온성 지질 성분으로서 DEPE를 포함한다.

[0619] 일부 구현예에서, 비-양이온성 지질은 중성 지질, 즉 본 약학적 조성물이 제형화되고/되거나 투여되는 조건하에서 순 전하를 띠지 않는 지질이다.

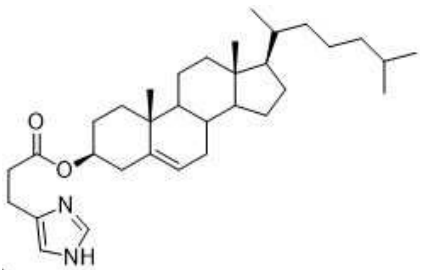
[0620] 일부 구현예에서, 이와 같은 비-양이온성 지질은 단독으로 사용될 수 있지만, 바람직하게는 다른 지질 예를 들어, 양이온성 지질과 조합하여 사용된다.

[0621] 일부 구현예에서, 비-양이온성 지질은 약학적 조성물에 존재하는 총 지질의 약 5% 내지 약 90%, 약 5% 내지 약 70%, 약 5% 내지 약 50%, 약 5% 내지 약 40%, 약 5% 내지 약 30%, 약 10% 내지 약 70%, 약 10% 내지 약 50%, 또는 약 10% 내지 약 40%의 몰 비율(mol%)로 존재할 수 있다. 일부 구현예에서, 총 비-양이온성 지질은 약학적 조성물에 존재하는 총 지질의 약 5% 내지 약 90%, 약 5% 내지 약 70%, 약 5% 내지 약 50%, 약 5% 내지 약 40%, 약 5% 내지 약 30%, 약 10% 내지 약 70%, 약 10% 내지 약 50%, 또는 약 10% 내지 약 40%의 몰 비율(mol%)로 존재할 수 있다. 일부 구현예에서, 지질 나노입자 내의 비-양이온성 지질의 백분율은 약 5 mol% 초과, 10 mol% 초과, 20 mol% 초과, 30 mol% 초과, 또는 40 mol% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 내의 총 비-양이온성 지질의 백분율은 약 5 mol% 초과, 10 mol% 초과, 20 mol% 초과, 30 mol% 초과, 또는 40 mol% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 내의 비-양이온성 지질의 백분율은 약 5 mol% 이하, 10 mol% 이하, 20 mol% 이하, 30 mol% 이하, 또는 40 mol% 이하이다. 일부 구현예에서, 지질 나노입자 내의 총 비-양이온성 지질의 백분율은 약 5 mol% 이하, 약 10 mol% 이하, 약 20 mol% 이하, 약 30 mol% 이하, 또는 약 40 mol% 이하일 수 있다.

[0622] 일부 구현예에서, 비-양이온성 지질은 약학적 조성물에 존재하는 총 지질의 약 5% 내지 약 90%, 약 5% 내지 약 70%, 약 5% 내지 약 50%, 약 5% 내지 약 40%, 약 5% 내지 약 30%, 약 10% 내지 약 70%, 약 10% 내지 약 50%, 또는 약 10% 내지 약 40%의 중량비(wt%)로 존재할 수 있다. 일부 구현예에서, 총 비-양이온성 지질은 약학적 조성물에 존재하는 총 지질의 약 5% 내지 약 90%, 약 5% 내지 약 70%, 약 5% 내지 약 50%, 약 5% 내지 약 40%, 약 5% 내지 약 30%, 약 10% 내지 약 70%, 약 10% 내지 약 50%, 또는 약 10% 내지 약 40%의 중량비(wt%)로 존재할 수 있다. 일부 구현예에서, 지질 나노입자 내의 비-양이온성 지질의 백분율은 약 5 wt% 초과, 약 10 wt% 초과, 약 20 wt% 초과, 약 30 wt% 초과, 또는 약 40 wt% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 내의 총 비-양이온성 지질의 백분율은 약 5 wt% 초과, 약 10 wt% 초과, 약 20 wt% 초과, 약 30 wt% 초과, 또는 약 40 wt% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 내의 비-양이온성 지질의 백분율은 약 5 wt% 이하, 약 10 wt% 이하, 약 20 wt% 이하, 약 30 wt% 이하, 또는 약 40 wt% 이하이다. 일부 구현예에서, 지질 나노입자 내의 총 비-양이온성 지질의 백분율은 약 5 wt% 이하, 약 10 wt% 이하, 약 20 wt% 이하, 약 30 wt% 이하, 또는 약 40 wt% 이하일 수 있다.

[0623] *콜레스테롤계 지질*

[0624] 일부 구현예에서, 지질 나노입자는 하나 이상의 콜레스테롤계 지질을 포함한다. 예를 들어, 적합한 콜레스테롤계 양이온성 지질은, 예를 들어, DC-Choi (N,N-디메틸-N-에틸카르복스아미도콜레스테롤), 1,4-비스(3-N-올레일아미노-프로필)피페라진(Gao 등, Biochem. Biophys. Res. Comm. 179, 280 (1991); Wolf 등, BioTechniques 23, 139 (1997)]; 미국 특허 제5,744,335호), 또는 국제 특허 공개 WO 2011/068810에 개시된 바와 같은, 다음의 구조를 갖는 이미다졸 콜레스테롤 에스테르(ICE)를 포함한다:




[0625] ("ICE").

[0626] 구현예에서, 콜레스테롤계 지질은 콜레스테롤이다.

[0627] 일부 구현예에서, 콜레스테롤계 지질은 지질 나노입자에 존재하는 총 지질의 약 1% 내지 약 30%, 또는 약 5% 내

지 약 20%의 몰비(mol%)를 포함할 수 있다. 일부 구현예에서, 지질 나노입자 중 콜레스테롤계 지질의 백분율은 약 5% 초과, 약 10% 초과, 약 20% 초과, 약 30% 초과, 또는 약 40% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 중 콜레스테롤계 지질의 백분율은 약 5 mol% 이하, 약 10 mol% 이하, 약 20 mol% 이하, 약 30 mol% 이하, 또는 약 40 mol% 이하일 수 있다.

[0628] 일부 구현예에서, 콜레스테롤계 지질은 지질 나노입자에 존재하는 총 지질의 약 1% 내지 약 30%, 또는 약 5% 내지 약 20%의 중량비(wt%)로 존재할 수 있다. 일부 구현예에서, 지질 나노입자 중 콜레스테롤계 지질의 백분율은 약 5 wt% 초과, 약 10 wt% 초과, 약 20 wt% 초과, 약 30 wt% 초과, 또는 약 40 wt% 초과일 수 있다. 일부 구현예에서, 지질 나노입자 중 콜레스테롤계 지질의 백분율은 약 5 wt% 이하, 약 10 wt% 이하, 약 20 wt% 이하, 약 30 wt% 이하, 또는 약 40 wt% 이하일 수 있다.

[0629] *PEG-변형 지질*

[0630] 일부 구현예에서, 지질 나노입자는 하나 이상의 PEG화 지질을 포함한다.

[0631] 예를 들어, 폴리에틸렌 글리콜(PEG)-변형 인지질, 및 N-옥타노일-스핑고신-1-[숙시닐(메톡시 폴리에틸렌 글리콜)-2000](C8 PEG-2000 세라미드)를 포함하는 유도체화된 세라미드(PEG-CER)와 같은 유도체화된 지질을 단독으로 사용하거나 바람직하게는 수송 비히클(예를 들어, 지질 나노입자)을 포함하는 다른 지질 약학적 조성물과 함께 조합하여 사용하는 것 또한 본 발명에서 고려된다.

[0632] 고려된 PEG-변형 지질은 $C_6$-$C_{20}$ 길이의 알킬 사슬(들)을 갖는 지질에 공유 부착된 최대 5 kDa 길이의 폴리에틸렌 글리콜 사슬을 포함하지만 이에 한정되지는 않는다. 일부 구현예에서, PEG-변형 또는 PEG화 지질은 PGE화 콜레스테롤 또는 PEG-2K이다. 이와 같은 성분을 첨가함으로써 복합체 응집을 예방할 수 있고, 또한 순환 수명을 늘리고 표적 조직에 대한 지질-핵산 조성물의 전달을 증가시키는 수단을 제공할 수 있거나(Klibanov 등 (1990) FEBS Letters, 268 (1): 235-237), 이들 성분은 생체 내에서 약학적 조성물로부터 신속하게 교환되도록 선택될 수 있다(미국 특허 제5,885,613호 참조). 특히 유용한 교환 가능한 지질은 더 짧은 아실 사슬(예컨대, $C_{14}$ 또는 $C_{18}$)을 갖는 PEG-세라미드이다. 본 발명과 함께 사용하기에 적합한 지질 나노입자는 1,2-디미리스톨-라세-글리세로-3-메톡시폴리에틸렌 글리콜-2000(DMG-PEG2K)와 같은 PEG-변형 지질을 일반적으로 포함한다.

[0633] 본 발명의 PEG-변형 인지질 및 유도된 지질은 리포솜 수송 비히클(예컨대, 본원에 개시된 지질 나노입자)에 존재하는 총 지질의 약 0% 내지 약 20%, 약 0.5% 내지 약 20%, 약 1% 내지 약 15%, 약 4% 내지 약 10%, 또는 약 2%의 몰비를 포함할 수 있다. 일부 구현예에서, 하나 이상의 PEG-변형 지질은 몰비 기준으로 총 지질의 약 4%를 구성한다. 일부 구현예에서, 하나 이상의 PEG-변형 지질은 몰비 기준으로 총 지질의 약 5%를 구성한다. 일부 구현예에서, 하나 이상의 PEG-변형 지질은 몰비 기준으로 총 지질의 약 6%를 구성한다. 폐 전달과 같은 특정 응용예의 경우, PEG-변형 지질 성분이 몰비 기준으로 총 지질의 약 5%를 구성하는 지질 나노입자가 특히 적합한 것으로 확인되었다.

[0634] *구별되는 지질 성분의 비*

[0635] 본 발명에 적합한 지질 나노입자는 본원에 기술된 양이온성 지질, 비-양이온성 지질, 콜레스테롤 지질, PEG-변형 지질, 양친매성 블록 공중합체, 및/또는 중합체 중 하나 이상을 다양한 비율로 포함할 수 있다. 일부 구현예에서, 지질 나노입자는 나노입자의 5개 이하의 구별되는 성분을 포함한다. 일부 구현예에서, 지질 나노입자는 나노입자의 4개 이하의 구별되는 성분을 포함한다. 일부 구현예에서, 지질 나노입자는 나노입자의 3개 이하의 구별되는 성분을 포함한다. 비제한적인 예로서, 적절한 지질 나노입자 약학적 조성물은, cKK-E12, DOPE, 콜레스테롤 및 DMG-PEG2K; C12-200, DOPE, 콜레스테롤 및 DMG-PEG2K; HGT4003, DOPE, 콜레스테롤 및 DMG-PEG2K; ICE, DOPE, 콜레스테롤 및 DMG-PEG2K; HGT4001, DOPE, 콜레스테롤 및 DMG-PEG2K; HGT4002, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-01D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-04D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-08D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-10D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; ICE, DOPE and DMG-PEG2K; HGT4001, DOPE 및 DMG-PEG2K; 또는 HGT4002, DOPE 및 DMG-PEG2K로부터 선택되는 조합을 포함할 수 있다.

[0636] 다양한 구현예에서, 양이온성 지질(예컨대, cKK-E12, C12-200, TL1-01D-DMA, TL1-04D-DMA, TL1-08D-DMA, TL1-10D-DMA, ICE, HGT4001, HGT4002 및/또는 HGT4003)은 몰비로 지질 나노입자의 약 30 내지 60%(예컨대, 약 30 내지 55%, 약 30 내지 50%, 약 30 내지 45%, 약 30 내지 40%, 약 35 내지 50%, 약 35 내지 45%, 또는 약 35 내지 40%)를 구성한다. 일부 구현예에서, 양이온성 지질(예컨대, cKK-E12, C12-200, TL1-01D-DMA, TL1-04D-DMA,

TL1-08D-DMA, TL1-10D-DMA, ICE, HGT4001, HGT4002 및/또는 HGT4003)의 백분율은 몰비로 지질 나노입자의 약 30% 이상, 약 35% 이상, 약 40 % 이상, 약 45% 이상, 약 50% 이상, 약 55% 이상, 또는 약 60% 이상이다.

[0637] 일부 구현예에서, 양이온성 지질(들) 대 비-양이온성 지질(들) 대 콜레스테롤계 지질(들) 대 PEG-변형 지질(들)의 비는 각각 약 30 내지 60:25 내지 35:20 내지 30:1 내지 15일 수 있다. 일부 구현예에서, 양이온성 지질(들) 대 비-양이온성 지질(들) 대 콜레스테롤계 지질(들) 대 PEG-변형 지질(들)의 비는 각각 대략 40:30:20:10이다. 일부 구현예에서, 양이온성 지질(들) 대 비-양이온성 지질(들) 대 콜레스테롤계 지질(들) 대 PEG-변형 지질(들)의 비는 각각 대략 40:30:25:5이다. 일부 구현예에서, 양이온성 지질(들) 대 비-양이온성 지질(들) 대 콜레스테롤계 지질(들) 대 PEG-변형 지질(들)의 비는 각각 대략 40:32:25:3이다. 일부 구현예에서, 양이온성 지질(들) 대 비-양이온성 지질(들) 대 콜레스테롤계 지질(들) 대 PEG-변형 지질(들)의 비는 대략 50:25:20:5이다.

[0638] 지질 나노입자가 3개 및 3개 미만의 구별되는 지질 성분을 포함하는 구현예에서, 총 지질 함량의 비(즉, 지질 성분(1):지질 성분(2):지질 성분(3)의 비)는 x:y:z로 표시될 수 있으며, 여기에서,

[0639] (y + z) = 100 - x이다.

[0640] 일부 구현예에서, "x", "y", 및 "z"의 각각은 3개의 구별되는 지질 성분의 몰 백분율을 나타내며, 비는 몰비이다.

[0641] 일부 구현예에서, "x", "y", 및 "z"의 각각은 3개의 구별되는 지질 성분의 중량%를 나타내며, 비는 중량비이다.

[0642] 일부 구현예에서, 변수 "x"로 표시된 지질 성분(1)은 스테롤계 양이온 지질이다.

[0643] 일부 구현예에서, 변수 "y"로 표시된 지질 성분(2)은 비-양이온성 지질이다.

[0644] 일부 구현예에서, 변수 "z"로 표시된 지질 성분(3)은 PEG 지질이다.

[0645] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 몰 백분율을 나타내는 변수 "x"는 적어도 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 95%이다.

[0646] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 몰 백분율을 나타내는 변수 "x"는 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 40%, 약 30%, 약 20%, 또는 약 10% 이하이다. 구현예에서, 변수 "x"는 약 65%, 약 60%, 약 55%, 약 50%, 또는 약 40% 이하이다.

[0647] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 몰 백분율을 나타내는 변수 "x"는: 약 50% 이상 및 약 95% 미만; 약 50% 이상 및 약 90% 미만; 약 50% 이상 및 약 85% 미만; 약 50% 이상 및 약 80% 미만; 약 50% 이상 및 약 75% 미만; 약 50% 이상 및 약 70% 미만; 약 50% 이상 및 약 65% 미만; 또는 약 50% 이상 및 약 60% 미만이다. 구현예에서, 변수 "x"는 적어도 약 50% 내지 약 70%; 적어도 약 50% 내지 약 65%; 또는 적어도 약 50% 내지 약 60%이다.

[0648] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 중량%를 나타내는 변수 "x"는 적어도 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 95%이다.

[0649] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 중량%를 나타내는 변수 "x"는 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 40%, 약 30%, 약 20%, 또는 약 10% 이하이다. 구현예에서, 변수 "x"는 약 65%, 약 60%, 약 55%, 약 50%, 또는 약 40% 이하이다.

[0650] 일부 구현예에서, 지질 성분(1)(예를 들어, 스테롤계 양이온 지질)의 중량%를 나타내는 변수 "x"는: 약 50% 이상 및 약 95% 미만; 약 50% 이상 및 약 90% 미만; 약 50% 이상 및 약 85% 미만; 약 50% 이상 및 약 80% 미만; 약 50% 이상 및 약 75% 미만; 약 50% 이상 및 약 70% 미만; 약 50% 이상 및 약 65% 미만; 또는 약 50% 이상 및 약 60% 미만이다. 구현예에서, 변수 "x"는 적어도 약 50% 내지 약 70%; 적어도 약 50% 내지 약 65%; 또는 적어도 약 50% 내지 약 60%이다.

[0651] 일부 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 몰 백분율을 나타내는 변수 "z"는 약 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 또는 25% 이하이다. 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 몰 백분율을 나타내는 변수 "z"는 약 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%이다. 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 몰 백분율을 나타내는 변수 "z"는 약 1% 내지 약 10%, 약 2% 내지 약 10%, 약 3%

내지 약 10%, 약 4% 내지 약 10%, 약 1% 내지 약 7.5%, 약 2.5% 내지 약 10%, 약 2.5% 내지 약 7.5%, 약 2.5% 내지 약 5%, 약 5% 내지 약 7.5%, 또는 약 5% 내지 약 10%이다.

[0652] 일부 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 중량%를 나타내는 변수 "z"는 약 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 또는 25% 이하이다. 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 중량%를 나타내는 변수 "z"는 약 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%이다. 구현예에서, 지질 성분(3)(예를 들어, PEG 지질)의 중량 백분율을 나타내는 변수 "z"는 약 1% 내지 약 10%, 약 2% 내지 약 10%, 약 3% 내지 약 10%, 약 4% 내지 약 10%, 약 1% 내지 약 7.5%, 약 2.5% 내지 약 10%, 약 2.5% 내지 약 7.5%, 약 2.5% 내지 약 5%, 약 5% 내지 약 7.5%, 또는 약 5% 내지 약 10%이다.

[0653] 3개 및 단 3개의 구별되는 지질 성분을 갖는 약학적 조성물의 경우, 변수 "x", "y", 및 "z"는 3개 변수의 합이 총 지질 함량의 100%가 되는 임의의 조합일 수 있다. 예를 들어, 본 발명과 함께 사용하기에 적합한 전형적인 3성분 지질 나노입자에서, 양이온성 지질 대 비-양이온성 지질 대 PEG-변형 지질의 몰비는 각각 약 55 내지 65:30 내지 40:1 내지 15일 수 있다. 일부 구현예에서, 60:35:5의 양이온성 지질(예를 들어, 스테롤계 지질) 대 비-양이온성 지질(예를 들어, DOPE 또는 DEPE) 대 PEG-변형 지질(예를 들어, DMG-PEG2K)의 몰비는 분무를 통한 지질 나노입자의 폐 전달에 특히 적합하다.

[0654] *예시적인 지질 나노입자 제형*

[0655] 본 발명에 따른 핵산의 생체 내 전달을 위한 예시적인 지질 나노입자는 양이온성 지질(예를 들어, cKK-E10), 비-양이온성 지질(예를 들어, DOPE), 콜레스테롤 및 PEG-변형 지질(예를 들어, DMG-PEG2K)을 포함한다. 특정 구현예에서, 본 발명은 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K로 이루어진 지질 성분을 40:30:28.5:1.5의 몰비로 갖는, 본 발명의 핵산의 전달을 위한 지질 나노입자를 제공한다. 실시예에 나타낸 바와 같이, 이러한 지질 나노입자 제형은 본 발명의 면역원성 조성물에 사용하기 위해, 특히 본 발명의 핵산을 포함하는 지질 나노입자의 근육내 투여에 특히 효과적인 것으로 밝혀졌다.

[0656] *적어도 2개의 핵산을 함유하는 지질 나노입자 조성물*

[0657] 일부 구현예에서, 본 발명의 상이한 최적화된 뉴클레오티드 서열을 포함하는 적어도 2개의 핵산은 동일한 지질 나노입자(예를 들어, cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 포함하는 지질 나노입자)에 캡슐화된다. 예를 들어, 본 발명의 제1 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산(예를 들어, mRNA)은 본 발명의 제2 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산(예를 들어, mRNA)과 조합될 수 있고, 동일한 지질 나노입자에 캡슐화될 수 있다.

[0658] 다른 구현예에서, 본 발명의 상이한 최적화된 뉴클레오티드 서열을 포함하는 적어도 2개의 핵산은 (일반적으로, 동일한 지질 나노입자, 예를 들어, cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K를 갖는 지질 나노입자 제형을 사용하여) 개별적으로 캡슐화된다. 예를 들어, 본 발명의 제1 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산(예를 들어, mRNA) 및 본 발명의 제2 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산(예를 들어, mRNA)은 각각 별도의 지질 나노입자 내에 캡슐화될 수 있고, 이에 이어서 이들은 제1 핵산을 캡슐화하는 지질 나노입자 및 제2 핵산을 캡슐화하는 지질 나노입자의 혼합물(일반적으로 1:1 비율임)을 제공하도록 조합될 수 있다.

[0659] 예를 들어, 본 발명에 따른 면역원성 조성물은 적어도 2개의 핵산을 포함할 수 있으며, 여기에서 제1 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고(예를 들어, 서열번호 44의 최적화된 뉴클레오티드 서열을 포함하는 mRNA, 또는 표 4에 제시된 예시적인 mRNA 작제물 1); 제2 핵산은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함한다(예를 들어, 서열번호 166의 최적화된 뉴클레오티드 서열을 포함하는 mRNA, 또는 표 4에 제시된 예시적인 mRNA 작제물 2). 일부 구현예에서, 제1 핵산은 제2 핵산과 조합될 수 있고 동일한 지질 나노입자 중 캡슐화될 수 있다. 다른 구현예에서, 제1 핵산 및 제2 핵산은 각각 별도의 지질 나노입자(일반적으로 동일한 지질 성분, 예를 들어, cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K로부터 형성됨) 중 캡슐화될 수 있다. 그런 다음, 제1 핵산을 캡슐화하는 지질 나노입자 및 제2 핵산을 캡슐화하는 지질 나노입자는 (통상적으로 1:1의 비율로) 조합된다.

[0660] ***약학적 조성물***

[0661] 본 발명에 따른 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산은 약학적 조성물(예를 들어, 면역원성 조성물 또는 백신)로 제공될 수 있다. 일반적인 구현예에서, 본 발명에 따른 약학적 조성물은 본 발명에 따른 핵산 및 지질 나노입자를 포함한다. 특정 구현예에서, 핵산은 지질 나노입자 중 캡슐화된다. 일부 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, PEG-변형 지질, 또는 이들의 조합 중 하나 이상을 포함할 수 있다. 일반적인 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 및 PEG-변형 지질을 포함한다. 일부 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 및 PEG-변형 지질을 포함한다.

[0662] *약학적으로 허용 가능한 부형제*

[0663] 핵산 및/또는 지질 나노입자를 안정화시키거나, 약학적 조성물의 투여를 용이하게 하고/하거나 본 발명의 핵산의 생체 내 발현을 향상시키기 위해, 핵산 및/또는 지질 나노입자는 하나 이상의 추가 핵산, 담체, 표적화 리간드, 안정화 시약, 및/또는 다른 약학적으로 허용 가능한 부형제와 조합하여 제형화될 수 있다. 약물의 제형화 및 투여를 위한 기술은 문헌["Remington's Pharmaceutical Sciences", Mack Publishing Co., Easton, Pa.,] 최신판에서 확인할 수 있다.

[0664] 일부 구현예에서, 약학적 조성물은 희석제와 함께 제형화된다. 일부 구현예에서, 희석제는 DMSO, 에틸렌 글리콜, 글리세롤, 2-메틸-2,4-펜탄디올(MPD), 프로필렌 글리콜, 수크로오스, 및 트레할로스로 이루어진 군으로부터 선택된다. 일부 구현예에서, 제형은 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19% 또는 20%의 희석제를 포함한다. 특정 구현예에서, mRNA는 희석제로서 10%의 트레할로스로 제형화된다.

[0665] *치료적 유효량*

[0666] 본 발명에 따른 핵산은 본원에 제공된 약학적 조성물 중 치료적 유효량으로 제공된다. 본원에서 사용되는 용어 "치료적 유효량"은 대체로 본 발명의 약학적 조성물에 함유된 치료제의 총량에 따라 결정된다. 대체적으로, 치료적 유효량은 대상체에서 의미있는 이익(예컨대, SARS-CoV-2 감염을 갖는 감염증의 치료 또는 예방)을 달성하기에 충분한 양이다. 예를 들어, 치료적 유효량은 본 발명의 면역원성 조성물을 사용하여 원하는 예방 효과를 달성하기에 충분한 양일 수 있다.

[0667] 일부 구현예에서, 본 발명에 따른 약학적 조성물(예를 들어, 면역원성 조성물)은 0.1 mg/mL 내지 10.0 mg/mL 범위의 농도로 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA를 포함한다. 일부 구현예에서, mRNA는 적어도 0.1 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.2 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.3 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.4 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.5 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.6 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.7 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.8 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 0.9 mg/mL의 농도로 존재한다. 일부 구현예에서, mRNA는 적어도 1.0 mg/mL의 농도로 존재한다. 전형적인 구현예에서, mRNA는 약 0.5 mg/mL 내지 약 1.0 mg/mL, 예를 들어 약 0.6 mg/mL 내지 약 0.8 mg/mL의 농도로 존재한다.

[0668] 일부 구현예에서, 본 발명에 따른 약학적 조성물(예를 들어, 면역원성 조성물)은 5 μg 내지 200 μg의 투여량으로 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA를 포함한다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 10 μg 내지 200 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 7 μg 내지 135 μg이다. 특정 구현예에서, 약학적 조성물 중의 mRNA 투여량은 15 μg 내지 135 μg(예를 들어, 15 μg 내지 45 μg)이다.

[0669] 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 5 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 10 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 15 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 20 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 25 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 30 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 35 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 40 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 45 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 50 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 75 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 100 μg이다. 일부 구현예에서, 약학적 조성물 중의 mRNA 투여량은 적어도 150 μg이다.

[0670] 특정 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 7.5 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 10 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 15 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 20 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 30 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 40 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 45 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량은 약 135 μg이다.

[0671] 일부 구현예에서, 본 발명에 따른 약학적 조성물(예를 들어, 면역원성 조성물)은 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열(예를 들어, SARS-SARS-CoV-2 S 단백질의 자연 발생 변이체를 암호화하는 2개의 mRNA 작제물)을 포함하는 하나 이상의 mRNA 작제물(예를 들어, 적어도 2개의 mRNA 작제물)을 포함한다. 따라서, 일부 구현예에서, mRNA 작제물의 총 투여량은 5 μg 내지 200 μg이다. 예를 들어, mRNA 작제물의 총 투여량은 10 μg 내지 200 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 7 μg 내지 135 μg이다. 특정 구현예에서, mRNA 작제물의 총 투여량은 15 μg 내지 135 μg(예를 들어, 15 μg 내지 45 μg)이다.

[0672] 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 5 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 10 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 15 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 20 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 25 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 30 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 35 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 40 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 45 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 50 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 75 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 100 μg이다. 일부 구현예에서, mRNA 작제물의 총 투여량은 적어도 150 μg이다.

[0673] 특정 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 7.5 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 10 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 15 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 20 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 30 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 40 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 45 μg이다. 또 다른 구현예에서, 약학적 조성물 중의 mRNA 투여량의 총 투여량은 약 135 μg이다.

[0674] *SARS-CoV-2 S 단백질의 조합*

[0675] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 스파이크 단백질을 암호화하는 하나 이상의 최적화된 뉴클레오티드 서열을 포함한다. 일부 구현예에서, 각각의 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질의 자연 발생 변이체를 암호화한다. 일부 구현예에서, 이들 최적화된 뉴클레오티드 서열 중 하나 이상은 자연 발생 SARS-CoV-2 스파이크 단백질에 비해 변형된 SARS-CoV-2 스파이크 단백질을 암호화한다. 특정 구현예에서, 변형은 전술한 바와 같은, 융합-전 형태의 SARS-CoV-2 스파이크 단백질을 안정화시킨다.

[0676] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서 제1 핵산은 서열번호 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 35, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84. 86, 88, 90, 92, 94, 96, 98, 104, 106, 108, 110, 118, 120, 123, 125, 127, 129, 131, 133, 135, 137, 139 또는 141로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 하나 이상의 추가의 핵산(들)은 서열번호 151, 153, 155, 157, 159, 161, 163, 165, 167, 169 또는 171로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0677] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 제2 핵산은 서열번호 157의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0678] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서, 제1 핵산은 서열번호 44의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적

으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 갖고; 제2 핵산은 서열번호 156의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 157을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 156의 핵산 서열을 갖는다.

[0679] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 제2 핵산은 서열번호 163의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0680] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서, 제1 핵산은 서열번호 44의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 갖고; 제2 핵산은 서열번호 162의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 163을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 162의 핵산 서열을 갖는다.

[0681] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 제2 핵산은 서열번호 167의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0682] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서, 제1 핵산은 서열번호 44의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 갖고; 제2 핵산은 서열번호 166의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 167을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166의 핵산 서열을 갖는다.

[0683] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 제2 핵산은 서열번호 171의 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0684] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 2개의 핵산을 포함하고, 여기에서, 제1 핵산은 서열번호 44의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 갖고; 제2 핵산은 서열번호 170의 뉴클레오티드 서열과 적어도 85%(예를 들어, 적어도 90%) 동일하고 서열번호 171을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 170의 핵산 서열을 갖는다.

[0685] 일부 구현예에서, 본 발명에 따른 면역원성 조성물은 SARS-CoV-2 감염의 예방에 사용하기 위한 적어도 3개, 적어도 4개 또는 적어도 5개의 핵산을 포함한다. 적용 가능한 경우, 제1, 제2, 제3, 제4 및 제5 핵산은 동일한 지질 나노입자 중 캡슐화될 수 있다. 대안적으로, 적용 가능한 경우, 제1, 제2, 제3, 제4 및 제5 핵산은 함께 혼합되어 본 발명에 따른 약학적 조성물을 형성하는 별도의 지질 나노입자 중 캡슐화될 수 있다.

[0686] *SARS-CoV-2 항원의 조합*

[0687] 일부 구현예에서, 본 발명에 따른 약학적 조성물은 SARS-CoV-2 항원을 암호화하는 하나 이상의 최적화된 뉴클레오티드 서열을 포함한다. 일부 구현예에서, 약학적 조성물은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산 및 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산을 포함할 수 있다. 일부 구현예에서, 약학적 조성물은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵

산 및 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산을 포함할 수 있다. 일부 구현예에서, 약학적 조성물은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산 및 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제2 핵산을 포함할 수 있다. 다른 구현예에서, 약학적 조성물은 SARS-CoV-2 S 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 제1 핵산, 및 제2, 제3 및/또는 제4 핵산을 포함할 수 있으며, 여기에서 전술한 제2 핵산은 SARS-CoV-2 M 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 전술한 제3 핵산은 SARS-CoV-2 N 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 전술한 제4 핵산은 SARS-CoV-2 E 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0688] 적용 가능한 경우, 제1, 제2, 제3, 및 제4 핵산은 동일한 지질 나노입자 중 캡슐화될 수 있다. 대안적으로, 적용 가능한 경우, 제1, 제2, 제3, 및 제4 핵산은 함께 혼합되어 본 발명에 따른 약학적 조성물을 형성하는 별도의 지질 나노입자 중 캡슐화될 수 있다.

[0689] *투여*

[0690] 일반적으로, 본 발명에 따른 약학적 조성물(예를 들어, 면역원성 조성물 또는 백신)은, 예를 들어, 정맥내, 피내, 피하, 또는 근육내 경로에 의해 비경구 투여된다. 가장 일반적으로, 투여는 근육내 투여이다. 투여는 주사, 예를 들어, 바늘이 없는 주사 및/또는 바늘 주사에 의한 것일 수 있다.

[0691] 예를 들어, Fenton 등(Adv Mater. 2017; 29(33))의 양이온성 지질 OF-Deg-Lin을 함유하는 지질 나노입자의 사용은 정맥내 주사를 통해 비장으로 캡슐화된 mRNA를 성공적으로 전달할 수 있었다. 이들은 총 단백질 생성의 85% 이상이 비장에서 발생하였음을 관찰하였다. 이들이 시험 동물의 비장을 분석했을 때, 지질 나노입자가 캡슐화된 mRNA를 주로 B 세포 및 단핵구/대식세포 집단에 전달하였음을 발견하였다. mRNA의 작은 백분율은 또한 호중구 및 T 세포 집단에 전달되는 것으로 나타났다. 본 명세서의 실시예에 나타낸 바와 같이, 40:30:28.5:1.5 몰비의 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K로 이루어진 지질 성분을 갖는 지질 나노입자를 포함하는 약학적 조성물은 특히, 캡슐화된 핵산(들)에 대한 면역 반응을 유도하는 데 효과적이며, 특히 근육내 투여될 경우 효과적이다.

[0692] *프라임-부스트 면역화*

[0693] 일부 구현예에서, 본 발명에 따른 약학적 조성물은 1회 투여된다. 일부 구현예에서, 본 발명에 따른 약학적 조성물은 적어도 2회 투여된다.

[0694] 예를 들어, β-코로나바이러스, 예를 들어, SARS-CoV-2로의 감염에 대해 이전에 면역화되지 않은 대상체의 전형적인 프라임-부스트 면역화는, 일반적으로 적어도 2회의 면역화를 포함한다. 일반적으로, 이들 2회의 면역화는 일정 간격으로 투여된다. 따라서, 일부 구현예에서, 본 발명에 따른 약학적 조성물은 2, 3, 4, 5, 6, 7 또는 8주의 간격으로 적어도 2회(예를 들어, 3회) 투여된다. 일부 구현예에서, 본 발명에 따른 약학적 조성물은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 또는 12주의 간격으로 2회 투여된다. 일반적인 구현예에서, 투여 간격은 2주 또는 4주(예를 들어, 1개월)이다. 다른 구현예에서, 투여 간격은 11주, 또는 12주(예를 들어, 약 3개월)이다. 따라서, 일 구현예에서, 본 발명은 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염을 예방하는 방법을 제공하며, 전술한 방법은 본 발명의 mRNA 작제물을 포함하는 면역원성 조성물의 제1 투여량, 및 본 발명의 면역원성 조성물의 제2 투여량을 대상체에게 투여하는 단계를 포함하되, 전술한 제1 투여량 및 제2 투여량은 서로 적어도 2주 간격으로 투여된다. 일부 구현예에서, 본 발명은 β-코로나바이러스(예를 들어, SARS-CoV-2)에 의해 야기되는 감염을 예방하는 방법을 제공하며, 전술한 방법은 본 발명의 mRNA 작제물을 포함하는 면역원성 조성물의 제1 투여량, 및 본 발명의 면역원성 조성물의 제2 투여량을 대상체에게 투여하는 단계를 포함하되, 전술한 제1 투여량 및 제2 투여량은 서로 3주 간격으로 투여된다.

[0695] 때때로, 최초 프라임-부스트 면역화 후에 적어도 하나의 추가 면역화가 이어져 최초 면역화 시리즈의 보호 효과를 갱신한다. 이러한 추가 면역화는 일반적으로 최초 프라임-부스트 면역화 후 수개월, 때로는 수년에 걸쳐 진행된다. 따라서, 일부 구현예에서, 본 발명에 따른 약학적 조성물은, 대상체가 β-코로나바이러스, 예를 들어 SARS-CoV-2와 같은 안지오텐신-전환 효소 2(ACE 2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스의 감염 예방을 위한 면역원성 조성물의 적어도 1회 투여량을 투여받은 후, 3 내지 18개월(예를 들어, 약 9개월 또는 약 12개월)에 적어도 1회 대상체에게 투여된다. 예를 들어, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2)의 감염 예방을 위해 적어도 1회 투여량의 면역원성 조성물을 투여받았을 수 있고, 3 내지 18개월(예를

들어, 약 9개월 또는 약 12개월) 후에 본 발명의 약학적 조성물이 투여될 수 있다. 보다 일반적으로, 대상체는 β-코로나바이러스(예를 들어, SARS-CoV-2) 감염 예방을 위해 면역원성 조성물의 2회 투여량을, 예를 들어, 제1 투여 및 적어도 2주 후, 제2 투여의 방식으로 투여받았을 수 있다. 제2 투여량을 투여받은 수 후 3 내지 18개월 차에, 대상체에게 본 발명의 약학적 조성물이 투여될 수 있다. 본 발명의 약학적 조성물의 투여는 대상체가 β-코로나바이러스(예를 들어, SARS-CoV-2) 감염 예방을 위해 면역원성 조성물의 제2 투여량을 투여받은 후, 적어도 9개월차(예를 들어, 약 12개월차)에 흔히 이루어질 수 있다.

[0696] 일부 구현예에서, 제1 및 제2 투여는 β-코로나바이러스(예를 들어, SARS-CoV-2) 감염 예방을 위한 면역원성 조성물, 예를 들어, 우한으로부터의 SARS-CoV-2 인덱스 계열의 S 단백질에 대한 중화 항체를 유도하는 백신(서열번호 1)일 수 있다. 예를 들어, 백신은, 전장 SARS-CoV-2 스파이크 단백질을 그의 융합-전 형태로 안정화시키기 위해 잔기 986 및 987을 프롤린으로 돌연변이시키도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 핵산을 포함할 수 있다. 중화 항체를 유도하는 백신은, 본원에 기술된 약학적 조성물(예를 들어, 본원에 개시된 면역원성 조성물)뿐만 아니라, 모데나(Moderna)(COVIDCOVID-19 Vaccine Moderna, 예컨대, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(AstraZeneca)(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 및 Novavax(NVX-CoV2373)에서 제조한 COVID-19 백신을 포함한다. 제1 투여 및 제2 투여는 동일한 백신을 포함할 수 있다. 제1 투여 및 제2 투여는 상이한 백신을 포함할 수 있다.

[0697] 특정 구현예에서, 3 내지 18개월 후 투여되는 본 발명의 약학적 조성물은, 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함하는 핵산(예를 들어, mRNA)을 포함한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열을 포함하는 핵산(예를 들어, mRNA)은 우한 인덱스 계열뿐만 아니라, 남아공, 일본, 브라질, UK, 인도 및 캘리포니아에서 관찰된 변이체를 포함하는 SARS-CoV-2의 자연 발생 변이체에 대해 광범위하게 중화 항체 반응을 유도할 수 있다. 일부 구현예에서, 최적화된 뉴클레오티드 서열을 포함하는 핵산(예를 들어, mRNA)은 SARS-CoV-1에 대한 중화 항체 반응을 유도할 수 있다. 특정 구현예에서, 최적화된 뉴클레오티드 서열을 포함하는 핵산(예를 들어, mRNA)은 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 대한 중화 항체 반응을 유도할 수 있다. 일부 구현예에서, 스파이크 단백질은 서열번호 1과 적어도 75%(예를 들어, 적어도 80%, 90%, 95% 또는 99%) 동일하다. 특정 구현예에서, 핵산(예를 들어, mRNA)은 서열번호 167을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166의 핵산 서열을 갖는다. 예를 들어, mRNA의 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 가질 수 있다.

[0698] 특정 일 구현예에서, 3 내지 18개월 후 투여되는 본 발명의 약학적 조성물은 적어도 2개의 핵산(예를 들어, 제1 mRNA 및 제2 mRNA)을 포함하고, 여기에서 제1 핵산(예를 들어, 제1 mRNA)은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고; 제2 핵산(예를 들어, 제2 mRNA)은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함한다. 특정 구현예에서, 제1 및 제2 mRNA를 포함하는 약학적 조성물은 우한 인덱스 계열뿐만 아니라, 남아공, 일본, 브라질, UK, 인도 및 캘리포니아에서 관찰된 변이체를 포함하는 SARS-CoV-2의 자연 발생 변이체에 대해 광범위하게 중화 항체 반응을 유도할 수 있다. 일부 구현예에서, 제1 및 제2 mRNA를 포함하는 약학적 조성물은 SARS-CoV-1에 대한 중화 항체 반응을 유도할 수 있다. 특정 구현예에서, 제1 및 제2 mRNA를 포함하는 약학적 조성물은 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 대한 중화 항체 반응을 유도할 수 있다. 일부 구현예에서, 스파이크 단백질은 서열번호 1과 적어도 75%(예를 들어, 적어도 80%, 90%, 95% 또는 99%) 동일하다. 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함할 수 있으며, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 갖는다. 제2 핵산은 서열번호 167을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166의 핵산 서열을 갖는다. 예를 들어, 제1 mRNA의 최적화된 뉴클레오티드 서열은 서열번호 148의 핵산 서열을 가질 수 있고, 제2 mRNA의 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 가질 수 있다. 일반적으로, 적어도

2개의 핵산은 지질 나노입자 중 캡슐화된다. 예를 들어, 제1 핵산 및 제2 핵산은 동일한 지질 나노입자 중 캡슐화될 수 있다. 대안적으로, 제1 핵산 및 제2 핵산은 개별적인 지질 나노입자 중 캡슐화될 수 있다.

[0699] 실시예에 나타낸 바와 같이, 우한으로부터 SARS-CoV-2 인덱스 계열의 S 단백질(서열번호 1)에 대한 중화 항체를 유도하는 백신으로 이전에 면역화된 대상체 및 SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체를 암호화하는 본 발명의 최적화된 뉴클레오티드 서열을 포함하는 mRNA 백신을 약 9개월 후에 투여받은 대상체는, 원래의 SARS-CoV-2 우한 계열 뿐만 아니라 다른 β-코로나바이러스, 특히 SARS-CoV-1과 같은 안지오텐신-전환 효소 2(ACE 2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 의해 발현되는 광범위한 S 단백질에 대해 효과적인 광범위 중화 항체 반응을 마운트할 수 있다.

[0700] 따라서, 일부 구현예에서, 본 발명의 약학적 조성물은 β-코로나바이러스, 특히 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 의해 야기되는 감염 예방에 사용하기 위한 것이다. 일부 구현예에서, 본 발명의 약학적 조성물은 β-코로나바이러스, 특히 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 의해 야기되는 감염 예방을 위한 의약의 제조에 사용하기 위한 것이다. 일부 구현예에서, 스파이크 단백질은 서열번호 1과 적어도 75%(예를 들어, 적어도 80%, 90%, 95% 또는 99%) 동일하다. 일반적인 구현예에서, β-코로나바이러스는 SARS-CoV-2(예를 들어, 남아공 변이체, 일본 변이체, 브라질 변이체, UK 변이체, 인도 변이체 또는 캘리포니아 변이체와 같은 우한 인덱스 계열의 자연 발생 변이체)이다.

[0701] 특정 일 구현예에서, 본 발명은 SARS-CoV-2에 의해 야기되는 감염을 예방하는 방법을 추가로 제공하며, 전술한 방법은 mRNA 작제물을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하며, 여기에서 전술한 mRNA 작제물은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이화하도록 서열번호 1의 자연 발생 전장 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 스파이크 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, L18F, D80A, D215G, L242-, A243-, L244-, K417N, E484K, N501Y, D614G 및 A701V 돌연변이를 추가로 포함하고, 전술한 면역원성 조성물은 대상체가 제1 COVID-19 백신 및 제2 COVID-19 백신으로 면역화된 후 적어도 3개월(예를 들어, 약 6개월, 약 9개월 또는 약 12개월)차에 대상체에게 투여되고, 전술한 제1 및 제2 COVID-19 백신은 서로 적어도 2주 간격을 두고 대상체에게 투여되며, 여기에서 전술한 제1 및 제2 COVID-19 백신은 SARS-CoV-2, 예를 들어 우한으로부터의 SARS-CoV-2 인덱스 계열의 S-단백질(서열번호 1)에 대한 중화 항체를 유도하도록 설계된다. 일부 구현예에서, 제1 및 제2 COVID-19 백신은 동일하다. 다른 구현예에서, 전술한 제1 및 제2 백신은 상이하다. 특정 구현예에서, 전술한 제1 및 제2 COVID-19 백신은 모데나(COVID-19 Vaccine Moderna, 예컨대, 예를 들어, mRNA-1273 또는 mRNA-1283), CureVac(CVnCoV), Johnson & Johnson(COVID-19 Vaccine Janssen), 아스트라제네카(Vaxzevria), Pfizer/BioNTech(Comirnaty), Sputnik(Gam-COVID-Vac), Sinovac(COVID-19 Vaccine (Vero Cell) 비활성화) 또는 Novavax(NVX-CoV2373)에서 제조된다.

[0702] 일부 구현예에서, 면역원성 조성물은 우한 인덱스 계열뿐만 아니라, 남아공, 일본, 브라질, UK, 인도 및 캘리포니아에서 관찰된 변이체를 포함하는 SARS-CoV-2의 자연 발생 변이체에 대해 광범위하게 중화 항체 반응을 유도할 수 있다. 일부 구현예에서, 면역원성 조성물은 SARS-CoV-1에 대한 중화 항체 반응을 유도할 수 있다. 특정 구현예에서, 면역원성 조성물은 안지오텐신-전환 효소 2(ACE2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 대한 중화 항체 반응을 유도할 수 있다. 특정 구현예에서, 스파이크 단백질은 서열번호 1과 적어도 75%(예를 들어, 적어도 80%, 90%, 95% 또는 99%) 동일하다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 167을 포함하는 아미노산 서열을 암호화하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 173의 핵산 서열을 갖는다. 특정 구현예에서, mRNA 작제물은 mRNA 작제물 2이다. 특정 구현예에서, 전술한 mRNA 작제물은 cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K로 이루어진 지질 성분을, 예를 들어, 40:30:28.5:1.5의 몰비로 지질 나노입자 중 캡슐화된다. 일부 구현예에서, 면역원성 조성물은 7 μg 내지 135 μg, 예를 들어 7.5 μg, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물을 포함한다.

[0703] **본 발명의 추가의 예시적인 구현예**

[0704] 일 양태에서, 본 발명은 SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산을 제공하며, 여기에서 최적화된 뉴클레오티드 서열은 10% 이상의 사용 빈도와 연관된 코돈으로 이루어지되; 최적화된 뉴클레오티드 서열은,

[0705] (i) 다음의 뉴클레오티드 서열 중 하나를 갖는 종결 신호를 함유하지 않고,

[0706] 5'-$X_1$ATCT$X_2$T$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, T 또는 G로부터 독립적으로 선택됨), 및 5'-$X_1$AUCU$X_2$U$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, U 또는 G로부터 독립적으로 선택됨);

[0707] (ii) 시스 조절 요소 및 음성 반복 요소를 함유하지 않으며;

[0708] (iii) 0.8을 초과하는 코돈 적용 인덱스를 가지되;

[0709] 중첩되지 않는 30개의 뉴클레오티드-길이의 부분으로 나누어질 경우, 상기 최적화된 뉴클레오티드 서열의 각 부분은 30% 내지 70%의 구아닌 시토신 함량 범위를 갖는다.

[0710] 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 다음의 서열 중 하나를 갖는 종결 신호를 함유하지 않는다: TATCTGTT; TTTTTT; AAGCTT; GAAGAGC; TCTAGA; UAUCUGUU; UUUUUU; AAGCUU; GAAGAGC; UCUAGA. 특정 구현예에서, 핵산은 mRNA 또는 DNA이다.

[0711] 다음에서, 변형된 SARS-CoV-2 스파이크 단백질 또는 이의 항원성 단편은 특정 최적화된 핵산 서열을 참조하여 기술된다. 이들 변형된 SARS-CoV-2 스파이크 단백질 또는 항원성 단편은 본 발명에서의 개시된 핵산 기반 백신의 맥락에서 특정한 유용성을 가질 수 있지만, 이들은 또한 단백질 기반 백신에서도 유용성을 가질 수도 있다는 것을 이해해야 한다. 또한, 최적화된 핵산 서열은 이러한 단백질 기반 백신의 효율적인 생산에 유용할 수 있다.

[0712] 소정의 양태에서, 본 발명의 핵산은 SARS-CoV-2 스파이크 단백질 또는 이의 항원성 단편을 암호화하는 최적화된 뉴클레오티드 서열이다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 전장 SARS-CoV-2 스파이크 단백질을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 1을 포함하는 아미노산 서열을 암호화한다. 다른 구현예에서, 본 발명의 핵산은 SARS-CoV-2 스파이크 단백질 또는 이의 항원성 단편의 엑토도메인을 암호화하는 최적화된 뉴클레오티드 서열이다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 2를 포함하는 아미노산 서열을 암호화한다. 소정의 구현예에서, 항원성 단편은 SARS-CoV-2 스파이크 단백질의 수용체 결합 도메인(RBD)을 포함한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 6을 포함하는 아미노산 서열을 암호화한다.

[0713] 소정의 구현예에서, 항원성 단편은 신호 서열을 추가로 포함한다. 소정의 구현예에서, 신호 서열은 서열번호 7이다. 다른 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 8을 포함하는 아미노산 서열을 암호화한다. 소정의 구현예에서, 신호 서열은 서열번호 142이다. 다른 구현예에서, 본 발명의 최적화된 뉴클레오티드 서열은 서열번호 143을 포함하는 아미노산 서열을 암호화한다. 본 발명의 추가적인 양태에서, 항원성 단편은 Fc 영역을 추가로 포함할 수 있다. 특정 구현예에서, Fc 영역은 서열번호 18의 아미노산 서열을 갖는다. 소정의 구현예에서, 항원성 단편은 신호 서열 및 Fc 영역을 추가로 포함한다.

[0714] 소정의 구현예에서, 항원성 단편은 신호 서열 및 Fc 영역에 작동가능하게 연결된 SARS-CoV-2 스파이크 단백질의 RBD로 이루어진다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 20을 포함하는 아미노산 서열을 암호화한다.

[0715] 다른 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 안정적인 융합-전 형태를 형성하도록 변형된다. 특정 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 활성화에 필요한 퓨린 절단 부위를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화한다. 추가의 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 9를 포함하는 아미노산 서열을 암호화한다.

[0716] 특정 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 잔기 985를 프롤린으로 돌연변이시키고/시키거나 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화한다. 추가의 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 10을 포함하는 아미노산 서열을 암호화한다. 추가의 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 118을 포함하는 아미노산 서열을 암호화한다.

[0717] 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 잔기 985, 986 및 987을 프롤린으로 돌연변이시키도록 자연

발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 92를 포함하는 아미노산 서열을 암호화한다.

[0718] 특정 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 (a) 잔기 985를 프롤린으로 돌연변이시키고/시키거나; (b) 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 11을 포함하는 아미노산 서열을 암호화한다. 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는다. 추가의 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 120을 포함하는 아미노산 서열을 암호화한다. 예를 들어, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질의 엑토도메인을 암호화한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 12를 포함하는 아미노산 서열을 암호화한다.

[0719] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 985, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질을 암호화한다. 추가의 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 94를 포함하는 아미노산 서열을 암호화한다.

[0720] 특정 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 118을 포함하는 아미노산 서열을 암호화한다.

[0721] 특정 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고, 잔기 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 120을 포함하는 아미노산 서열을 암호화한다.

[0722] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 129를 포함하는 아미노산 서열을 암호화한다.

[0723] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 131을 포함하는 아미노산 서열을 암호화한다.

[0724] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 133을 포함하는 아미노산 서열을 암호화한다.

[0725] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 135를 포함하는 아미노산 서열을 암호화한다.

[0726] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 123을 포함하는 아미노산 서열을 암호화한다. 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 137을 포함하는 아미노산 서열을 암호화한다.

[0727] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 ER 회수 신호를 돌

연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 125를 포함하는 아미노산 서열을 암호화한다.

[0728] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 127을 포함하는 아미노산 서열을 암호화한다.

[0729] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 139를 포함하는 아미노산 서열을 암호화한다.

[0730] 소정의 구현예에서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 141을 포함하는 아미노산 서열을 암호화한다.

[0731] 특정 구현예에서, 항원성 단편은 SARS-CoV-2 스파이크 단백질의 S1, S2 또는 S2' 서브유닛을 포함하거나 이로 이루어진다. 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 3, 서열번호 4 또는 서열번호 5를 포함하는 아미노산 서열을 암호화한다.

[0732] 소정의 구현예에서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 S 단백질의 하나 이상의 항원성 단편을 포함하는 융합 펩티드를 암호화한다. 특정 구현예에서, SARS-CoV-2 S 단백질의 하나 이상의 항원성 단편은 서열번호 21의 아미노산 서열, 서열번호 22의 아미노산 서열, 서열번호 23의 아미노산 서열 및/또는 서열번호 24의 아미노산 서열을 갖는다.

[0733] 소정의 구현예에서, 하나 이상의 항원성 단편은 링커 서열, 예를 들어, GGGGS에 의해 연결된다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 25 또는 서열번호 27을 포함하는 융합 펩티드를 암호화한다. 소정의 구현예에서, 융합 펩티드는 N 말단 신호 서열을 포함한다; 예를 들어, 최적화된 뉴클레오티드 서열은 서열번호 51 또는 서열번호 53을 포함하는 융합 펩티드를 암호화한다. 소정의 구현예에서, 융합 펩티드는 C-말단 Fc 도메인을 포함한다. 다른 구현예에서, 융합 펩티드는 N 말단 신호 서열 및 C-말단 Fc 도메인을 포함한다. 특정 구현예에서, 최적화된 뉴클레오티드 서열은 서열번호 55 또는 서열번호 57을 포함하는 융합 펩티드를 암호화한다.

[0734] 다른 양태에서, 전술한 바와 같은 본 발명의 핵산은 치료에 사용하기 위한 것이다. 예를 들어, 전술한 바와 같은 본 발명의 핵산은 SARS-CoV-2 감염 예방을 위한 의약의 제조에 사용될 수 있다. 다른 양태에서, SARS-CoV-2 감염 예방에 사용하기 위한 본 발명의 핵산을 포함하는 면역원성 조성물이 제공된다. 본 발명은 또한 SARS-CoV-2 감염을 치료하거나 예방하는 방법을 제공하며, 전술한 방법은 본 발명의 핵산을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함한다.

[0735] 다른 양태에서, SARS-CoV-2 감염 예방에 사용하기 위한, 적어도 2개의 핵산을 포함하는 본 발명에 따른 면역원성 조성물이 제공되며, 여기에서 제1 핵산은 서열번호 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 35, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84. 86, 88, 90, 92, 94, 96, 98, 104, 106, 108, 110, 118, 120, 123, 125, 127, 129, 131, 133, 135, 137, 139 또는 141로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 하나 이상의 추가의 핵산(들)은 서열번호 151, 153, 155, 157, 159, 161, 163, 165, 167, 169 또는 171로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함한다.

[0736] 다른 양태에서, SARS-CoV-2 감염 예방에 사용하기 위한, 적어도 2개의 핵산을 포함하는 본 발명에 따른 면역원성 조성물이 제공되며, 여기에서 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 가지며, 하나 이상의 추가의 핵산(들)은 다음으로부터 선택된다:

[0737] (a) 서열번호 157로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을

포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 156의 핵산 서열을 갖는, 핵산; 및

[0738] (b) 서열번호 163으로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 162의 핵산 서열을 갖는, 핵산; 및

[0739] (c) 서열번호 167로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166의 핵산 서열을 갖는, 핵산; 및

[0740] (d) 서열번호 171로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 170의 핵산 서열을 갖는, 핵산.

[0741] 소정의 양태에서, 본 발명은 i) 본 발명의 핵산 및 ii) 지질 나노입자를 포함하는 약학적 조성물을 제공한다. 소정의 구현예에서, 핵산은 지질 나노입자 중 캡슐화된다. 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, PEG-변형 지질, 또는 이들의 조합 중 하나 이상을 포함할 수 있다. 소정의 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 및 PEG-변형 지질을 포함한다.

[0742] 소정의 구현예에서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 및 PEG-변형 지질을 포함한다. 소정의 구현예에서, 지질 나노입자는 다음을 포함한다:

[0743] (a) DOTAP(1,2-디올레일-3-트리메틸암모늄 프로판), DODAP(1,2-디올레일-3-디메틸암모늄 프로판), DOTMA(N-[1-(2,3-디올레일옥시)프로필]-N,N,N-트리메틸암모늄 염화물), DLinKC2DMA, DLin-KC2-DM, C12-200, cKK-E12, cKK-E10, HGT5000, HGT5001 , HGT4003, ICE, HGT4001, HGT4002, TL1-01D-DMA, TL1-04D-DMA, TL1-08D-DMA, TL1-10D-DMA, OF-Deg-Lin 및 OF-02로부터 선택되는 양이온성 지질;

[0744] (b) DSPC(1,2-디스테라로일-sn-글리세로-3-포스포콜린), DPPC (1,2-디팔미토일-sn-글리세로-3-포스포콜린), DOPE(1,2-디올레일-sn-글리세로-3-포스포에탄올아민), DEPE 1,2-디에루코일-sn-글리세로-3-포스포에탄올아민, DOPC(1,2-디올레일-sn-글리세로-3-포스포티딜콜린), DPPE(1,2-디팔미토일-sn-글리세로-3-포스포에탄올아민), DMPE(1,2-디미리스토일-sn-글리세로-3-포스포에탄올아민), 및 DOPG(1,2-디올레일-sn-글리세로-3-포스포-(1'-rac-글리세롤))로부터 선택되는 비-양이온성 지질;

[0745] (c) DC-Choi(N,N-디메틸-N-에틸카르복사미도콜레스테롤), 1,4-비스(3-N-올레일아미노-프로필)피페라진, 또는 이미다졸 콜레스테롤 에스테르(ICE)로부터 선택되는 콜레스테롤계 지질; 및/또는

[0746] (d) PEG화 콜레스테롤 및 DMG-PEG-2K로부터 선택되는 PEG-변형 지질.

[0747] 약학적 조성물의 소정의 구현예에서,

[0748] a. 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고;

[0749] b. 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고;

[0750] c. 콜레스테롤계 지질은 콜레스테롤이고;

[0751] d. PEG-변형 지질은 DMG-PEG-2K이다.

[0752] 소정의 구현예에서, 양이온성 지질은, 몰비로 지질 나노입자의 약 30 내지 60%, 예를 들어, 약 35 내지 40%를 구성한다. 특정 구현예에서, 양이온성 지질 대 비-양이온성 지질 대 콜레스테롤계 지질 대 PEG-변형 지질의 비율은 대략 30 내지 60:25 내지 35:20 내지 30:1 내지15의 몰비이거나, 양이온성 지질 대 비-양이온성 지질 대 PEG-변형 지질의 비율은 대략 55 내지 65:30 내지 40:1 내지 15의 몰비이다.

[0753] 소정의 구현예에서, 지질 나노입자는, cKK-E12, DOPE, 콜레스테롤 및 DMG-PEG2K; cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-Deg-Lin, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-02, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-01D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-04D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-08D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-10D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; ICE, DOPE 및 DMG-PEG2K; HGT4001, DOPE 및 DMG-PEG2K; 또는 HGT4002, DOPE 및 DMG-PEG2K로부터 선택되는 양이온성 지질, 비-양이온성 지질, PEG-변형 지질 및 선택적으로 콜레스테롤의 조합을 포함한다.

[0754] 소정의 구현예에서, 지질 나노입자는 150 nm 미만, 예를 들어, 100 nm 미만의 평균 크기를 갖는다. 특정 구현예에서, 지질 나노입자는 약 50 내지 70 nm, 예를 들어, 약 55 내지 65 nm의 평균 크기를 갖는다.

[0755] 소정의 구현예에서, 지질 나노입자는 주사용수 중 10% 트레할로스에 현탁된다. 특정 구현예에서, 핵산은 약 0.5

mg/mL 내지 약 1.0 mg/mL 농도의 mRNA이다.

[0756] 소정의 양태에서, 본 발명은 i) 본 발명의 최적화된 핵산(예를 들어, mRNA) 및 ii) 지질 나노입자를 포함하는 약학적 조성물을 제공한다. 이러한 약학적 조성물은 SARS-CoV-2 감염의 치료 또는 예방에 사용되기 위한 것이다. 소정의 구현예에서, 약학적 조성물은 비경구 투여된다. 소정의 구현예에서, 약제학적 조성물은 정맥내, 피내, 피하, 또는 근육내 투여된다. 특정 구현예에서, 약학적 조성물은 정맥내 또는 근육내 투여된다.

[0757] 소정의 구현예에서, 약학적 조성물은 적어도 1회 투여된다. 특정 구현예에서, 약학적 조성물은 적어도 2회 투여된다. 보다 구체적인 구현예에서, 투여 사이의 기간은 적어도 2주, 예를 들어 1개월이다. 일부 구현예에서, 투여 사이의 기간은 약 3주이다.

[0758] 소정의 양태에서, 본 발명은 SARS-CoV-2 항원을 제공한다. 예를 들어, SARS-CoV-2 항원은, SARS-CoV-2 스파이크 단백질, 항원성 단편, 또는 특정 최적화된 핵산 서열을 참조하여 전술하거나 아래에서 더욱 상세히 기술되는 항원성 단편의 융합 단백질 중 어느 하나일 수 있다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 1의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 10의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 9의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 11의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 2의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 12의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 3의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 8의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 20의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 17의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 14의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 16의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 66의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 15의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 82의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 84의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 74의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 76의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 78의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 80의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 68의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 70의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 96의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 86의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 88의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 90의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 92의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 94의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 118의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다. 일부 구현예에서, SARS-CoV-2 항원은 서열번호 120의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드이다.

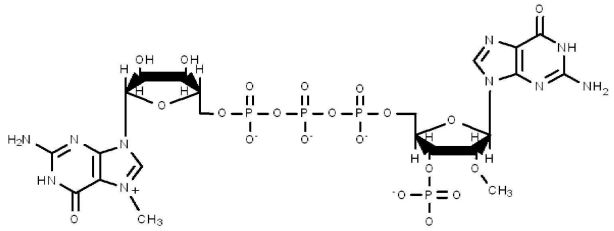
[0759] 추가의 양태에서, 본 발명은 SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 포함하는 펩티드 융합 작제물을 제공하며, 여기에서 하나 이상의 항원성 영역은 다음의 성분을 포함하거나 이로 이루어진다: FP, D1, D2 및/또는 B1, 여기에서 FP는 SARS-CoV-2 S 단백질의 잔기 815 내지 833을 포함하고, D1은 SARS-CoV-2 S 단백질의 잔기 820 내지 846을 포함하고, D2는 SARS-CoV-2 S 단백질의 잔기 1078 내지 1111을 포함하고, B1은 SARS-CoV-2 S 단백질의 잔기 798 내지 829를 포함한다. 펩티드 융합 작제물은 다음의 구조를 가질 수 있다: D1- 링커- FP - 링커 - D2 - 링커 - D1. D1은 서열번호 22의 서열을 가질 수 있다. FP는 서열번호 21의 서열을 가질 수 있다. 링커는 아미노산 서열 GGGGS를 포함하거나 이로 이루어진다. 예를 들어, 펩티드 융합 작제물은 서열번호 25 또

는 51, 55의 서열을 포함하거나 이로 이루어질 수 있다. 대안적으로, 펩티드 융합 작제물은 다음의 구조를 가질 수 있다: FP - 링커 - FP - 링커 - FP, D1 - 링커 - D1- 링커 - D1, 또는 FP/D1- 링커 - FP/D1- 링커- FP/D1. FP/D1 부분은 서열번호 99의 서열을 가질 수 있다. 링커는 아미노산 서열 GGGGS를 포함하거나 이로 이루어질 수 있다. 예를 들어, 펩티드 융합 작제물은 서열번호 27 또는 53, 57의 서열을 포함하거나 이로 이루어질 수 있다.

[0760] 본 발명은 또한 본 발명의 SARS-CoV-2 항원 또는 펩티드 융합 작제물을 포함하는 약학적 조성물을 제공한다. 일부 구현예에서, 약학적 조성물은 애쥬번트를 추가로 포함한다. 특정 구현예에서, 애쥬번트는 명반, CpG, PolyI:C, MF59, AS01, AS02, AS03, AS04, AF03, 편모, ISCOM 및 ISCOMMATRIX로부터 선택된다. 일부 구현예에서, 약학적 조성물은 SARS-CoV-2 감염의 치료 또는 예방에 사용되기 위한 것이다. 일부 구현예에서, 약학적 조성물은 비경구 투여된다. 일부 구현예에서, 약학적 조성물은 피내, 피하 또는 근육내 투여된다. 일부 구현예에서, 약학적 조성물은 근육내 투여된다. 일부 구현예에서, 약학적 조성물은 적어도 1회 투여된다. 일부 구현예에서, 약학적 조성물은 적어도 2회 투여된다. 일부 구현예에서, 투여 사이의 기간은 적어도 2주, 예를 들어 1개월이다. 일부 구현예에서, 투여 사이의 기간은 약 3주이다.

[0761] 특정 구현예에서, 본 발명은 다음의 구조적 요소로 이루어진 mRNA 작제물을 제공한다:

[0762] (i) 다음의 구조를 갖는 5' 캡:

[0763] 

[0764] (ii) 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

[0765] (iii) 서열번호 148의 핵산 서열을 갖는 단백질 코딩 영역;

[0766] (iv) 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

[0767] (v) 폴리A 꼬리.

[0768] 특정 구현예에서, 본 발명은 전술한 mRNA 작제물을 캡슐화하는 지질 나노입자를 제공한다. 지질 나노입자는 양이온성 지질(예를 들어, cKK-E12, cKK-E10, OF-Deg-Lin 또는 OF-02), 비-양이온성 지질(예를 들어, DOPE 또는 DEPE), 콜레스테롤계 지질(예를 들어, 콜레스테롤) 및 PEG-변형 지질(예를 들어, DMG-PEG-2K)을 포함할 수 있다. 특정 구현예에서, mRNA 작제물 또는 이를 캡슐화하는 지질 나노입자는 면역원성 조성물로서 제공된다. 일부 구현예에서, 면역원성 조성물은 10 μg 내지 200 μg의 mRNA 작제물을 포함한다. 특정 구현예에서, 면역원성 조성물은 15 μg 내지 135 μg(예를 들어, 15 μg 내지 45 μg)의 mRNA 작제물을 포함한다. 일부 구현예에서, 면역원성 조성물은 적어도 20 μg, 적어도 25 μg, 적어도 30 μg, 적어도 35 μg, 적어도 40 μg, 또는 적어도 45 μg의 mRNA 작제물을 포함할 수 있다. 특정 구현예에서, 면역원성 조성물은 15 μg, 45 μg 또는 135 μg의 mRNA 작제물을 포함한다. 본 발명은 추가적으로, SARS-CoV-2 감염을 치료하거나 예방하는 방법을 제공하며, 전술한 방법은 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함한다. 일부 구현예에서, 면역원성 조성물은 적어도 2회 대상체에게 투여된다. 일부 구현예에서, 투여 사이의 기간은 적어도 2주이다. 일부 구현예에서, 투여 사이의 기간은 약 3주이다.

[0769] 소정의 구현예에서, 본 발명은 다음의 번호로 매겨진 구현예에 의해 추가로 기술된다:

[0770] 1. SARS-CoV-2 항원을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산으로서, 상기 최적화된 뉴클레오티드 서열은 10% 이상인 사용 빈도와 연관된 코돈으로 이루어지되, 상기 최적화된 뉴클레오티드 서열은,

[0771] (i) 다음의 뉴클레오티드 서열 중 하나를 갖는 종결 신호를 함유하지 않고,

[0772] 5'-$X_1$ATCT$X_2$T$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, T 또는 G로부터 독립적으로 선택됨), 및 5'-$X_1$AUCU$X_2$U$X_3$-3'(여기에서, $X_1$, $X_2$ 및 $X_3$은 A, C, U 또는 G로부터 독립적으로 선택됨);

[0773] (ii) 시스 조절 요소 및 음성 반복 요소를 함유하지 않으며;

[0774] (iii) 0.8을 초과하는 코돈 적용 인덱스를 가지되;

[0775] 중첩되지 않는 30개의 뉴클레오티드-길이의 부분으로 나누어질 경우, 상기 최적화된 뉴클레오티드 서열의 각 부분은 30% 내지 70%의 구아닌 시토신 함량 범위를 갖는, 핵산.

[0776] 2. 구현예 1에 있어서, 최적화된 뉴클레오티드 서열은, TATCTGTT; TTTTTT; AAGCTT; GAAGAGC; TCTAGA; UAUCUGUU; UUUUUU; AAGCUU; GAAGAGC; UCUAGA의 서열 중 하나를 갖는 종결 신호를 함유하지 않는, 핵산.

[0777] 3. 구현예 1 또는 2에 있어서, 핵산은 mRNA인, 핵산.

[0778] 4. 구현예 1 또는 2에 있어서, 핵산은 DNA인, 핵산.

[0779] 5. 구현예 1 내지 4 중 어느 하나에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질 또는 이의 항원성 단편을 암호화하는, 핵산.

[0780] 6. 구현예 5에 있어서, 최적화된 핵산은 전장 SARS-CoV-2 스파이크 단백질을 암호화하는, 핵산.

[0781] 7. 구현예 5 또는 구현예 6에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 1을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0782] 8. 구현예 5에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질 또는 이의 항원성 단편의 엑토도메인을 암호화하는, 핵산.

[0783] 9. 구현예 8에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 2를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0784] 10. 구현예 5에 있어서, 항원성 단편은 SARS-CoV-2 스파이크 단백질의 수용체 결합 도메인(RBD)을 포함하는, 핵산.

[0785] 11. 구현예 10에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 6을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0786] 12. 구현예 10 또는 11에 있어서, 항원성 단편은 신호 서열을 추가로 포함하는, 핵산.

[0787] 13. 구현예 12에 있어서, 신호 서열은 서열번호 7인, 핵산.

[0788] 14. 구현예 12 또는 구현예 13에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 8을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0789] 15. 구현예 12에 있어서, 신호 서열은 서열번호 142인, 핵산.

[0790] 16. 구현예 12 또는 구현예 13에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 143을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0791] 17. 구현예 10 내지 16에 있어서, 항원성 단편은 Fc 영역을 추가로 포함하는, 핵산.

[0792] 18. 구현예 17에 있어서, Fc 영역은 서열번호 18인, 핵산.

[0793] 19. 구현예 10 내지 18에 있어서, 항원성 단편은 신호 서열 및 Fc 영역을 추가로 포함하는, 핵산.

[0794] 20. 구현예 10 내지 18에 있어서, 항원성 단편은 신호 서열 및 Fc 영역에 작동가능하게 연결된 SARS-CoV-2 스파이크 단백질의 RBD로 이루어지는, 핵산.

[0795] 21. 구현예 20에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 20을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0796] 22. 구현예 5, 구현예 6 또는 구현예 8 중 어느 하나에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 안정화된 융합-전 형태를 취하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0797] 23. 구현예 22에 있어서, SARS-CoV-2 스파이크 단백질, 엑토도메인 또는 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0798] 24. 구현예 23에 있어서, 최적화된 뉴클레오티드 서열은 활성화에 필요한 퓨린 절단 부위를 제거하도록 자연 발

생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화하는, 핵산.

[0799] 25. 구현예 23에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 9를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0800] 26. 구현예 22 내지 25에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 985를 프롤린으로 돌연변이시키고/시키거나 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0801] 27. 구현예 26에 있어서, 최적화된 뉴클레오티드 서열은 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 스파이크 단백질을 암호화하는, 핵산.

[0802] 28. 구현예 27에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 10 또는 서열번호 118을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0803] 29. 구현예 26에 있어서, 최적화된 뉴클레오티드 서열은 잔기 985, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 스파이크 단백질을 암호화하는, 핵산.

[0804] 30. 구현예 29에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 92를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0805] 31. 구현예 22 내지 30에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고

[0806] (a) 잔기 985를 프롤린으로 돌연변이시키고/시키거나;

[0807] (b) 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0808] 32. 구현예 31에 있어서, SARS-CoV-2 스파이크 단백질, SARS-CoV-2 스파이크 단백질의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0809] 33. 구현예 32에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질을 암호화하는, 핵산.

[0810] 34. 제33항에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 11 또는 서열번호 120을 포함하는 아미노산 서열을 암호화하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 44 또는 서열번호 148의 핵산 서열을 갖는, 핵산.

[0811] 35. 구현예 32에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질의 엑토도메인을 암호화하는, 핵산.

[0812] 36. 구현예 35에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 12를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0813] 37. 구현예 31에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 985, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0814] 38. 구현예 37에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질을 암호화하는, 핵산.

[0815] 39. 구현예 38에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 94를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0816] 40. 구현예 22 내지 39에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 변형되는, 핵산.

[0817] 41. 구현예 40에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 118을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0818] 42. 구현예 22 내지 41에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고, 잔기 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 변형되는,

핵산.

[0819] 43. 구현예 42에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 120을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0820] 44. 구현예 22 내지 43에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0821] 45. 구현예 44에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 129를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0822] 46. 구현예 22 내지 45에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 활성화에 필요한 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0823] 47. 구현예 46에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 131을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0824] 48. 구현예 22 내지 47에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0825] 49. 구현예 48에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 133을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0826] 50. 구현예 22 내지 49에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 D614G 돌연변이를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0827] 51. 구현예 50에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 135를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0828] 52. 구현예 22 내지 51에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0829] 53. 구현예 52에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 123을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0830] 54. 구현예 22 내지 53에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0831] 55. 구현예 54에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 137을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0832] 56. 구현예 22 내지 55에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 ER 회수 신호를 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0833] 57. 구현예 56에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0834] 58. 구현예 57에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 125를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0835] 59. 구현예 56에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0836] 60. 구현예 59에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 127을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0837] 61. 구현예 56에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0838] 62. 구현예 61에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 139를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0839] 63. 구현예 56에 있어서, SARS-CoV-2 스파이크 단백질, 이의 엑토도메인 또는 이의 항원성 단편은 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 돌연변이시키고 ER 회수 신호를 제거하고 연장된 N-말단 신호 펩티드를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형되는, 핵산.

[0840] 64. 구현예 63에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 141을 포함하는 아미노산 서열을 암호화하는, 핵산.

[0841] 65. 구현예 5에 있어서, 항원성 단편은 SARS-CoV-2 스파이크 단백질의 S1, S2 또는 S2' 서브유닛을 포함하거나 이로 이루어지는, 핵산.

[0842] 66. 구현예 65에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 3, 서열번호 4 또는 서열번호 5를 포함하는 아미노산 서열을 암호화하는, 핵산.

[0843] 67. 구현예 1 내지 4에 있어서, 최적화된 뉴클레오티드 서열은 SARS-CoV-2 스파이크 단백질의 하나 이상의 항원성 단편을 포함하는 융합 펩티드를 암호화하는, 핵산.

[0844] 68. 구현예 67에 있어서, SARS-CoV-2 스파이크 단백질의 하나 이상의 항원성 단편은 서열번호 21의 아미노산 서열, 서열번호 22의 아미노산 서열, 서열번호 23의 아미노산 서열 및/또는 서열번호 24의 아미노산 서열을 갖는, 핵산.

[0845] 69. 구현예 67 또는 68에 있어서, 하나 이상의 항원성 단편은 링커 서열, 예를 들어, GGGGS에 의해 연결되는, 핵산.

[0846] 70. 구현예 69에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 25 또는 서열번호 27을 포함하는 융합 단백질을 암호화하는, 핵산.

[0847] 71. 구현예 67 내지 70에 있어서, 융합 펩티드는 N 말단 신호 서열을 포함하는, 핵산.

[0848] 72. 구현예 71에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 51 또는 서열번호 53을 포함하는 융합 단백질을 암호화하는, 핵산.

[0849] 73. 구현예 67 내지 72에 있어서, 융합 펩티드는 C-말단 Fc 도메인을 포함하는, 핵산.

[0850] 74. 구현예 67 내지 73에 있어서, 융합 펩티드는 N 말단 신호 서열 및 C-말단 Fc 도메인을 포함하는, 핵산.

[0851] 75. 구현예 74에 있어서, 최적화된 뉴클레오티드 서열은 서열번호 55 또는 서열번호 57을 포함하는 융합 단백질을 암호화하는, 핵산.

[0852] 76. 구현예 1 내지 75 중 어느 하나에 있어서, 치료 요법에 사용하기 위한, 핵산.

[0853] 77. SARS-CoV-2 감염 예방에 사용하기 위한, 구현예 1 내지 76 중 어느 하나의 핵산을 포함하는, 면역원성 조성물.

[0854] 78. SARS-CoV-2 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 구현예 1 내지 76 중 어느 하나의 핵산을 포함하는 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

[0855] 79. 약학적 조성물로서, i) 구현예 1 내지 76 중 어느 하나의 핵산, 및 ii) 지질 나노입자를 포함하는, 약학적 조성물.

[0856] 80. 구현예 79에 있어서, 핵산은 지질 나노입자 중 캡슐화되는, 약학적 조성물.

[0857] 81. 구현예 79 또는 80에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, PEG-변형 지질, 또는 이들의 조합 중 하나 이상을 포함하는, 약학적 조성물.

[0858] 82. 구현예 81에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질, 및 PEG-변형 지질을 포함하는, 약학적 조성물.

[0859] 83. 구현예 79에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 및 PEG-변형 지질을 포함하는, 약학적 조성물.

[0860] 84. 구현예 79 내지 83 중 어느 하나에 있어서, 지질 나노입자는,

[0861] a. DOTAP(1,2-디올레일-3-트리메틸암모늄 프로판), DODAP(1,2-디올레일-3-디메틸암모늄 프로판), DOTMA(N-[1-(2,3-디올레일옥시)프로필]-N,N,N-트리메틸암모늄 염화물), DLinKC2DMA, DLin-KC2-DM, C12-200, cKK-E12, cKK-E10, HGT5000, HGT5001 , HGT4003, ICE, HGT4001, HGT4002, TL1-01D-DMA, TL1-04D-DMA, TL1-08D-DMA, TL1-10D-DMA, OF-Deg-Lin 및 OF-02로부터 선택되는 양이온성 지질;

[0862] b. DSPC(1,2-디스테라로일-sn-글리세로-3-포스포콜린), DPPC (1,2-디팔미토일-sn-글리세로-3-포스포콜린), DOPE(1,2-디올레일-sn-글리세로-3-포스포에탄올아민), DEPE 1,2-디에루코일-sn-글리세로-3-포스포에탄올아민, DOPC(1,2-디올레일-sn-글리세로-3-포스포티딜콜린), DPPE(1,2-디팔미토일-sn-글리세로-3-포스포에탄올아민), DMPE(1,2-디미리스토일-sn-글리세로-3-포스포에탄올아민), 및 DOPG(1,2-디올레일-sn-글리세로-3-포스포-(1'-rac-글리세롤))로부터 선택되는 비-양이온성 지질;

[0863] c. DC-Choi(N,N-디메틸-N-에틸카르복사미도콜레스테롤), 1,4-비스(3-N-올레일아미노-프로필)피페라진, 또는 이미다졸 콜레스테롤 에스테르(ICE)로부터 선택되는 콜레스테롤계 지질; 및/또는

[0864] d. PEG화 콜레스테롤 및 DMG-PEG-2K로부터 선택되는 PEG-변형 지질을 포함하는, 약학적 조성물.

[0865] 85. 구현예 82에 있어서,

[0866] a. 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고;

[0867] b. 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고;

[0868] c. 콜레스테롤계 지질은 콜레스테롤이고;

[0869] d. PEG-변형 지질은 DMG-PEG-2K인, 약학적 조성물.

[0870] 86. 구현예 79 내지 85 중 어느 하나에 있어서, 양이온성 지질은 약 30 내지 60%의 몰비로, 예를 들어, 약 35 내지 40%의 몰비로 지질 나노입자를 구성하는, 약학적 조성물.

[0871] 87. 구현예 79 내지 86 중 어느 하나에 있어서, 양이온성 지질 대 비-양이온성 지질 대 콜레스테롤계 지질 대 PEG-변형 지질의 비율은 대략 30 내지 60:25 내지 35:20 내지 30:1 내지15의 몰비이거나, 양이온성 지질 대 비-양이온성 지질 대 PEG-변형 지질의 비율은 대략 55 내지 65:30 내지 40:1 내지 15의 몰비인, 약학적 조성물.

[0872] 88. 구현예 79 내지 87 중 어느 하나에 있어서, 지질 나노입자는, cKK-E12, DOPE, 콜레스테롤 및 DMG-PEG2K; cKK-E10, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-Deg-Lin, DOPE, 콜레스테롤 및 DMG-PEG2K; OF-02, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-01D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-04D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-08D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; TL1-10D-DMA, DOPE, 콜레스테롤 및 DMG-PEG2K; ICE, DOPE 및 DMG-PEG2K; HGT4001, DOPE 및 DMG-PEG2K; 또는 HGT4002, DOPE 및 DMG-PEG2K로부터 선택되는 양이온성 지질, 비-양이온성 지질, PEG-변형 지질 및 선택적으로 콜레스테롤의 조합을 포함하는, 약학적 조성물.

[0873] 89. 구현예 79 내지 88 중 어느 하나에 있어서, 지질 나노입자는 150 nm 미만, 예를 들어, 100 nm 미만의 평균 크기를 갖는, 약학적 조성물.

[0874] 90. 구현예 89에 있어서, 지질 나노입자는 약 50 내지 70 nm, 예를 들어, 약 55 내지 65 nm의 평균 크기를 갖는, 약학적 조성물.

[0875] 91. 구현예 79 내지 90 중 어느 하나에 있어서, 지질 나노입자는 주사용수 중 10% 트레할로스에 현탁되는, 약학적 조성물.

[0876] 92. 구현예 79 내지 91 중 어느 하나에 있어서, 핵산은 약 0.5 mg/mL 내지 약 1.0 mg/mL 농도의 mRNA인, 약학적 조성물.

[0877] 93. 구현예 79 내지 92 중 어느 하나에 있어서, SARS-CoV-2 감염을 치료하거나 예방하는 데 사용하기 위한, 약

학적 조성물.

[0878] 94. 구현예 79 내지 93에 따른 용도에 있어서, 약학적 조성물은 비경구 투여되는, 약학적 조성물.

[0879] 95. 구현예 79 내지 93에 따른 용도에 있어서, 약학적 조성물은 정맥내, 피내, 피하 또는 근육내 투여되는, 약학적 조성물.

[0880] 96. 구현예 95에 따른 용도에 있어서, 약학적 조성물은 정맥내 투여되는, 약학적 조성물.

[0881] 97. 구현예 95에 따른 용도에 있어서, 약학적 조성물은 근육내 투여되는, 약학적 조성물.

[0882] 98. 구현예 79 내지 97 중 어느 하나에 따른 용도에 있어서, 약학적 조성물은 적어도 1회 투여되는, 약학적 조성물.

[0883] 99. 구현예 98에 따른 용도에 있어서, 약학적 조성물은 적어도 2회 투여되는, 약학적 조성물.

[0884] 100. 구현예 99에 따른 용도에 있어서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월인, 약학적 조성물.

[0885] 101. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 1의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0886] 102. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 10의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0887] 103. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 9의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0888] 104. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 11의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0889] 105. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 2의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0890] 106. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 12의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0891] 107. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 3의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0892] 108. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 8의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0893] 109. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 20의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0894] 110. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 17의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0895] 111. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 14의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0896] 112. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 16의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0897] 113. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 66의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0898] 114. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 15의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0899] 115. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 82의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0900] 116. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 84의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0901] 117. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 74의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0902] 118. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 76의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0903] 119. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 78의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0904] 120. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 80의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0905] 121. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 68의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0906] 122. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 70의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0907] 123. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 96의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0908] 124. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 86의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0909] 125. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 88의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0910] 126. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 90의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0911] 127. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 92의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0912] 128. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 94의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0913] 129. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 118의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0914] 130. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 120의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0915] 131. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 123의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0916] 132. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 125의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0917] 133. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 127의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0918] 134. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 129의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0919] 135. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 131의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0920] 136. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 133의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0921] 137. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 135의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0922] 138. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 139의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0923] 139. SARS-CoV-2 항원으로서, 상기 SARS-CoV-2 항원은 서열번호 141의 아미노산 서열을 포함하거나 이로 이루어진 폴리펩티드인, SARS-CoV-2 항원.

[0924] 140. SARS-CoV-2 S 단백질의 하나 이상의 항원성 영역을 포함하는 펩티드 융합 작제물로서, 상기 하나 이상의 항원성 영역은 다음의 성분을 포함하거나 이로 이루어지는, 융합 단백질: FP, D1, D2 및/또는 B1, 여기에서 FP는 SARS-CoV-2 S 단백질의 잔기 815 내지 833을 포함하고, D1은 SARS-CoV-2 S 단백질의 잔기 820 내지 846을 포함하고, D2는 SARS-CoV-2 S 단백질의 잔기 1078 내지 1111을 포함하고, B1은 SARS-CoV-2 S 단백질의 잔기 798 내지 829를 포함함.

[0925] 141. 구현예 140에 있어서, 펩티드 융합 작제물은 다음의 구조를 갖는, 펩티드 융합 작제물: D1- 링커- FP - 링커 - D2 - 링커 - D1.

[0926] 142. 구현예 141에 있어서, D1은 서열 번호 22의 서열을 갖는, 펩티드 융합 작제물.

[0927] 143. 구현예 140 또는 141에 있어서, FP는 서열 번호 21의 서열을 갖는, 펩티드 융합 작제물.

[0928] 144. 구현예 140, 141 및 142 중 어느 하나에 있어서, 링커는 아미노산 서열 GGGGS를 포함하거나 이로 이루어지는, 펩티드 융합 작제물.

[0929] 145. 구현예 140 내지 144 중 어느 하나에 있어서, 서열번호 25 또는 51, 55의 서열을 포함하거나 이로 이루어지는, 펩티드 융합 작제물.

[0930] 146. 구현예 140에 있어서, 펩티드 융합 작제물은 다음의 구조를 갖는, 펩티드 융합 작제물: FP - 링커 - FP - 링커 - FP, D1 - 링커 - D1- 링커 - D1, 또는 FP/D1- 링커 - FP/D1- 링커- FP/D1.

[0931] 147. 구현예 146에 있어서, FP/D1 부분은 서열 번호 99의 서열을 갖는, 펩티드 융합 작제물.

[0932] 148. 구현예 146 또는 147에 있어서, 링커는 아미노산 서열 GGGGS를 포함하거나 이로 이루어지는, 펩티드 융합 작제물.

[0933] 149. 구현예 146 내지 148 중 어느 하나에 있어서, 서열번호 27 또는 53, 57의 서열을 포함하거나 이로 이루어지는, 펩티드 융합 작제물.

[0934] 150. 구현예 101 내지 131 중 어느 하나의 SARS-CoV-2 항원 또는 구현예 146 내지 149 중 어느 하나의 펩티드 융합 작제물을 포함하는, 약학적 조성물.

[0935] 151. 구현예 150에 있어서, 애쥬번트를 추가로 포함하는, 약학적 조성물.

[0936] 152. 구현예 151에 있어서, 애쥬번트는 명반, CpG, PolyI:C, MF59, AS01, AS02, AS03, AS04, AF03, 편모, ISCOM 및 ISCOMMATRIX로부터 선택되는, 약학적 조성물.

[0937] 153. 구현예 150 내지 152 중 어느 하나에 있어서, SARS-CoV-2 감염을 치료하거나 예방하는 데 사용하기 위한, 약학적 조성물.

[0938] 154. 구현예 153에 따른 용도에 있어서, 약학적 조성물은 비경구 투여되는, 약학적 조성물.

[0939] 155. 구현예 154에 따른 용도에 있어서, 약학적 조성물은 피내, 피하 또는 근육내 투여되는, 약학적 조성물.

[0940] 156. 구현예 155에 따른 용도에 있어서, 약학적 조성물은 근육내 투여되는, 약학적 조성물.

[0941] 157. 구현예 153 내지 156 중 어느 하나에 따른 용도에 있어서, 약학적 조성물은 적어도 1회 투여되는, 약학적 조성물.

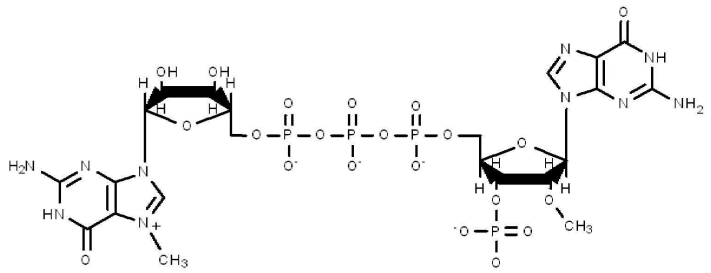
[0942] 158. 구현예 153 내지 156 중 어느 하나에 따른 용도에 있어서, 약학적 조성물은 적어도 2회 투여되는, 약학적 조성물.

[0943] 159. 구현예 158에 따른 용도에 있어서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월인, 약학

적 조성물.

[0944] 160. mRNA 작제물로서, 다음의 구조 요소:

[0945] (i) 다음의 구조를 갖는 5' 캡:

[0946] 

[0947] (ii) 서열번호 144의 핵산 서열을 갖는 5' 미번역 영역(5' UTR);

[0948] (iii) 서열번호 148의 핵산 서열을 갖는 단백질 코딩 영역;

[0949] (iv) 서열번호 145의 핵산 서열을 갖는 3' 미번역 영역(3' UTR); 및

[0950] (v) 폴리A 꼬리로 이루어지는, mRNA 작제물.

[0951] 161. 구현예 160의 mRNA 작제물을 캡슐화하는, 지질 나노입자.

[0952] 162. 구현예 161에 있어서, 지질 나노입자는 양이온성 지질, 비-양이온성 지질, 콜레스테롤계 지질 및 PEG-변형 지질을 포함하는, 지질 나노입자.

[0953] 163. 구현예 161 또는 162에 있어서, 양이온성 지질은 cKK-E12, cKK-E10, OF-Deg-Lin 및 OF-02로부터 선택되고; 비-양이온성 지질은 DOPE 및 DEPE로부터 선택되고; 콜레스테롤계 지질은 콜레스테롤이며; PEG-변형 지질은 DMG-PEG-2K인, 지질 나노입자.

[0954] 164. 구현예 160의 mRNA 작제물 또는 구현예 161 내지 163 중 어느 하나의 지질 나노입자를 포함하는, 면역원성 조성물.

[0955] 165. 구현예 164에 있어서, 10 μg 내지 200 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0956] 166. 구현예 165에 있어서, 15 μg 내지 135 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0957] 167. 구현예 166에 있어서, 적어도 20 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0958] 168. 구현예 166에 있어서, 적어도 25 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0959] 169. 구현예 166에 있어서, 적어도 35 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0960] 170. 구현예 166에 있어서, 적어도 40 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0961] 171. 구현예 166에 있어서, 적어도 45 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0962] 172. 구현예 166에 있어서, 15 μg, 45 μg 또는 135 μg의 mRNA 작제물을 포함하는, 면역원성 조성물.

[0963] 173. SARS-CoV-2 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 구현예 164 내지 172 중 어느 하나의 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

[0964] 174. 구현예 173에 있어서, 면역원성 조성물은 적어도 2회 대상체에게 투여되는, 방법.

[0965] 175. 구현예 174에 있어서, 투여 사이의 기간은 적어도 2주, 예를 들어, 3주, 또는 1개월인, 방법.

[0966] 176. SARS-CoV-2 감염 예방에 사용하기 위한, 적어도 2개의 핵산을 포함하는 면역원성 조성물로서, 여기에서 제 1 핵산은 서열번호 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 35, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84. 86, 88, 90, 92, 94, 96, 98, 104, 106, 108, 110, 118, 120, 123, 125, 127, 129, 131, 133, 135, 137, 139 또는 141로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하고, 하나 이상의 추가의 핵산(들)은 서열번호 151, 153, 155, 157, 159, 161, 163, 165, 167, 169 또는 171로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을

포함하는, 면역원성 조성물.

[0967] 177. SARS-CoV-2 감염 예방에 사용하기 위한, 적어도 2개의 핵산을 포함하는 면역원성 조성물로서, 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 상기 최적화된 뉴클레오티드 서열은 서열번호 44의 핵산 서열을 가지며,

[0968] 여기에서 하나 이상의 추가 핵산(들)은 다음으로부터 선택되는, 면역원성 조성물:

[0969] (a) 서열번호 157로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 156의 핵산 서열을 갖는, 핵산; 및

[0970] (b) 서열번호 163으로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 162의 핵산 서열을 갖는, 핵산; 및

[0971] (c) 서열번호 167로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 166의 핵산 서열을 갖는, 핵산; 및

[0972] (d) 서열번호 171로부터 선택되는 서열을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 선택적으로, 최적화된 뉴클레오티드 서열은 서열번호 170의 핵산 서열을 갖는, 핵산.

[0973] 178. 구현예 176 또는 구현예 177에 있어서, 적어도 2개의 핵산은 mRNA인, 면역원성 조성물.

[0974] 179. 구현예 178에 있어서, 제1 핵산은 서열번호 11을 포함하는 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 포함하되, 최적화된 뉴클레오티드 서열은 서열번호 148의 핵산 서열을 갖는, 면역원성 조성물.

[0975] 180. 구현예 176 내지 178에 있어서, 핵산은 지질 나노입자 중 캡슐화되는, 면역원성 조성물.

[0976] 181. SARS-CoV-2 감염을 치료하거나 예방하는 방법으로서, 상기 방법은 구현예 176 내지 179 중 어느 하나의 면역원성 조성물의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법.

[0977] **실시예**

[0978] **실시예 1. 최적화된 뉴클레오티드 서열 생성**

[0979] 본 실시예는, 시험관 내 합성 동안 전장 전사체를 수득하도록 최적화되고, 암호화된 단백질의 높은 발현 수준을 야기하는, 본 발명에 따른 최적화된 뉴클레오티드 서열을 생성하는 프로세스를 예시한다.

[0980] 프로세스는 도 1a의 코돈 최적화 방법을 도 1b에 도시된 필터링 단계의 순서와 조합하여 최적화된 뉴클레오티드 서열의 목록을 생성한다. 구체적으로, 도 1a에 도시된 바와 같이, 프로세스는 관심 아미노산 서열 및 주어진 유기체(즉, 본 실시예의 맥락에서의 인간 코돈 사용 선호도)에서의 각 코돈의 빈도를 반영하는 제1 코돈 사용 테이블을 수령한다. 그런 다음, 프로세스는 임의의 코돈이 임계 빈도(10%) 미만인 코돈 사용 빈도와 연관되는 경우, 제1 코돈 사용 테이블로부터 해당 코돈을 제거한다. 제1 단계에서 제거되지 않은 코돈의 코돈 사용 빈도는 정규화되어 정규화된 코돈 사용 테이블을 생성한다.

[0981] 코돈 사용 테이블을 정규화하는 단계는 각각의 제거된 코돈에 대한 사용 빈도 값을 재분배하는 단계를 포함하고; 제거된 특정 코돈에 대한 사용 빈도는 제거된 코돈이 아미노산을 공유하는 다른 코돈의 사용 빈도에 추가된다. 이 실시예에서, 재분배는 테이블로부터 제거되지 않은 코돈의 사용 빈도의 크기에 비례한다. 프로세스는 정규화된 코돈 사용 테이블을 사용하여 최적화된 뉴클레오티드 서열의 목록을 생성한다. 최적화된 뉴클레오티드 서열 각각은 관심 아미노산 서열을 암호화한다.

[0982] 도 1b에 도시된 바와 같이, 최적화된 뉴클레오티드 서열의 목록은 최적화된 뉴클레오티드 서열의 업데이트된 목록을 생성하기 위해, 모티프 스크린 필터, 구아닌-시토신(GC) 함량 분석 필터, 및 코돈 적용 인덱스(CAI) 분석 필터를 이 순서로 적용함으로써 추가로 처리된다.

[0983] 다음의 실시예에 예시되는 바와 같이, 이러한 프로세스는 관심 아미노산 서열을 암호화하는 최적화된 뉴클레오티드 서열을 생성한다. 뉴클레오티드 서열은 시험관 내 합성 동안 전장 전사체를 수득하고, 암호화된 단백질의 높은 수준의 발현을 야기한다(실시예 2 참조).

[0984] **실시예 2. 높은 CAI 점수를 갖는 뉴클레오티드 서열을 생성하기 위한 코돈 최적화는 단백질 수율을 개선시킴.**

[0985] 본 실시예는 약 0.8 이상의 코돈 적용 인덱스(CAI)를 갖는 코돈 최적화된 단백질 코딩 서열이 0.8 미만의 CAI를 갖는 코돈 최적화된 단백질 코딩 서열을 능가한다는 것을 입증한다.

[0986] 코돈 최적화는 인간 에리트로포이에틴(hEPO)의 야생형 아미노산 서열에 대해 수행하였다. hEPO는 낮은 세포 산소 수준(저산소증)에 반응하여 신장에 의해 분비되는 단백질 호르몬이다. hEPO는 적혈구 생성, 즉 적혈구위 생산에 필수적이다. 재조합 hEPO는 만성 신장 질환을 앓고 있는 대상체 또는 암 화학요법을 받고 있는 대상체에서 발생할 수 있는 낮은 적혈구 또는 헤모글로빈 수를 특징으로 하는 병태인 빈혈의 치료에 흔히 사용된다.

[0987] 상이한 코돈 최적화 알고리즘을 사용하여, hEPO(#1 내지 #5)를 암호화하는 총 5개의 새로운 코돈 최적화된 뉴클레오티드 서열을 생성하였다. 뉴클레오티드 서열 #4 및 #5를 도 1a 및 도 1b에 도시된 바와 같은 코돈 최적화 방법에 따라 생성하였다. 참조로서, 시험관 내 및 생체 내 모두에서 실험적으로 이전에 검증된 코돈 최적화된 hEPO 코딩 서열을 갖는 뉴클레오티드 서열이 제공되었다. 참조 뉴클레오티드 서열은 야생형 뉴클레오티드 서열 및 hEPO 단백질을 암호화하는 다른 코돈 최적화된 뉴클레오티드 서열에 비해 우월한 단백질 수율을 제공하는 것으로 밝혀졌다.

**[표 5] hEPO 암호화 뉴클레오티드 서열**

| | |
|---|---|
| **서열번호 112** | ATGGGTGTGCACGAATGTCCTGCTTGGCTGTGGCTCCTTCTCTC<br>CCTGCTGTCCCTGCCTCTTGGACTCCCGGTGCTTGGAGCACCCC<br>CGAGACTGATCTGCGACAGCAGGGTGCTCGAGCGCTACCTCCT<br>GGAAGCCAAGGAAGCCGAAAACATCACTACTGGCTGCGCCGA<br>ACACTGCTCCCTGAACGAGAACATCACCGTGCCGGACACCAAG<br>GTCAACTTCTACGCGTGGAAGAGAATGGAGGTCGGACAGCAA<br>GCCGTGGAAGTGTGGCAGGGACTTGCGCTCCTGTCGGAAGCCG<br>TGCTGAGGGGACAAGCCCTGCTCGTGAACAGCTCACAGCCTTG<br>GGAGCCCCTGCAGCTGCATGTCGACAAGGCCGTGTCCGGACTG<br>CGCTCACTGACCACTCTGCTGAGGGCCTTGGGTGCCCAGAAAG<br>AGGCTATTTCCCCACCGGATGCAGCCTCGGCAGCTCCTCTGCG<br>GACCATTACGGCGGACACCTTTCGGAAGCTGTTCCGCGTCTAC<br>AGCAATTTCCTCCGGGGGGAAGTTGAAACTGTATACCGGCGAAG<br>CCTGTCGGACTGGCGATCGCTGA |
| **서열번호 113** | ATGGGGGTTCATGAGTGCCCAGCTTGGCTTTGGCTCCTGCTCAG<br>CTTGCTTAGTCTCCCTTTGGGCCTGCCCGTGCTGGGCGCCCCTC<br>CACGCTTGATCTGTGACAGCAGGGTCTTGGAACGGTATTTGCTT<br>GAAGCTAAAGAAGCTGAGAACATAACAACGGGATGTGCTGAA<br>CATTGCTCCTTGAACGAAAACATCACAGTTCCCGACACAAAAG<br>TCAATTTTTACGCATGGAAGCGGATGGAGGTTGGCCAGCAAGC<br>TGTGGAGGTCTGGCAAGGGCTGGCTCTTCTCAGTGAAGCCGTG<br>CTGCGCGGACAAGCACTCTTGGTGAACTCCAGCCAGCCCTGGG<br>AGCCCCTTCAGCTCCATGTCGATAAAGCAGTTAGCGGCCTCCG<br>ATCATTGACTACCCTCCTTAGGGCTTTGGGTGCACAAAAAGAG<br>GCCATTTCACCACCGGACGCGGCAAGTGCTGCTCCGTTGCGAA<br>CTATAACTGCTGACACCTTCCGGAAACTTTTTCGGGTATATTCC<br>AACTTTCTCAGGGGGAAACTCAAGCTCTACACCGGCGAGGCGT<br>GCCGAACTGGAGACCGCTGA |

[0988]

[0989]

| | |
|---|---|
| 서열번호 114 | ATGGGCGTACATGAATGCCCGGCATGGCTTTGGCTGCTGCTGT<br>CCCTGCTGAGTTTGCCGCTGGGCCTCCCCGTCCTCGGCGCTCCC<br>CCGAGACTCATTTGCGACTCTAGGGTCCTCGAACGCTATCTGCT<br>GGAAGCAAAAGAAGCTGAGAACATAACTACAGGATGCGCTGA<br>GCACTGTTCCTTGAATGAGAATATCACAGTACCTGACACTAAG<br>GTGAATTTTTACGCATGGAAACGCATGGAAGTGGGTCAGCAGG<br>CCGTGGAAGTGTGGCAGGGCCTGGCGCTGCTGTCCGAGGCTGT<br>TCTTAGAGGCCAAGCCTTGTTGGTCAATTCCTCTCAACCCTGGG<br>AGCCCCTCCAGCTGCATGTTGATAAAGCCGTCTCTGGTCTCCGG<br>TCCCTTACCACCCTGCTCAGGGCACTTGGCGCACAGAAGGAAG<br>CTATCTCCCCCCCAGACGCTGCCAGTGCCGCCCCCCTCCGGACT<br>ATTACCGCCGATACTTTCAGGAAACTGTTTCGAGTCTATAGCAA<br>TTTTCTCCGCGGGAAACTGAAGCTGTATACAGGTGAGGCCTGC<br>AGGACAGGAGATCGCTGA |
| 서열번호 115 | ATGGGCGTGCACGAATGTCCTGCTTGGCTGTGGCTGCTGCTGA<br>GTCTGCTGTCTCTGCCTCTGGGACTGCCTGTTCTTGGAGCCCCT<br>CCTAGACTGATCTGCGACAGCAGAGTGCTGGAAAGATACCTGC<br>TGGAAGCCAAAGAGGCCGAGAACATCACAACAGGCTGTGCCG<br>AGCACTGCAGCCTGAACGAGAATATCACCGTGCCTGACACCAA<br>AGTGAACTTCTACGCCTGGAAGCGGATGGAAGTGGGACAGCA<br>GGCTGTGGAAGTTTGGCAAGGACTGGCCCTGCTGTCTGAAGCT<br>GTTCTGAGAGGACAGGCTCTGCTGGTCAATAGCTCTCAGCCTT<br>GGGAACCTCTCCAGCTGCATGTGGATAAGGCCGTGTCTGGCCT<br>GAGAAGCCTGACAACACTGCTGAGAGCCCTGGGAGCCCAGAA<br>AGAGGCCATTTCTCCACCTGATGCTGCCAGCGCTGCCCCTCTGA<br>GAACAATCACCGCCGACACCTTCAGAAAGCTGTTCCGGGTGTA<br>CAGCAACTTCCTGCGGGGCAAGCTGAAACTGTACACCGGCGAA<br>GCCTGCAGAACCGGCGATAGATAA |
| 서열번호 116 | ATGGGGGTGCACGAGTGCCCTGCCTGGCTGTGGTTGCTGCTGT<br>CCCTGCTGTCTCTGCCACTGGGACTGCCAGTGCTGGGAGCTCCA<br>CCTAGGCTGATCTGCGACAGCCGGGTCCTGGAGAGGTACCTGC<br>TCGAGGCCAAGGAGGCCGAGAACATTACCACAGGCTGCGCCG<br>AGCACTGCAGCCTGAACGAGAACATTACAGTGCCCGATACAAA |

| | |
|---|---|
| | GGTGAACTTCTACGCCTGGAAGAGGATGGAGGTGGGCCAGCA<br>GGCCGTGGAGGTGTGGCAGGGGCTGGCCCTGCTGAGCGAGGCC<br>GTGCTGAGGGGCCAAGCCCTGCTGGTCAACAGCAGCCAGCCTT<br>GGGAGCCCCTGCAGCTCCACGTGGACAAGGCTGTGTCTGGCTT<br>GAGGTCTCTCACAACATTGCTGAGGGCCCTGGGCGCACAGAAA<br>GAAGCTATCAGCCCACCTGATGCCGCTAGTGCCGCTCCACTGC<br>GGACAATTACCGCCGATACCTTTAGAAAATTGTTCAGGGTCTA<br>CTCCAACTTTTTGCGCGGGAAGCTGAAGCTCTATACCGGCGAG<br>GCCTGCCGGACAGGGGACAGATGA |
| 서열번호 117 | ATGGGAGTGCACGAATGTCCTGCATGGCTCTGGCTCCTGCTGTC<br>TCTCCTGAGCCTGCCACTGGGACTCCCAGTGCTGGGAGCACCC<br>CCTAGGCTGATCTGCGATTCTCGGGTGCTGGAGCGCTACCTGCT<br>CGAGGCTAAGGAGGCCGAGAATATCACTACTGGGTGTGCCGAA<br>CACTGTAGCCTCAATGAAAACATTACAGTCCCAGATACCAAGG<br>TGAACTTTTATGCATGGAAGAGGATGGAGGTCGGGCAGCAGGC<br>AGTGGAGGTGTGGCAGGGACTGGCTCTGCTGTCCGAAGCCGTG<br>CTCAGAGGTCAGGCCCTGCTGGTTAATTCCAGCCAGCCTTGGG<br>AACCTCTGCAGCTGCATGTGGACAAGGCAGTGTCTGGCCTGAG<br>ATCCCTTACTACACTGCTGAGAGCACTGGGGGCTCAGAAAGAA<br>GCTATTTCCCCACCAGACGCCGCCTCAGCAGCACCTCTCCGGA<br>CCATCACTGCTGACACCTTCCGCAAGCTCTTTAGGGTGTACTCC<br>AACTTCCTGCGCGGGAAGCTCAAGCTGTACACCGGCGAAGCCT<br>GCAGGACCGGGGATCGCTGA |

[0990]

[0991] CAI, GC 함량, 코돈 빈도 분포(CFD)뿐만 아니라 음의 CIS 요소 및 음의 반복 요소의 존재와 관련하는 5개의 뉴클레오티드 서열 각각의 특성이 표 6에 요약되어 있다.

**[표 6] hEPO를 암호화하는 최적화된 뉴클레오티드 서열의 특성**

| 뉴클레오티드 서열 | 서열번호 | CAI | GC 함량 % | CFD % | 음성 CIS 요소 | 음성 반복 요소 |
|---|---|---|---|---|---|---|
| 기준 | 서열번호 112 | 0.79 | 61.06% | 3% | 0 | 0 |
| #1 | 서열번호 113 | 0.69 | 54.12% | 2% | 0 | 0 |
| #2 | 서열번호 114 | 0.76 | 56.23% | 1% | 0 | 0 |
| #3 | 서열번호 115 | 0.90 | 57.28% | 0% | 0 | 0 |
| #4 | 서열번호 116 | 0.89 | 60.95% | 0% | 0 | 0 |
| #5 | 서열번호 117 | 0.86 | 59.56% | 0% | 0 | 0 |

[0992]

[0993] 각각의 코돈 최적화된 서열의 단백질 수율을 시험하기 위해, 동일한 3' 및 5' 비번역 서열(3' 및 5' UTR)이 측면에 위치한 hEPO 단백질을 암호화하는 6개의 뉴클레오티드 서열 중 하나를 함유하고 RNA 중합효소 프로모터가 선행하는 발현 카세트를 각각 포함하는 6개의 핵산 벡터를 제조하였다. 이들 핵산 벡터는 시험관 내 전사 반응을 위한 템플릿으로서 기능하여 6개의 코돈 최적화된 뉴클레오티드 서열(참조 및 뉴클레오티드 서열 #1 내지 #5)을 함유하는 mRNA의 6개의 배치를 제공하였다. 캡핑 및 테일링을 별도로 수행하였다. 각각의 캡핑된 mRNA 및 꼬리가 달린 mRNA를 세포주(HEK293) 내에 별도로 형질감염시켰다. 암호화된 hEPO 단백질의 발현 수준을 ELISA로 평가하였다. 본 실험의 결과를 도 2에 요약하였다.

[0994] 도 2에서 알 수 있는 바와 같이, 가장 높은 수준의 발현은 뉴클레오티드 서열 #3에서 관찰되었으며, 이는 실험

적으로 검증된 참조 뉴클레오티드 서열에 비해 거의 2배의 hEPO 단백질을 제공하였다. 보다 높은 단백질 수율에 대한 경향을 CAI에 따르는 서열에 대해 관찰할 수 있다(표 6 참조). 가장 높은 단백질 수율을 갖는 뉴클레오티드 서열 #3은 가장 높은 CAI를 가졌다. 제2 및 제3 최고 수율의 뉴클레오티드 서열 #4 및 #5는 제3 및 제4 최고 CAI를 가졌다. 또한, 최저 성능의 뉴클레오티드 서열 #1 및 #2는 최저 CAI를 가졌다. 부수적으로, 이들은 또한 GC 함량이 가장 낮은 뉴클레오티드 서열이었다. 그러나, GC 함량만은 결정적이지 않았다. 참조 뉴클레오티드 서열은 시험된 모든 코돈 최적화된 서열 중 가장 높은 GC 함량(61%)을 가졌지만, 보다 낮은 GC 함량을 갖는 뉴클레오티드 서열 #3, #4 및 #5만큼 양호하게 기능하지 않았다. 특히, 최저 성능의 뉴클레오티드 서열 #1 및 #2 또한 보다 높은 CFD를 가졌다.

[0995] 종합하면, 본 실시예의 데이터는 약 0.8 이상의 CAI를 달성하기 위한 치료적으로 관련된 뉴클레오티드 서열의 코돈 최적화가, 예를 들어, 가능한 최고 GC 함량을 갖는 뉴클레오티드 서열을 달성하기 위한 코돈 최적화보다 더 큰 단백질 수율을 야기한다는 것을 입증한다.

[0996] **실시예 3. 최적화된 핵산 작제물을 사용하는 생성된 스파이크 단백질의 검출**

[0997] 본 실시예는 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열은 배양된 세포에서 형질감염 후 높은 수준으로 성공적으로 발현된다는 것을 입증한다. 이는 또한 발현된 단백질이 기대한 바와 같이 세포에 의해 처리된다는 것을 입증한다.

[0998] 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산 작제물을 도 1a 및 도 1b에 도시된 바와 같은 코돈 최적화 방법에 따라 생성하였다. 최적화된 뉴클레오티드 서열은 표 7에 제시된다.

**[표 7] SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 핵산**

| 작제물 번호 | 최적화된 핵산 서열 | 아미노산 서열 | 단백질 설명 |
|---|---|---|---|
| A | 서열번호 29 | 서열번호 1 | 천연 전장 SARS-CoV-2 스파이크 단백질 |
| B | 서열번호 44 | 서열번호 11 | 퓨린 절단 부위를 제거하고 잔기 986 및 987 을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질 |
| C | 서열번호 43 | 서열번호 10 | 잔기 986 및 987 을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질 |
| D | 서열번호 42 | 서열번호 9 | 퓨린 절단 부위를 제거하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 SARS-CoV-2 S 단백질 |

[0999]

[1000] 배양된 세포의 형질감염을 위해, 150 μL OptiMEM 환원 혈청 배지를 0.5 μg(도 7) 또는 1 μg(도 5 및 도 6) mRNA 및 2.5 μL 리포펙타민 2000과 함께 1.5 mL 에펜도르프(Eppendorf) 튜브에 첨가하여 mRNA를 형질감염 시약에 복합체화시켰다. 각각의 튜브를 와류 상에서 부드럽게 혼합하고 마이크로 원심분리기에서 잠깐 회전시켜 내용물을 수집하였다. 복합체를 실온에서 10 ± 2분 동안 인큐베이션하였다. 그런 다음, HEK293 세포 단층(웰당 $5 \times 10^5$)을 흐트러뜨리지 않도록, 전체 복합체 부피를 12 웰 플레이트의 웰에 조심스럽게 첨가하였다. 세포를 37℃ 인큐베이터에 다시 넣고 18 ± 2시간 동안 인큐베이션한 후 수확하였다.

[1001] 배양 배지를 제거하고 250 μL의 CelLytic M(Sigma) + 1x HALT를 첨가하여 각 웰의 내용물을 수확하였다. 세포 현탁액을 얼음 위에서 20분 동안 방치하여 세포가 완전히 용해되도록 한 다음, 1.5 mL 에펜도르프 튜브에 용해물을 수집하였다. 용해물을 13,000 RPM에서 3분 동안 원심분리하여 잔류물을 펠릿화하였다. 상청액을 깨끗한 1.5 mL 에펜도르프 튜브에 옮겼다. 이 시점부터, 샘플을 항상 얼음 위에 보관하였다.

[1002] 웨스턴 블롯팅을 위해, 15 μL의 각 세포 용해물을 1X NuPage 시료 환원제가 보충된 5 μL의 4x Novex NuPAGE LDS 시료 완충액과 조합하였다. 샘플을 85℃에서 5분 동안 인큐베이션한 다음, 얼음 위에서 냉각시켰다. 전체 샘플 부피를 3 μg의 I-56578SS/겔을 갖는 Novex WedgeWell 12-웰 6% 트리스-글리신 미니 겔 내에 로딩하고 165 V에서 1 내지 1.5시간 동안 작동시켰다. PVDF 전달 팩을 갖는 TransBlot Turbo를 이송에 사용하였고, 멤브레인을 0.2% iBlock(Thermo)에서 1x PBS 중 0.05% Tween-20으로 차단하였다. 멤브레인을 차단 완충액에 명시된 대로 희석된 일차 항체(항-토끼 HRP #W401B)와 함께 ≥ 1시간 동안 인큐베이션하였다. 그런 다음, 이들을 1x TBST(Thermo)로 2회 세척하였다. 그런 다음, 차단 완충액 중에서 1:10,000으로 희석된 종에 적합한 이차 항체와 함께 멤브레인을 ≥ 1시간 동안 인큐베이션하였다. 그런 다음, 이들을 1x TBST로 4회 세척하였다. 그런 다음, SuperSignal Pico West 기질을 사용하여 멤브레인 상에 필름을 생성시켰다.

[1003] 표 7에 기술된 최적화된 뉴클레오티드 서열을 함유하는 mRNA의 형질감염은 배양된 HEK293 세포에서 단백질 발현 수준을 야기하였다. 도 5 및 도 6은 전처리된 전장 S 단백질에 상응하는 약 170 내지 180 kDa 밴드를 나타낸다. 도 5는 또한 S1 및 S2 서브유닛 밴드의 존재를 나타내며, 이는 미처리 전장 SARS-CoV-2 S 단백질(작제물 A)이 세포에 의해 올바르게 처리됨을 입증한다. 세포가 작제물 B를 발현할 때, 완전히 당화된 성숙 단백질에 상응하는 큰 밴드가 관찰되었다. 작제물 B는, 퓨린 절단 부위가 결여되고(따라서, S1 및 S2 서브유닛을 형성하도록 절단되지 않음), 프롤린을 잔기 986 및 987로서 함유하도록(따라서, 이의 융합-전 형태에서 단백질을 안정화함), 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 암호화한다.

[1004] 도 7은 또한 약 170 내지 180 kDa의 전장 S 단백질 밴드를 나타낸다. 이러한 밴드는 시험된 4개의 작제물 모두에서 관찰되었다. S1 및 S2 서브유닛 밴드를 작제물 A 및 작제물 C로 검출하였다. 작제물 C는, 프롤린을 잔기 986 및 987로서 함유하도록(따라서, 이의 융합-전 형태에서 단백질을 안정화함), 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 발현한다. 다시, 완전히 당화된 성숙 단백질을 작제물 B 및 작제물 D를 갖는 강한 밴드로서 검출하였다. 작제물 D는, 퓨린 절단 부위가 결여되도록(따라서, S1 및 S2 서브유닛을 형성하도록 절단되지 않음) 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 변이체 SARS-CoV-2 S 단백질을 암호화한다.

[1005] 본 실시예는 전장 SARS-CoV-2 S 단백질 또는 이의 변이체를 암호화하는 최적화된 핵산 서열이 높은 수준으로 발현된다는 것을 입증한다. 이는 또한 발현된 단백질이 기대한 바와 같이 세포에 의해 처리된다는 것을 입증한다.

[1006] **실시예 4. 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 서열 최적화 mRNA을 사용하는 면역화에 대한 중화 항체 반응**

[1007] 본 실시예는 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA가 마우스에서 중화 항체 반응을 유도하는 데 효과적임을 입증한다.

[1008] 실시예 3의 표 7에 기술된 최적화된 뉴클레오티드 서열을 함유하는 4개의 mRNA 각각을 지질 나노입자(LNP)에 캡슐화하였다. BALB/c 마우스군에는 4가지 제형 중 하나의 0.4 μg 투여량으로 3주 간격으로 2회 면역화하였다. 혈청 샘플에서의 결합 항체 활성을 효소-결합 면역흡착 검정(Enzyme-Linked Immunosorbent Assay, ELISA)을 통해 평가하였다. 중화 항체의 역가를 결정하기 위해, 슈도바이러스 기반 중화 검정을 사용하였다.

[1009] ELISA를 위해, 2019-nCoV 스파이크 단백질(S1 + S2) 엑토도메인(Sino Biological, Cat# 40589-V08B1)을 기질로서 사용하고, 중탄산염 완충액 중 2 μg/mL 농도로 4℃에서 밤새 코팅하였다. 플레이트를 비색 기질인 Sure Blue TMB 1-성분(SERA CARE, KPL Cat# 5120-0077)을 사용하여 현상시켰고, 정지(Stop) 용액으로 정지시켰다(SERA CARE Sure Blue, KPL Cat# 5120-0024). 각 시료에 대한 종말점 항체 역가를 배경보다 3배 더 높은 OD 값을 갖는 최고 희석으로 결정하였다.

[1010] 슈도바이러스 기반 중화 검정을 위해, 혈청 샘플을 배지(FluoroBrite 페놀 레드 유리 DMEM + 10% FBS + 10 mM HEPES + 1% PS + 1% Glutamax)에서 1:4로 희석시키고, 56℃에서 0.5시간 동안 열 불활성화시켰다. 또한, 열 불활성화 혈청의 2배 희석 시리즈를 제조하고, 리포터 바이러스 입자(RVP)-GFP(Integral Molecular)와 혼합하고, 웰 당 300개의 감염성 입자를 함유하도록 희석시키고, 37℃에서 1시간 동안 인큐베이션하였다. 75 μL 부피의 50% 융합성 293T-hsACE2 클론 세포의 96-웰 플레이트를 50 μL의 혈청/바이러스 혼합물로 접종하고 37℃에서 72시간 동안 인큐베이션하였다. 인큐베이션 종료 시, 고함량 이미저 상에서 플레이트를 스캔하고, 개별 GFP 발현 세포를 계수하였다. 억제 희석 역가(ID50)는 시험에서 바이러스 플라크의 수를 50%까지 감소시키는 희석의 역수로서 기록되었다. 각 시험 샘플에 대한 ID50은 50% 중화점 미만의 플라크 수를 갖는 최종 희석액 및 50% 중화점을 초과하는 플라크 수를 갖는 최초 희석액을 사용하여 기울기 및 절편을 계산함으로써 보간되었다. ID50 역가

= (50% 중화점 - 절편)/기울기.

[1011] 4개의 mRNA 제형 모두는 1차 백신접종 후 14일차에 유사한 수준의 결합 항체를 유도하였으며, 반응은 28일차에서의 2차 투여 후 1주차에 더욱 강화되었다. 35일차에, 슈도바이러스 중화 검정에 의해 결정된 중화 항체에 대한 기하 평균 역가(GMT)는, 작제물 A의 경우 152, 작제물 B의 경우 354, 작제물 C의 경우 195, 및 작제물 D의 경우 1005였다. 작제물 D 변이체의 중화 가능성은 작제물 B보다 약간 더 높은 경향을 나타냈다.

[1012] ELISA에서 결합에 대해 검출된 혈청학적 항체 역가는 슈도바이러스에 의해 결정된 중화 역가를 예측하지 못했다. 작제물 A 및 작제물 C 군의 일부 마우스는 중화 검정에서 혈청 변환되지 않았지만, ELISA에서의 이들의 종말점 적정 역가는 해당 군의 다른 마우스와 비슷하였다. 작제물 B 및 D는 중화 항체의 유도에 대한 면역원성에서 유사할 가능성이 높았다.

[1013] 본 실시예는 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA가 천연 전장 SARS-CoV-2 S 단백질을 암호화하는 mRNA보다 중화 항체 역가를 유도하는 데 효과적임을 입증한다. 잔기 986 및 987을 프롤린으로 돌연변이시키는 것 외에도 퓨린 절단 부위를 차단하는 것은 융합-전에서 융합-후로의 전환을 방지하기 위한 또 다른 층을 추가한다. 융합-전 형태의 중요성을 고려하여, 추가의 전임상 평가에 대해서는, (퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화하는) 작제물 B가 선택되었다.

[1014] **실시예 5. mRNA-캡슐화 지질 나노입자의 제조**

[1015] 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA를 시험관 내에서 합성하였다. RNA 중합효소 프로모터 서열에 작동가능하게 연결된 다음의 핵산 서열을 포함하는 템플릿 플라스미드를 사용하여 mRNA를 제조하였다:

[1016]
```
   1 GGACAGATCG CCTGGAGACG CCATCCACGC TGTTTTGACC TCCATAGAAG
  51 ACACCGGGAC CGATCCAGCC TCCGCGGCCG GGAACGGTGC ATTGGAACGC
 101 GGATTCCCCG TGCCAAGAGT GACTCACCGT CCTTGACACG ATGTTCGTCT
 151 TCCTCGTGCT GCTCCCACTC GTTTCTTCCC AGTGTGTCAA CCTGACAACT
 201 AGGACTCAGC TGCCACCAGC CTACACCAAC TCCTTCACCA GAGGCGTGTA
 251 TTACCCAGAC AAGGTGTTTA GAAGCAGCGT GCTGCACTCT ACCCAGGACC
 301 TCTTTCTGCC CTTTTTCAGC AACGTGACAT GGTTTCACGC AATTCACGTG
 351 TCCGGCACTA ATGGCACAAA GCGGTTCGAC AATCCAGTCC TGCCTTTCAA
 401 CGATGGCGTC TACTTTGCAT CTACTGAGAA ATCCAATATC ATTAGGGGAT
 451 GGATCTTCGG CACAACCCTG GATTCTAAGA CCCAGAGCCT GCTGATCGTC
 501 AACAACGCCA CAAACGTGGT CATTAAGGTT TGCGAGTTTC AGTTCTGTAA
 551 CGATCCTTTT CTGGGCGTGT ATTATCATAA GAACAATAAG AGCTGGATGG
 601 AGTCCGAGTT TAGAGTGTAT AGCTCTGCAA ATAATTGTAC CTTTGAGTAC
 651 GTGAGCCAGC CCTTTCTGAT GGACCTGGAG GGAAAACAAG GAAACTTCAA
 701 AAACCTGCGG GAATTCGTTT TCAAAAACAT CGACGGCTAT TTCAAGATCT
 751 ATAGCAAGCA TACCCCAATC AACCTCGTGA GGGACCTCCC CCAGGGCTTT
 801 AGCGCACTGG AGCCACTGGT TGACCTGCCT ATCGGCATTA ATATCACAAG
 851 ATTTCAGACC CTGCTGGCAC TGCATAGAAG CTATCTGACC CCTGGAGACT
 901 CCTCTAGTGG GTGGACTGCC GGCGCCGCTG CCTACTATGT GGGCTATCTG
 951 CAGCCACGGA CATTCCTGCT GAAATACAAT GAGAACGGGA CAATCACAGA
1001 TGCTGTTGAT TGCGCACTCG ACCCCCTGTC CGAGACAAAG TGCACTCTCA
1051 AGAGCTTTAC CGTCGAGAAG GGCATCTATC AGACCTCAAA CTTCAGGGTG
1101 CAGCCCACAG AATCTATCGT GCGCTTCCCT AATATCACTA ACCTGTGTCC
1151 TTTCGGTGAA GTGTTCAACG CCACCAGGTT TGCTAGCGTG TATGCCTGGA
1201 ACAGGAAGAG GATCTCTAAC TGCGTCGCCG ACTATTCCGT GCTGTATAAC
1251 AGCGCCTCCT TCTCCACATT CAAATGCTAT GGAGTGAGCC CGACAAAACT
1301 GAACGATCTC TGCTTTACAA ATGTCTACGC CGACTCTTTT GTGATCAGAG
1351 GGGACGAGGT CCGGCAGATC GCACCAGGAC AGACAGGCAA GATTGCTGAC
1401 TACAACTATA AGCTGCCTGA CGACTTCACA GGATGTGTGA TCGCATGGAA
```

1451 CTCAAACAAT CTGGACTCCA AAGTCGGGGG CAACTATAAT TACCTGTATC
1501 GCCTGTTCCG GAAGTCCAAC CTGAAGCCCT TCGAGAGGGA CATCAGTACA
1551 GAGATCTATC AGGCTGGCTC CACCCCTTGC AATGGCGTCG AAGGCTTTAA
1601 TTGTTATTTT CCCCTGCAGT CTTACGGGTT TCAGCCTACT AATGGAGTTG
1651 GGTACCAGCC ATACAGAGTG GTCGTGCTCA GCTTCGAGCT CCTGCATGCT
1701 CCAGCTACAG TTTGCGGGCC AAAGAAGTCC ACTAACCTGG TGAAGAATAA
1751 GTGCGTCAAC TTCAACTTTA ACGGGCTCAC CGGCACCGGC GTGCTGACTG
1801 AGAGCAACAA GAAGTTTCTG CCATTTCAAC AGTTTGGACG GGACATTGCC
1851 GACACCACCG ATGCCGTTCG GGATCCACAG ACCCTGGAAA TTCTGGACAT
1901 TACACCGTGC AGCTTCGGGG GCGTGAGCGT GATCACACCC GGAACCAATA
1951 CAAGCAACCA GGTTGCCGTC CTGTATCAGG ATGTCAATTG CACAGAAGTG
2001 CCAGTTGCTA TCCACGCAGA CCAGCTGACT CCCACATGGC GGGTGTATAG
2051 CACCGGATCC AACGTGTTTC AGACCCGCGC CGGATGTCTC ATTGGGGCCG
2101 AGCACGTGAA TAACAGCTAC GAGTGCGACA TCCCCATTGG CGCCGGCATT
2151 TGTGCGTCTT ACCAGACTCA GACCAACTCT CCTGGCTCCG CCTCTTCCGT
2201 TGCTAGTCAG TCTATTATTG CCTATACCAT GAGCCTCGGA GCTGAGAATA
2251 GCGTGGCCTA CTCCAATAAT TCCATCGCAA TCCCTACTAA CTTCACTATT
2301 TCTGTGACCA CCGAGATCCT GCCTGTGTCT ATGACTAAGA CTAGCGTTGA
2351 TTGTACCATG TATATTTGTG GCGACTCTAC CGAATGTTCT AACCTGCTGC
2401 TTCAGTACGG CTCATTTTGC ACACAGCTGA ACAGAGCCCT GACTGGGATC
2451 GCTGTGGAGC AGGACAAGAA CACACAGGAG GTGTTTGCAC AGGTGAAGCA
2501 GATCTATAAG ACCCCTCCTA TTAAGGATTT CGGCGGATTC AATTTCTCAC
2551 AGATTCTGCC AGACCCCAGT AAGCCTTCCA AGAGGAGCTT CATCGAGGAT
2601 CTCCTGTTTA ACAAGGTGAC CCTGGCAGAC GCCGGCTTTA TTAAGCAATA
2651 TGGGGATTGC CTGGGCGACA TTGCTGCCAG AGACCTGATT TGCGCCCAGA
2701 AATTCAATGG CCTCACAGTG CTGCCACCTC TGCTGACCGA CGAGATGATC
2751 GCTCAATACA CTAGCGCACT GCTGGCCGGA ACCATCACAT CAGGCTGGAC
2801 CTTCGGGGCC GGAGCAGCAC TGCAGATTCC ATTCGCCATG CAGATGGCCT
2851 ATAGATTCAA CGGCATTGGC GTCACACAGA ACGTGCTGTA CGAAAACCAG
2901 AAGCTCATCG CTAACCAGTT TAATTCCGCA ATTGGAAAGA TCCAAGATTC
2951 ACTCAGCTCA ACCGCCTCTG CACTCGGAAA GCTGCAGGAC GTGGTCAACC
3001 AGAATGCTCA GGCCCTGAAC ACACTCGTCA AGCAGCTGTC CTCTAACTTT
3051 GGCGCTATCA GCTCCGTTCT GAACGACATT CTGAGCCGCC TGGATCCCCC
3101 AGAGGCTGAA GTCCAGATTG ACCGCCTGAT TACCGGCCGG CTGCAGTCTC
3151 TGCAAACATA CGTGACCCAG CAGCTGATCA GAGCAGCCGA GATCCGGGCA
3201 TCCGCAAATC TGGCAGCAAC TAAGATGAGC GAATGCGTGC TGGGCCAGTC
3251 CAAGCGGGTG GACTTTTGTG GCAAGGGCTA CCACCTGATG AGCTTCCCCC
3301 AGAGCGCCCC ACATGGCGTT GTTTTTCTGC ACGTGACCTA TGTCCCTGCT
3351 CAGGAAAAGA ACTTTACAAC TGCTCCTGCT ATCTGCCATG ACGGCAAGGC
3401 CCACTTCCCA CGGGAGGGAG TGTTTGTGTC CAATGGCACA CACTGGTTCG

[1017]

3451 TGACCCAGAG GAACTTCTAT GAACCCCAGA TCATCACCAC TGACAATACC
3501 TTCGTGTCTG GAAATTGCGA CGTCGTGATC GGCATCGTTA ACAACACCGT
3551 GTACGACCCT CTCCAGCCAG AGCTGGACTC CTTTAAGGAG GAACTGGATA
3601 AGTATTTTAA GAACCACACA AGCCCAGATG TGGATCTCGG GGACATCTCC
3651 GGAATTAACG CCTCCGTGGT GAATATCCAG AAGGAGATTG ACCGCCTAAA
3701 TGAAGTTGCC AAGAACCTCA ATGAGTCTCT GATTGATCTG CAGGAACTGG
3751 GCAAGTATGA GCAGTATATC AAATGGCCCT GGTACATTTG GCTGGGGTTT
3801 ATCGCCGGAC TGATTGCCAT CGTCATGGTG ACCATCATGC TGTGTTGCAT
3851 GACCTCCTGT TGTTCCTGTC TGAAGGGCTG CTGTAGTTGC GGCTCTTGCT
3901 GTAAATTCGA CGAAGATGAT AGCGAGCCCG TGCTGAAGGG CGTGAAGCTG
3951 CATTATACCT GACGGGTGGC ATCCCTGTGA CCCCTCCCCA GTGCCTCTCC
4001 TGGCCCTGGA AGTTGCCACT CCAGTGCCCA CCAGCCTTGT CCTAATAAAA
4051 TTAAGTTGCA TCAAGCT (서열번호 149)

[1018]

[1019] 변형되지 않은 뉴클레오티드의 템플릿-의존성 RNA 합성은, 서열번호 148의 최적화된 핵산 서열을 포함하는, 서열번호 147의 핵산 서열을 갖는 폴리뉴클레오티드를 제공하였다. 다단계의 효소 촉매화 프로세스에서, 최종 mRNA 생성물이 합성되었으며, 이를 정제하여 효소 시약 및 조기 중단된 합성 생성물("단량체")을 제거하였다.

[1020] 최종 mRNA는 표 4에 나타낸 구조적 요소를 가졌다. SARS-CoV-2 S 단백질 코딩 서열의 측면에는 각각 140 및 105 뉴클레오티드의 5' 및 3' 미번역 영역(UTR)이 위치한다. mRNA는 또한 역 5'5' 삼인산 가교를 통해 5'의 제1 뉴클레오시드에 연결된 7-메틸 구아노신(m7G) 잔기로 이루어진 5' 캡 구조를 함유하였으며, 이는 자체적으로 2'O 리보스 메틸화에 의해 변형된다. 5' 캡은 리보솜에 의한 번역 개시에 필수적이다. 전체 선형 구조는 대략 100 내지 500개의 아데노신 뉴클레오시드(폴리A)의 관에 의해 3' 말단에서 종료된다. 폴리A 영역은 mRNA에 안정성을 부여하고 번역을 향상시키는 것으로도 여겨진다. 이들 구조적 요소 모두는 SARS-CoV-2 스파이크 mRNA의 효율적인 번역을 위해 필요한 자연 발생 성분이다.

[1021] 정제된 mRNA를 독점적인 양이온성 지질, 비-양이온성 지질(DOPE), 콜레스테롤계 지질(콜레스테롤) 및 PEG-변형 지질(DMG-PEG-2K)을 포함하는 지질 나노입자(LNP) 중 캡슐화하였다. 최종 mRNA-LNP 제형은 수성 현탁액이었다.

[1022] **실시예 6. 마우스에서의 중화 항체 반응의 유도**

[1023] 본 실시예는, 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물은 마우스에서의 SARS-CoV-2 S 단백질에 대한 결합 및 중화 항체의 강력한 반응을 유도함을 입증한다.

[1024] 실시예 5에서 제조된 LNP 제형을 사용하여, 0일차 및 21일차에 근육내 주사(IM)로 마우스를 2회 면역화하였다(도 9c 참조). 8마리의 6 내지 8주령 BALB/c 마우스로 이루어진 4개의 군을 각각 투여량 당 0.2 μg, 1 μg, 5 μg 또는 10 μg의 mRNA로 면역화하였다. (음성 대조군으로서 역할을 한) 제5 군의 마우스에게는 mRNA-LNP 조성물의 희석물만 투여하였다. 면역화 전 7일(제7일)에, SARS-CoV-2 S 단백질에 대한 항체의 베이스라인 수준을 결정하기 위해 각 마우스로부터 혈액 샘플을 채취하였다. 14일차, 21일차, 28일차 및 35일차에 추가 혈액 샘플을 채취하였다. 마우스 실험은 미국 국립보건원(US National Institutes of Health)의 모든 관련 규정 및 미국 펜실베니아주 덴버 소재 Covance Inc의 동물 보호 및 사용 위원회(Animal Care and Use Committee)의 승인을 준수하여 수행하였다.

[1025] ELISA 검정을 사용하여 SARS-CoV-2 S 단백질에 대한 항체 역가를 결정하였다. 96-웰 플레이트를 상업적으로 이용 가능한 SARS-CoV-2 S 단백질(SinoBio)로 코팅하고, -7일차, 14일차, 21일차, 28일차 및 35일차부터의 연속 희석된 마우스 혈청과 함께 인큐베이션하고, 결합된 총 마우스 IgG를 검출하기 위해 이차 항체로 프로브하였다.

[1026] 중화 항체의 역가를 결정하기 위해, 슈도바이러스 기반 검정을 사용하였다. 경증, 강하고 중증인 증상이 있는 COVID-19 환자로부터 유래된 39개의 개별 전환기 혈청 샘플을 양성 대조군으로 사용하였다. 혈청 샘플을 배지(FluoroBrite 페놀 레드 유리 DMEM + 10% FBS + 10 mM HEPES + 1% PS + 1% Glutamax)에서 1:4로 희석시키고, 56℃에서 30분 동안 열 불활성화시켰다. 열 불활성화된 혈청의 추가의 2배, 9-포인트 연속 희석 시리즈를 동일한 배지에서 수행하였다. 희석된 혈청 샘플을 웰당 약 300개의 감염성 입자를 함유하도록 희석된 리포터 바이러스 입자(RVP) - 녹색 형광 단백질(GFP)(Integral Molecular)의 부피와 혼합하고, 37℃에서 1시간 동안 인큐베이션하였다. 75 μL 부피의 대략 50% 융합성 293T-hsACE2 클론 세포의 96-웰 플레이트를 싱글톤 중 50 μL의 혈청/바이러스 혼합물로 접종하고 37℃에서 72시간 동안 인큐베이션하였다. 72-시간의 인큐베이션 종료 시, 고함량 이미저 상에서 플레이트를 스캔하고, 개별 GFP 발현 세포를 계수하였다. 중화 항체 역가는, 시험에서 바이러스 플라크의 수를 50%까지 감소시키는 희석의 역수로서 기록된다(도 9b 참조).

[1027] 이러한 마우스 면역화 실험의 결과를 도 9a 및 도 9b에 요약하였다. 단일 접종 후에도, 시험된 모든 투여량에 대해 14일차에서의 강력한 항체 반응을 ELISA로 관찰하였다(도 9a 참조). 제2 접종은 항체 반응의 상당한 증가를 야기하였고 중화 항체의 역가를 극적으로 개선시켰다(도 9b 참조). 1 μg, 5 μg 또는 10 μg mRNA의 2회 투여량을 투여한 결과, 35일차에 ELISA에 의해 결정했을 때 유사한 항체 역가가 나타났다. 도 9b에 나타낸 바와 같이, 0.2 μg mRNA의 2회 투여는 35일차에 중화 항체를 유도하는 데 약간 덜 효과적이었던 반면, 1 μg, 5 μg 또는 10 μg mRNA의 2회 투여량은, SARS-CoV-2로 이전에 감염된 인간 환자의 전환기 혈청에서 관찰된 중화 항체의 역가를 초과하면서, 35일차에 서로 유사한 항체 역가를 유도하였다.

[1028] 본 실시예는 본 실시예에서 시험된 면역원성 조성물은 2회 투여 후 강력한 중화 항체 반응을 유도한다는 것을 입증한다. 반응의 크기는 투여량 의존적이었다. 본 결과는 면역원성 조성물이 전환기 인간 환자에서 중화 항체 역가와 유사한 중화 항체 역가를 유도할 수 있음을 나타낸다.

[1029] **실시예 7. 마우스에서 Th1-편향 T 세포 반응의 유도**

[1030] Th1-편향 면역을 촉진하는 백신은 일반적으로 그렇지 않은 백신보다 바이러스 병원균에 대해 보다 많은 보호를 제공한다. IFN-γ와 같은 Th1 사이토카인의 분비는 바이러스에 감염된 세포의 사멸을 유도할 수 있는, T 세포의 서브 그룹인 세포독성 T 림프구(CTL)를 활성화한다. 본 실시예는 실시예 6에서 시험된 면역원성 조성물이 마우스에서 Th1-편향 T 세포 반응을 유도한다는 것을 입증한다.

[1031] 실시예 6에서 시험된 백신의 면역 반응의 품질을 추가로 평가하기 위해, 해당 실시예에 기술된 실험을 각각 5 μg 또는 10 μg mRNA로 IM 주사에 의해 마우스의 군을 2회 면역화함으로써 반복하였다. 혈액을 -4일차(베이스라인), 14일차, 21일차, 28일차 및 35일차에 샘플링하였다(도 10c 참조). 마우스 실험은 미국 국립보건원(US National Institutes of Health)의 모든 관련 규정 및 미국 펜실베니아주 덴버 소재 Covance Inc의 동물 보호 및 사용 위원회(Animal Care and Use Committee)의 승인을 준수하여 수행하였다. 35일차에 마우스를 희생시키고, 비장을 제거하였다. 단리된 비장을 균질화하고 비장세포를 후술하는 바와 같이 단리하였다. 펩티드-자극 비장세포에 의한 IFN-γ 및 IL-5 분비를 ELISPOT 검정으로 결정하였다.

[1032] 수확된 비장을 얼음 위에서 5 mL의 냉각된 배지에 보관하였다. 처리 직전, 비장을 배지를 함유하는 멸균 페트리 접시에 넣었다. 10 cc 주사기 플런저의 후면을 사용하여 비장을 균질화시켰다. 균질물을 필터에 통과시켜 멸균 튜브 내로 옮겼다. 그런 다음, 균질물을 1200 rpm으로 8 내지 10분 동안 원심분리하여 펠릿화하였다. 상등액을 부드럽게 붓고 깨끗한 종이 타월로 튜브의 가장자리를 블롯팅하였다. ACK 용해 완충액을 첨가하여 적혈구를 용해시키고, 세포를 실온에서 5분 동안 인큐베이션하였다. 튜브를 1200 rpm으로 8 내지 10분 동안 원심분리하였다. 상등액을 붓고 펠릿을 2 mM L-글루타민 CTL-시험 배지에 재현탁하였다. 현탁액을 새로운 15 mL 원뿔형 튜브 내로 여과하였다. 사용 시까지, 세포를 가습된 5% $CO_2$ 인큐베이터에서 37℃에서 유지하였다.

[1033] 시험 배지를 사용하여 PepMix™ SARS-CoV-2(Spike Glycoprotein, Cat# PM-WCPV-S-1) 펩티드 풀 1 및 펩티드 풀 2를 갖는 용액을 제조하였다. 검정에서의 각 펩티드의 최종 농도는 2 μg/ml였다. 양성 대조군으로서, 시험 배지 중 1 μg/ml의 ConA를 사용하였다. 이들 항원/미토겐 용액을 100 μL/웰에 플레이팅하였다. 항원/미토겐 용액을 함유하는 플레이트를 37℃ 인큐베이터 내에 10 내지 20분 동안 넣은 다음, 세포를 플레이팅하여 pH 및 온도가 세포에 최적이 되도록 하였다. 세포 농도를 원하는 농도로 조정하였다. 0.3 x $10^6$/100 μl/웰 비장세포를 항원/미토겐 용액을 갖는 플레이트에 첨가하였다. 완료된 후, 플레이트를 부드럽게 테이핑하고 37℃의 가습된 5% $CO_2$ 인큐베이터에 넣고 밤새 인큐베이션하였다. 플레이트를 PBS로 2회 세척한 다음, 0.05% Tween-PBS, 200 μL/웰로 2회 세척하였다.

[1034] 마우스 IFN-γ/IL-5 이중-색상 효소 ELISPOT 키트(CTL Shaker Heights, Cleveland,)를 제조사의 프로토콜에 따라 사용하였다. 검출 용액을 제조사의 지침에 따라 제조하고, 80 μL를 각 웰에 첨가하였다. 그런 다음, 플레이트를 RT에서 2시간 동안 인큐베이션하였다. 플레이트를 0.05% Tween-PBS, 200 μL/웰로 3회 세척하였다. 웰 당 80 μL로 삼차 용액을 첨가하고, 플레이트를 RT에서 30분 동안 인큐베이션하였다. 플레이트를 0.05% Tween-PBS로 2회 세척한 다음, 웰 당 200 μL의 증류소로 2회 세척하였다. 웰 당 80 μL으로 현상액을 웰에 첨가하고, RT에서 15분 동안 인큐베이션하였다. 수돗물로 멤브레인을 3회 부드럽게 헹구어 반응을 중단시켰다. 플레이트를 공기 건조시키고 CTL 분석기를 사용하여 스캔하였다. 백만 세포당 사이토카인 생성 세포의 수를 기록하였다(도 10 참조).

[1035] 도 10a에 나타낸 바와 같이, 5 μg 또는 10 μg의 mRNA로 2회 면역화된 마우스로부터 35일차에 단리된 비장세포는 다량의 Th1 사이토카인 IFN-γ를 분비하였다. 그러나, 도 10b에 나타낸 바와 같이, 이들 세포는 검출 가능한 양의 Th2 사이토카인 IL-5를 분비하지 않았다.

[1036] 본 실시예는 시험된 면역원성 조성물이 마우스에서 Th1-편향 T 세포 반응을 유도하는 데 효과적임을 입증하며, 이는 이러한 면역원성 조성물을 사용하는 백신접종이 SARS-CoV-2-감염 세포를 인식하고 제거하는 CTL 반응을 유도할 수 있음을 나타낸다.

[1037] **실시예 8. 시노몰구스 원숭이에서의 중화 항체 반응의 유도**

[1038] 본 실시예는, 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물은 시노몰구스 원숭이에서의 SARS-CoV-2 S 단백질에 대한 결합 및 중화 항체의 강력한 반응을 유도함을 입증한다.

[1039] 실시예 5에서 제조된 LNP 제형을 사용하여, 0일차 및 21일차에 IM 투여로 원숭이를 2회 면역화시켰다(도 11d 참조). 3 내지 4 연령 시노몰구스 원숭이로 이루어진 3개의 군을 투여량 당 15 μg, 45 μg 또는 135 μg의 mRNA로 각각 면역화하였다. 면역화 전 4일(제-4일)에, SARS-CoV-2 S 단백질에 대한 항체의 베이스라인 수준을 결정하기 위해 각 원숭이로부터 혈액 샘플을 채취하였다. -4일차, 2일차, 7일차, 14일차, 21일차, 23일차, 28일차 및 35일차 및 42일차에 추가 혈액 샘플을 채취하였다. 시노몰구스 원숭이 실험은 미국 국립보건원(US National Institutes of Health)의 모든 관련 규정 및 New Iberia Research Center의 동물 보호 및 사용 위원회(Animal Care and Use Committee)의 승인을 준수하여 수행하였다.

[1040] ELISA 검정을 사용하여 시노몰구스 원숭이로부터 수득한 혈액 샘플 중 SARS-CoV-2 S 단백질에 대한 항체 역가를 결정하였다. 가볍고, 강하고 심각한 증상을 가진 COVID-19 환자로부터의 39개의 개별 혈청 샘플은 양성 대조군으로서 기능한다. Nunc 마이크로웰 플레이트를 PBS 중 0.5 ug/ml로 SARS-CoV S-GCN4 단백질(Expi 293 세포주에서 발현된 GeneArt)로 4℃에서 밤새 코팅하였다. PBS-Tween 0.1%로 플레이트를 3회 세척한 후, PBS-Tween 0.1% 중 1% BSA로 1시간 동안 차단하였다. 샘플을 1:450의 초기 희석으로 플레이팅한 후, 차단 완충액에서 3배, 7-포인트의 연속 희석으로 플레이팅하였다. 실온에서 1시간 인큐베이션한 후 플레이트를 3회 세척한 다음, 50 ul의 1:5000 토끼 항-인간 IgG(Jackson Immuno Reserarch)를 각각의 웰에 첨가하였다. 플레이트를 실온에서 1시간 동안 인큐베이션하고 3회 세척하였다. Pierce 1-Step Ultra TMB-ELISA 기질 용액을 사용하여 플레이트를 6분 동안 현상하고 TMB STOP 용액으로 정지시켰다. 450 nm의 SpectraMax 플레이트 판독기에서 플레이트를 판독하였다. 0.2 OD 컷오프와 동일한 최고 희석도로 항체 역가가 기록되었다.

[1041] 슈도바이러스 기반 검정을 사용하여 시노몰구스 원숭이의 혈청 중 중화 항체의 역가를 결정하였다. 가볍고, 강하고 심각한 증상을 가진 COVID-19 환자로부터의 39개의 개별 전환기 혈청 샘플은 양성 대조군으로서 기능한다. 혈청 샘플을 배지(FluoroBrite 페놀 레드 유리 DMEM + 10% FBS + 10 mM HEPES + 1% PS + 1% Glutamax)에서 1:4로 희석시키고, 56℃에서 30분 동안 열 불활성화시켰다. 열 불활성화된 혈청의 추가의 2배, 9-포인트 연속 희석 시리즈를 배지에서 수행하였다. 희석된 혈청 샘플을 웰당 약 300개의 감염성 입자를 함유하도록 희석된 리포터 바이러스 입자(RVP) - GFP(Integral Molecular)의 부피와 혼합하고, 37℃에서 1시간 동안 인큐베이션하였다. 75 μL 부피의 대략 50% 융합성 293T-hsACE2 클론 세포의 96-웰 플레이트를 싱글톤 중 50 μL의 혈청/바이러스 혼합물로 접종하고 37℃에서 72시간 동안 인큐베이션하였다. 인큐베이션 종료 시, 고함량 이미저 상에서 플레이트를 스캔하고, 개별 GFP 발현 세포를 계수하였다. 중화 항체 역가는, 시험에서 바이러스 플라크의 수를 50%까지 감소시키는 희석의 역수로서 기록된다(도 11b 참조).

[1042] 또한, 39개의 인간 전환기 혈청을 양성 대조군으로 사용하여 각 원숭이 샘플의 마이크로중화 역가를 결정하였다. Vero E6 세포를 96-웰 평평한 바닥 세포 배양 플레이트에, 사용 1일 전 웰 당 0.1 mL 중 2 x $10^4$개의 세포 농도로 씨딩하였다. 실험 당일, 1:10 희석에서 시작하여, 열 불활성화 원숭이 또는 인간 혈청의 2배 연속 희석물을 37℃ 인큐베이터에서 60 ± 5분 동안 SARS-CoV-2 바이러스(예를 들어, USA-WA1/2020 [BEI Resources; 카탈로그# NR-52281])와 함께 인큐베이션하였다. 성장 배지를 Vero E6 세포로부터 무균적으로 제거하고, 시험 샘플(혈청 및 바이러스)을 Vero E6-씨딩된 플레이트에 첨가하고, 37℃ 인큐베이터에서 30 ± 5분 동안 배양하였다. 후속하여, 100 μL의 성장 배지를 기존의 접종물을 제거하지 않고 모든 플레이트의 모든 웰에 첨가하였다. 그런 다음, 플레이트를 인큐베이터에 다시 넣고 2일 동안 인큐베이션하였다. 감염 후 2일차에, 세포를 고정시키고, 1차 항체(SARS-CoV 항-핵단백질 마우스 단클론 항체(SinoBio 카탈로그# 40143-MM05 또는 이의 등가물)로 염색한 다음, HRP-태그된 2차 항체(고추냉이 과산화효소(HRP)-접합 염소 항-마우스 면역글로불린 G (IgG) 항체(Jackson ImmunoResearch Laboratories, 카탈로그# 115-035-062 또는 이의 등가물)로 염색하였다.

[1043] 도 11은 이들 검정의 결과를 요약한다. 시험된 15 μg의 최저 mRNA 투여량에서도, 2회 접종 후 강력한 결합 및 중화 항체 반응이 관찰되었다(도 11a 및 도 11b 참조). 15 μg, 45 μg 또는 135 μg mRNA의 2회 투여량을 투여한 결과, 28, 35 및 42일차에 ELISA에 의해 결정했을 때 유사한 항체 역가가 나타났다(도 11a 참조). 15 μg 또는 45 μg mRNA의 2회 접종은 또한 이들 일차에 유사한 수준의 중화 항체를 생성하였다(도 11b 참조). 135 μg mRNA의 2회 접종은, 28, 35 및 42일차에 SARS-CoV-2로 감염된 인간 환자의 전환기 혈청에서 관찰된 중화 항체의 역가를 초과한, 항체의 역가를 유도하였다. 마이크로중화 역가 검정은 유사한 역가를 초래하는 15 μg 및 45 μg mRNA 투여량, 및 SARS-CoV-2로 감염된 인간 환자의 전환기 혈청에서 관찰된 역가를 초과하는 135 μg 투여량에서 유사한 결과를 제공하였다(도 11c 참조).

[1044] 본 실시예는, 투여 사이의 간격이 적어도 2주(특히 약 3주)인 경우, 2회 접종 후 15 μg의 최저 투여량에서도

시험된 면역원성 조성물이 강력한 중화 항체 반응을 유도한다는 것을 입증한다. 데이터는 보호 중화 항체 반응을 유도하기 위한 인간 환자에서의 본 시험 조성물의 사용을 뒷받침한다.

[1045] **실시예 9. 시노몰구스 원숭이에서 Th1-편향 T 세포 반응의 유도**

[1046] 본 실시예는 실시예 8에서 시험된 면역원성 조성물이 시노몰구스 원숭이에서 Th1-편향 T 세포 반응을 유도한다는 것을 입증한다.

[1047] 실시예 8에서 시험된 백신의 면역 반응의 품질을 추가로 평가하기 위해, PBMC를 시노몰구스 혈액 샘플로서 단리하였다. 단리된 PBMC를 극저온 바이알에 보관하였다. ELISPOT 검정을 사용하여, 펩티드 자극 PBMC에 의한 IFN-γ 및 IL-13 분비를 결정함으로써 T 세포 반응을 평가하였다. 미처리 PBMC는 비활성화, 비자극 세포에서의 IFN-γ 또는 IL-13 분비의 베이스라인 수준을 확립하기 위한 대조군으로서 기능하였다. 도 12는 결과를 요약한다.

[1048] 검정을 수행하기 위해, 원숭이 PBMC(DMEM1640+10% 열-불활성화 FCS)에 대해 완전 배지를 37℃ 수조에서 예열하였다. PBMC 극저온 바이알을 37℃ 수조에서 신속하게 해동하고, 그 내용물을 원뿔형 튜브에서 예열된 배지 내로 천천히 적가하였다. 그런 다음, 튜브를 1500 RPM으로 5분 동안 원심분리하였다. 세포 펠릿을 예열된 완전 배지로 1회 세척하고, 1500 RPM으로 15분 동안 재-펠렛화하였다. 상청액을 버리고, PBMC를 완전 배지로 재현탁하고, Guava 세포 계수기를 사용하여 계수하였다.

[1049] 원숭이 IFN-γ ELISPOT 키트(CTL, cat# 3421M-4APW) 및 IL-13 ELISPOT 키트(CTL, cat# 3470M-4APW)를 사용하여 펩티드-자극 PBMC에 의한 IFN-γ 및 IL-13 분비의 수준을 결정하였다. 키트와 함께 제공된 사전 코팅된 플레이트를 멸균 PBS로 4회 세척한 다음, 200 μl/웰 완전 배지로 차단하였다. 37℃ 인큐베이터 중 적어도 30분 동안 차단 단계를 수행하였다. 펩티드당 2 μg/ml의 최종 농도의 PepMix™ SARS-CoV-2(JPT Cat# PM-WCPV-S-1) 펩티드 풀 1 및 풀 2를 검정에서의 회수 항원으로서 사용하였다. 2 μg/ml의 Concanavalin A(Sigma, cat#C5275)를 양성 대조군으로서 사용하였다. 50 μl의 회수 항원 및 50 μl 중 300,000개의 PBMC를 자극을 위한 각각의 웰에 첨가하였다. 그런 다음, 플레이트를 37℃, 5% $CO_2$의 가습된 인큐베이터에 24시간 동안 두었다. 24시간 인큐베이션 후, 플레이트를 PBS로 5회 세척하였다. 1% 송아지 태아 혈청을 함유하는 PBS에서 제조된 100 μl/웰의 비오틴화 항-IFN-γ 또는 항-IL-13 검출 항체(1 μg/ml)를 첨가하고, 플레이트를 실온에서 2시간 동안 인큐베이션하였다. 그런 다음, 플레이트를 전과 같이 PBS로 5회 세척하고, 1% 송아지 태아 혈청을 함유하는 PBS 중 1:1000로 희석된 100 μl/웰의 스트렙타비딘과 함께 실온에서 1시간 동안 인큐베이션하였다. 플레이트를 PBS로 다시 5회 세척하고, 반점이 보일 때까지 100 μl/웰 BCIP/NBT 기질 용액을 사용하여 현상하였다. 플레이트를 수돗물로 세척하여 발색을 중단시켰다. 그런 다음, 플레이트를 밤새 건조시키고, 스캔하고, CTL 분석기를 사용하여 반점을 계수하였다. 데이터는 백만 PBMC당 반점 형성 세포(SFC)로서 기록된다(도 12 참조).

[1050] 도 12a(펩티드 풀 S1) 및 도 12c(펩티드 풀 S2)에서 나타내는 바와 같이, 15 μg, 45 μg 또는 135 μg mRNA의 2회 접종으로 면역화된 원숭이로부터 42일차에 단리된 PBMC는 SARS-CoV-2 S 단백질로부터 유래된 펩티드를 사용하는 자극에 반응하여 다량의 Th1 사이토카인 IFN-γ를 분비하였다. 대조적으로, 이들 세포는 펩티드 자극에 반응하여 베이스라인 양만의 Th2 사이토카인 IL-13을 분비하였다(도 12b(펩티드 풀 S1) 및 도 12d(펩티드 풀 S2) 참조).

[1051] 본 실시예는 시험된 면역원성 조성물이 시노몰구스 원숭이에서 Th1-편향 T 세포 반응을 유도하는 데 효과적임을 입증하며, 이는 이러한 면역원성 조성물을 사용하는 백신접종이 SARS-CoV-2-감염 세포를 인식하고 제거하는, 인간에서의 CTL 반응을 유도할 수 있음을 나타낸다.

[1052] **실시예 10. 투여량 모델링**

[1053] 본 실시예는 실시예 6 및 8에서 시험된 면역원성 조성물의 낮은 mRNA 투여량은 COVID-19 환자로부터의 전환기 혈청의 대조군 패널에서 관찰된 상응하는 역가보다 유의하게 더 높은 중화 항체 역가를 수득하는 데 효과적임을 입증한다.

[1054] 실시예 6에 기술된 면역화된 마우스의 1 μg, 5 μg 및 10 μg 군 사이에서, 35일차의 슈도바이러스 중화 역가는 통계적으로 유의한 차이가 없었으며, 이는 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 시험된 최적화된 뉴클레오티드 서열을 포함하는 1 μg 이상의 mRNA의 투여량 포화 효과를 시사한다. 마우스에서의 35일차 피크 슈도바이러스 중화 역가는 COVID-19 환자의 전환기 혈청의 대조군 패널에서 관찰된 상응하는 역가보다 유의하게 더 높았다(도 13a 참조).

[1055] 실시예 8에 기술된 시노몰구스 원숭이 실험에 대한 슈도바이러스 중화 검정 및 마이크로중화 검정 둘 모두의 결

과는 매우 높은 상관 관계를 가졌다(도 13b). 투여량 수준에 상관없이, 35일차 슈도바이러스 및 마이크로중화 역가는 면역 전 동물의 역가보다 약 130배 더 높았다. COVID-19 환자로부터의 93개의 전환기 혈청을 갖는 완전한 데이터 세트에 대한 추가 통계 분석에서, 각각 15 μg, 45 μg 및 135 μg의 mRNA 투여량으로 수득된 역가는 전환기 인간 혈청에서 관찰된 상응하는 역가보다 유의하게 더 높은 것으로 밝혀졌다(모든 P 값은 0.005 미만임; 도 13c 및 도 13d).

[1056] 본 실시예는 실시예 5에서 제조된 면역원성 조성물의 안전성 및 효능을 조사하는 인간 임상 시험을 위한 10 μg 내지 200 μg의 mRNA 투여량 범위를 뒷받침한다. 실제로, 15 μg 내지 45 μg의 투여량은 효과적인 중화 항체 반응을 유도하는 동시에 양호한 내약성을 유도하기에 충분할 수 있다.

[1057] **실시예 11. 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 mRNA의 면역원성.**

[1058] 본 실시예는, 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이(2P/GSAS)시키도록 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화하는 mRNA는 다른 전장 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 mRNA보다 중화 항체 반응을 유도하는 데 보다 효과적임을 입증한다.

[1059] 면역원성에 대한 융합-전 형태의 SARS-CoV-2 S 단백질을 안정화하는 돌연변이의 영향을 결정하기 위해, 7개의 mRNA 작제물 - 야생형 SARS-CoV-2 S 단백질(WT) 및 상응하는 융합-전 안정화된 SARS-CoV-2 S 단백질(각각 2P, GSAS, 2P/GSAS, 2P/GSAS/ALAYT, 6P 및 6P/GSAS, respectively)을 실시예 5에 기술된 바와 같이 mRNA 백신으로서 지질 입자(LNP) 중에 제형화하였다. WT, 2P/GSAS, 2P, GSAS는 실시예 3에서의 작제물 A 내지 D에 각각 상응한다. 2P/GSAS/KLHYT는 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 대체하고, ER 회수 신호를 돌연변이시키도록 돌연변이된 SARS-CoV-2 S 단백질이며, 이는 서열번호 124의 최적화된 핵산 서열 및 서열번호 125의 아미노산 서열을 갖는다. 6P는 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 대체하도록 돌연변이된 SARS-CoV-2 S 단백질이며, 이는 서열번호 128의 최적화된 핵산 서열 및 서열번호 129의 아미노산 서열을 갖는다. 6P/GSAS는 퓨린 절단 부위를 제거하고 잔기 817, 892, 899, 942, 986 및 987을 프롤린으로 대체하도록 돌연변이된 SARS-CoV-2 S 단백질이며, 이는 서열번호 130의 최적화된 핵산 서열 및 서열번호 131의 아미노산 서열을 갖는다.

[1060] 면역 평가를 위해 2개의 동물 모델을 사용하였다. 각각의 5가지 제형(WT, 2P, GSAS, 2P/GSAS, 2P/GSAS/ALAYT)의 투여량 당 0.4 μg으로 3주 간격으로 2회의 면역화를 BALB/c 마우스에게 수행하였다. 이와 병행하여, 6가지의 S mRNA 백신(2P, GSAS, 2P/GSAS, 2P/GSAS/ALAYT, 6P 및 6P/GSAS)의 투여량 당 5 μg으로 동일한 면역화 일정을 사용하여 비인간 영장류(NHP)를 면역화하였다.

[1061] 기능적 항체, 예를 들어 nAb 역가를 평가하기 위해, 인간 ACE2를 안정적으로 과발현하는 HEK-293T 세포에서의 GFP 리포터 슈도바이러스 입자(RVP)의 감염성을 중화하는 면역 혈청의 능력을 시험하였다. SARS CoV-2 S 단백질을 발현하는 RVP는, 진입 시 GFP 발현에 의해 표시되는, 단일 라운드 감염이 가능하다. 중화 효능은 RVP 진입의 50% 억제를 달성할 수 있는 혈청 희석으로 결정하였다(ID50). 또한, GCN4 나선 다발에 의해 항원으로서 삼량체화된 재조합 가용성 S-단백질을 사용하여 효소-결합 면역흡착 검정(ELISA) 역가를 평가하였다.

[1062] 1차 면역화 후 14일차에 몇몇 동물에게 중화 역가가 발생하였지만, 역가는 일반적으로 낮았다. 예상대로, 대부분의 시험 동물은 제2 면역화 후 중화 역가를 나타냈다(도 14). 35일차에, 마우스에서의 슈도바이러스(PsV) nAb 역가에 대한 95% 신뢰 구간(95% CI)을 갖는 기하 평균 역가(GMT)는 야생형(WT)의 경우 152(36; 645), 2P의 경우 195(44; 870), GSAS의 경우 1005(261; 3877), 2P/GSAS의 경우 354(129; 976), 및 2P/GSAS/ALAYT의 경우 940이었다. 특히 35일차 및 42일차에, WT 및 2P 작제물의 것과 비교했을 때 GSAS 돌연변이를 갖는 3개의 작제물의 경우 보다 높은 GMT에 대한 경향이 있었다.

[1063] NHP에서, 제2 면역화 후에도 각 군 내에서 다양한 중화 역가가 관찰되었다(도 14). 2P 및 6P/GSAS 백신은 35일차에 각각 78 및 10의 GMT를 갖는 다른 작제물보다 낮은 면역원성을 나타냈다. 6P 백신은 임의의 검출 가능한 중화 역가를 유도하지 못했다. 마우스 연구에서의 관찰과 일관되게, 6P/GSAS를 제외한 모든 GSAS 작제물은 2차 투여 후 보다 높은 중화 역가를 유도하였으며, D35에서의 GMT(95% CI)는 2P 백신 군의 것과 비교 시, GSAS의 경우 425(48; 3769), 2P/GSAS의 경우 772(116; 5121), 2P/GSAS/ALAYT의 경우 280(11; 6970)으로 기록되었다. 마우스 및 NHP 둘 모두에서의 GMT 경향은 다른 작제물 대비 2P/GSAS의 우월한 면역원성을 시사한다. 또한, 마우스 및 NHP에서 2P/GSAS 변이체에 대한 피크 PsVNa 역가(35일차)는 COVID-19 환자로부터의 93개의 전환기 혈청 패널에서 관찰된 역가와 비슷하거나 보다 높았다.

[1064] 본 실시예는 GSAS 돌연변이가 백신 면역원성에 유익함을 입증한다. S 단백질의 융합-전 형태의 안정화를 위해

도입된 2P 돌연변이는 GSAS 돌연변이의 맥락에서 유익한 것으로 나타난 반면, ALAYT는 2P/GSAS의 맥락에서, 특히 NHP에서 보다 적은 면역원성에 대한 영향을 나타냈다. 따라서, 본 실시예는, 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 mRNA는 다른 융합-전 안정화된 SARS-CoV-2 S 단백질을 암호화하는 mRNA보다 중화 항체 반응을 유도하는 데 보다 효과적일 수 있다는 확증을 추가로 제공한다.

[1065] **실시예 12. 시리아 골든 햄스터에서의 보호 효능**

[1066] 본 실시예는, 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키도록 자연 발생 SARS-CoV-2 S 단백질에 대해 변형된 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화된 mRNA의 면역원성 조성물은 폐의 바이러스 감염을 감소시키고 폐 병리학을 방지함으로써, COVID-19의 동물 모델에서 보호 효능을 가질 수 있음을 입증한다.

[1067] 시리아 골든 햄스터에서의 SARS-CoV-2 감염은 병리학 모델이며, 여기에서 바이러스 감염은 감염 후 2일차(DPI)에 폐 및 비강 상피에서 피크 역가를 갖는 높은 수준의 바이러스 복제, 7 DPI에서의 폐 질환의 조직병리학적 증거, 및 7 DPI 부근에서의 약 8 내지 15%의 체중 감소와 연관된다.

[1068] 바이러스 감염 및 질환으로부터의 보호에 대한 실시예 5에서 제조된 LNP 제형의 가능성을 평가하기 위해, 시리아 골든 햄스터를 D21에서의 단일 IM 면역화, 또는 0일차 및 21일차에서의 2회의 IM 투여에 대해, 투여량 당 0.15, 1.5, 4.5 또는 13.5 μg의 4가지 백신 제형 투여량 수준으로 면역화하였다. 49일차에 SARS-CoV-2의 비강내(IN) 접종을 통해 동물을 접종하고, 8 DPI에서의 체중 감소로서 질환의 임상 증상에 대해 모니터링하였다. 조직병리학, 및 하위 게놈 RNA RT-PCR 분석에 의한 바이러스 복제의 정량화를 위해 폐 및 비강 조직을 4 또는 7 DPI에 채취하였다.

[1069] 실시예 5의 LNP 제형은 제1 백신접종 후 강력한 투여량 의존적 중화 항체 반응을 유도하였으며, 이는 제2 면역화에 의해 상당히 향상되었다. 1차 면역화 후, 0.15 μg 투여량 군을 제외한 모든 동물은 야생형 SARS-CoV-2 바이러스에 대한 플라크 감소 중화 역가(PRNT)로서 기록된 중화 항체를 발생시켰다. 단일 투여 면역화 스케줄에 대한 35일차 PRNT50 GMT는 1.5, 4.5 및 13.5 μg 투여량에 대해 각각 237, 410 및 711인 반면, 2회 투여 군에 대한 상응하는 값은 3219, 2446 및 3219였다. 투여량이 증가함에 따라 역가가 더 높아지는 것으로 관찰되는 경향에도 불구하고, 1.5, 4.5 및 13.5 μg 군에서의 역가 간의 차이는 통계적으로 유의하지 않았다.

[1070] 백신접종의 보호 효과를 시험하기 위해, 모든 군을 비강내 접종하였다. 각 동물에 대한 체중을 7일 동안 매일 모니터링하였다(도 15a). 샴(희석제) 백신 접종 동물은 7 DPI에서 10%를 초과하는 감소값을 갖는 가장 유의한 체중 감소를 보였다. 1.5, 4.5, 및 13.5 μg의 백신접종 요법은, 1회 투여 또는 2회 투여 요법에 관계없이, 동물을 체중 감소로부터 보호하였고, 대부분의 동물은 5% 미만의 체중 감소를 경험하였으며, 손실은 대부분 2 내지 3 DPI 부근에서 최고치에 달하였다. 이들 군 간의 체중 비교에 있어서 유의한 차이는 없었다. 샴 군과 비교하여, 체중 감소의 정도가 유사한 유일한 군은 단일 면역화가 이루어진 0.15 μg 투여량 군이었다.

[1071] 바이러스 감염에 의해 야기된 병리학을 평가하기 위해, 4 또는 7개의 DPI에서 각 군의 동물 4마리로부터 폐 샘플을 채취하고, 고정된 조직을 절단하고, 조직병리학적 검사를 위해 무작위 배정하고 맹검화하였다. 조직 손상의 중증도에 기초하여, 보다 높은 점수는 보다 중증인 병리학을 반영하는, 0 내지 3의 병리학 점수를 각 샘플에 할당하였다. 폐 절편의 25% 미만에서 조직병리학적 소견이 드러나는 경우 1점을 부여하였다. 유사하게, 유조직(parenchyma)의 25% 초과 50% 미만이 관련된 경우, 2점을 부여하였다. 전체 섹션의 50% 이상이 영향을 받은 섹션에는 3점을 부여하였다. SARS-CoV-2가 접종된 샴 백신 접종 햄스터는 COVID-19 환자에서 검출된 중증 폐렴의 보고와 유사한 광범위한 폐 조직병리학을 나타냈다(도 15b). 미처리 햄스터의 폐는 조직학적으로 특이하지 않았다. 단일 백신접종의 0.15 μg 투여량 군으로부터의 폐 샘플에서 유사한 병변이 관찰될 수 있었으며, 이는 맹검 검사에서 3점을 부여받았다. 반대로, 단일 백신접종의 13.5 μg 투여량 군으로부터의 폐 샘플은 건강한 대조군과 유사하게 이러한 병변을 나타내지 않았으며, 둘 모두 0점을 부여받았다(도 15c).

[1072] 폐 병리학은 실시예 5의 LNP 제형을 1회 또는 2회 투여받은 햄스터에서 현저하게 약화되었고, 4 및 7 DPI 모두에서 투여량 의존적 효과가 있는 것으로 나타났다(도 15b). 감염에 의해 야기된 1.5, 4.5 및 13.5 μg의 단일 백신접종은 실질적으로 병리학을 약화시켰지만, 1.5, 4.5 및 13.5 μg의 2회 투여 백신접종은 거의 완전한 병리학 예방을 제공하였다. 0.15 μg의 매우 낮은 투여량 수준은 단일 투여량 처방에서 사용 시 보호를 나타내지 않았지만, 2회 투여의 백신접종 처방에서는 약간의 미미한 보호를 나타냈다.

[1073] 실시예 5의 LNP 제형을 사용하는 면역화가 햄스터에서 바이러스 감염에 영향을 미칠 수 있는지 여부를 평가하기

위해, RT PCR에 의한 폐 및 비강 샘플로부터의 바이러스 서브게놈 mRNA(sgRNA)를 측정하였다. 해당 군의 절반(n = 4)의 폐 및 비강 샘플을 4 또는 7 DPI에서 채취하고, RT-PCR에 의한 sgRNA의 검출을 위해 총 RNA를 처리하였다(도 15d). 4 및 7 DPI에 채취된 폐 샘플의 경우, 샴 백신접종된 군은 각각 약 108 및 105의 그램 조직 당 카피를 제공한 반면, 13.5 μg 2회 투여 요법을 받은 조직은 두 시점 모두에서 검출 수준 미만이었다. 1.5 μg 및 4.5 μg 2회 투여 요법을 받은 이들로부터의 폐 샘플은 4 DPI에서의 바이러스 sgRNA 카피에서 거의 3 로그 감소를 가졌고 7 DPI에서는 검출 수준 미만이었다. 1.5, 4.5 및 13.5 μg의 단일-투여 백신접종을 받은 이들로부터의 폐 샘플의 경우, 4 DPI에서의 바이러스 부하는 샴 백신접종 군의 바이러스 부하와 유사한 반면, 7 DPI에서의 부하는 검출 임계값 미만이었다. 특히, 0.15 μg의 1회 투여 또는 2회 투여 요법을 받은 폐 샘플은 4 또는 7 DPI에서 샴 백신접종된 군의 것과 비교하여 유사하거나 심지어 보다 높은 바이러스 부하를 가졌다. 그러나, 바이러스 부하(sgRNA)는 모든 군에서 4 DPI에서 보다 다양했으며, 대부분의 군에서 1마리 또는 2마리의 동물이 음성이었다. 7 DPI에서의 비강 샘플에서 바이러스 sgRNA의 제거를 달성한 유일한 군은 13.5 μg의 2회 투여 백신접종 군이었다.

[1074] 본 실시예는 실시예 5에서 제조된 면역원성 조성물이 COVID-19의 동물 모델에서 폐의 바이러스 감염을 감소시키고 폐 병리학을 예방할 수 있음을 입증한다. 실시예 5에서 제조된 면역원성 조성물을 사용하는 면역화는 기간이 단축되고 상부 호흡기로부터 바이러스가 더 많이 배출되기 때문에 전염에 영향을 미칠 수 있다.

[1075] **실시예 13. mRNA-캡슐화 지질 나노입자의 제조**

[1076] 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고, L18F, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 2 + D614G)를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA를 시험관 내에서 합성하였다. RNA 중합효소 프로모터 서열에 작동가능하게 연결된 서열번호 166을 포함하는 템플릿 플라스미드를 사용하여 mRNA를 제조하였다.

[1077] 변형되지 않은 뉴클레오티드의 템플릿-의존성 RNA 합성은, 서열번호 173의 최적화된 핵산 서열을 포함하는, 서열번호 172의 핵산 서열을 갖는 폴리뉴클레오티드를 제공하였다. 다단계의 효소 촉매화 프로세스에서, 최종 mRNA 생성물이 합성되었으며, 이를 정제하여 효소 시약 및 조기 중단된 합성 생성물("단량체")을 제거하였다.

[1078] 최종 mRNA는 표 4의 mRNA 작제물 2에 나타낸 구조적 요소를 가졌다. SARS-CoV-2 S 단백질 코딩 서열의 측면에는 각각 140 및 105 뉴클레오티드의 5' 및 3' 미번역 영역(UTR)이 위치한다. mRNA는 또한 역 5'5' 삼인산 가교를 통해 5'의 제1 뉴클레오시드에 연결된 7-메틸 구아노신(m7G) 잔기로 이루어진 5' 캡 구조를 함유하였으며, 이는 자체적으로 2'O리보스 메틸화에 의해 변형된다. 5' 캡은 리보솜에 의한 번역 개시에 필수적이다. 전체 선형 구조는 대략 100 내지 500개의 아데노신 뉴클레오시드(폴리A)의 관에 의해 3' 말단에서 종료된다. 폴리A 영역은 mRNA에 안정성을 부여하고 번역을 향상시키는 것으로도 여겨진다. 이들 구조적 요소 모두는 SARS-CoV-2 스파이크 mRNA의 효율적인 번역을 위해 필요한 자연 발생 성분이다.

[1079] 몰비로 40% cKK-E10, 30% DOPE, 28.5% 콜레스테롤 및 1.5% DMG-PEG-2K를 포함하는 지질 나노입자(LNP)에 정제된 mRNA를 캡슐화하였다. 최종 mRNA-LNP 제형은 수성 현탁액이었다.

[1080] **실시예 14. SARS-CoV-2의 변이체 계열에 대해 효과적인 중화 항체 반응**

[1081] 본 실시예는, 이전에 실시예 5의 LNP 제형의 2회 투여로 면역화된 비인간 영장류(NHP)는, 원래의 우한 계열로부터 유래된 SARS-CoV-2 S 뿐만 아니라, 남아공, 일본/브라질 및 캘리포니아에서 관찰된 SARS-CoV-2 S 단백질의 자연 발생 변이체, 및 퓨린 절단 부위를 제거하고 잔기 986 및 987을 프롤린으로 돌연변이시키고, SARS-CoV-2의 남아공 변이체의 L18F, D80A, D215G, ΔL242, ΔA243, ΔL244, K417N, E484K, N501Y, D614G 및 A701V 돌연변이(남아공 변이체 2 + D614G)를 함유하도록 자연 발생 SARS-CoV-2 스파이크 단백질에 대해 변형된 전장 SARS-CoV-2 S 단백질을 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 LNP-캡슐화 mRNA의 면역원성 조성물을 사용한 노출에 대한 반응으로 SARS-CoV-1로부터 유래된 S 단백질에 대한 효과적인 중화 항체 반응을 마운트한다. 실시예 13에 기술된 바와 같이 면역원성 조성물을 제조하였다.

[1082] SARS-CoV-2의 다양한 순환성 변이체 및 SARS-CoV-1에서 유래된 S 단백질을 표적화하는 광범위한 면역 반응을 유도하기 위해. 비인간 영장류(NHP) 모델(시노몰구스 원숭이)을 사용하여, 실시예 5에 기술된 (SARS-CoV-2 단백질의 융합-전 안정화된 우한 변이체를 암호화하는) mRNA 백신에 의해 유도된 원래의 우한 계열에 대한 원래의 항원 특이성이, 단독으로, 또는 실시예 5의 mRNA 백신(우한)과 조합하여, SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공(SA) 변이체를 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA 백신을 사용하는 후속적인 면

역화에 의해 극복될 수 있는지의 여부를 조사하였다. 시노몰구스 원숭이(n = 4)를 실시예 5에서 제조된 LNP 제형의 각각 15 μg, 45 μg 또는 135 μg으로 3주 간격(0일차 및 21일차)으로 2회 면역화시켰다. 315일차에 동물을 무작위 배정한 뒤, 2개의 군으로 나누고 면역화시켰다. 1군을 실시예 13에 기술된, SARS-CoV-2의 남아공 변이체(SA 단독)로부터 유래된 돌연변이를 함유하는 mRNA 백신으로 면역화시켰다. 2군은 실시예 5로부터의 원래 mRNA 백신을 함유하는 제형 + 1군(우한 + SA)에 주어진 변이체로 면역화시켰다. 1군 및 2군 둘 모두 10 μg의 총 mRNA 투여량을 투여받았다. 연구는 항원 반응을 넓히기에 2가 면역원성 조성물(우한 + SA)이 필요한지의 여부, 또는 비-우한 변이체(SA 단독)로부터 유래된 SARS-CoV-2 S 단백질을 포함하는 1가 면역원성 조성물이 항원 반응을 넓히기에 충분한지의 여부를 평가하기 위해 설계되었다.

[1083] 면역화-전 및 부스트-전(4일차, 308일차) 동물로부터의 혈청 샘플뿐만 아니라, 14일차, 21일차, 28일차, 35일차, 42일차, 90일차, 308일차 및 329일차에 채취된 샘플을 우한 S-단백질-발현 슈도바이러스(PsV) 중화 검정으로 시험하였다. 35일차, 308일차 및 329일차에 채취한 혈청 샘플을 슈도바이러스(PsV) 중화 검정으로 시험하였다. 도 16에 도시된 바와 같이, 시험된 PsV는 SARS-CoV-2 계열 우한, 남아공(SA 20C 및 SA 20H), 일본/브라질(Jap/Braz) 또는 캘리포니아로부터 유래된 S 단백질, 또는 SARS-CoV-1 계열로부터 유래된 S 단백질을 발현하였다. 혈청 샘플을 배지(FluoroBrite 페놀 레드 유리 DMEM + 10% FBS + 10 mM HEPES + 1% PS + 1% Glutamax)에서 희석시키고, 56℃에서 30분 동안 열 불활성화시켰다. 열 불활성화된 혈청의 추가의 2배, 11-포인트 연속 희석 시리즈를 배지에서 수행하였다. 희석된 혈청 샘플을 웰 당 대략 300개의 감염성 입자를 함유하도록 리포터 바이러스 입자(RVP)-GFP(Integral Molecular)와 혼합하고, 37℃에서 1시간 동안 인큐베이션하였다. 75 μL 부피의 대략 50% 융합성 293T-hsACE2 클론 세포의 96-웰 플레이트를 50 μL의 혈청 + RVP 혼합물로 싱글톤으로 접종하고 37℃에서 72시간 동안 인큐베이션하였다. 인큐베이션 종료 시, 고함량 이미저 상에서 플레이트를 스캔하고, 개별 GFP 발현 세포를 계수하였다. 억제 희석 역가($ID_{50}$)는 시험에서 바이러스 플라크의 수를 50%까지 감소시키는 희석의 역수로서 기록되었다. 각 시험 샘플에 대한 $ID_{50}$은 50% 중화점($ID_{50}$ 역가 = (50% 중화점 - 절편)/기울기) 미만의 플라크 수를 갖는 최종 희석액 및 50% 중화점을 초과하는 플라크 수를 갖는 최초 희석액을 사용하여 기울기 및 절편을 계산함으로써 보간되었다. 도 17은 결과를 요약한다.

[1084] 도 17에서 도시되는 바와 같이, 최초 2회 용량의 프라임-부스트 면역화 후 약 9개월차에 SARS-CoV-2 S 단백질의 융합-안정화된 남아공 변이체를 암호화하는 최적화된 뉴클레오티드 서열을 포함하는 mRNA 백신을 사용하는 NHP로 부스터 면역화된 2개의 군 모두에서, 원래 우한 계열의 SARS-CoV-2 S 단백질을 발현하는 우한 PsV에 대한 높은 중화 효능을 제공하였다. 이들 데이터는, SARS-CoV 2 S 단백질의 남아공 변이체를 암호화하는 mRNA 백신에 대한 노출이 원래 mRNA 백신에 의해 암호화된 SARS-CoV-2 S 단백질에 대한 중화 항체 반응을 촉진한다는 것을 시사한다. SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체 및 우한 계열로부터 유래된 융합-전 안정화된 S 단백질을 암호화하는 원래의 mRNA의 혼합물에 대한 노출은, SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체를 암호화하는 mRNA 백신 단독에 대한 노출보다 원래 우한 계열의 S 단백질에 대한 중화 항체 반응을 부스팅하는 데 더 이상 효과적이지 않았다.

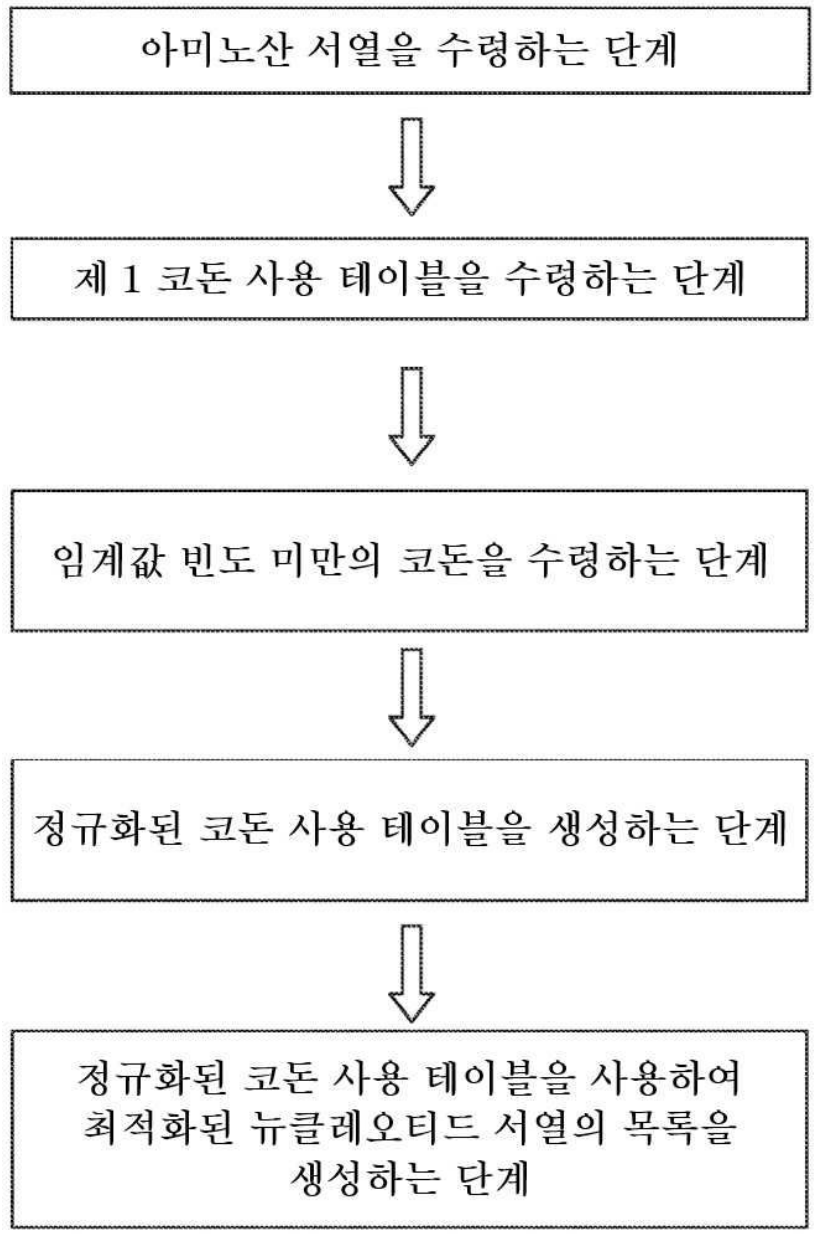
[1085] 흥미롭게도, SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체를 암호화하는 mRNA 백신을 사용한 면역화는 또한, 남아공에서 관찰된 SARS-CoV-2 S 단백질의 자연 발생 변이체 및 일본/브라질 및 캘리포니아에서 관찰된 SARS-CoV-2 S 단백질의 자연 발생 변이체를 발현하는 시험된 다른 모든 PsV에 대한 높은 중화 효능을 야기하였다. 놀랍게도, 항원 반응은 매우 광범위하여, SARS-CoV-1의 S 단백질을 발현하는 PsV 또한 NHP 시험 혈청에 의해 효과적으로 중화되었다. SARS-CoV-1의 S 단백질은 SARS-CoV-2 우한의 S 단백질과 76%만 동일하기 때문에 이는 예상하지 못한 것이었다.

[1086] 도 17에 도시된 바와 같이, 대부분의 경우 중화 항체 반응은 원래 우한 계열로부터 유래된 S 단백질만큼 변이체 S 단백질에 대해 효과적이었다. 또한, SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체를 암호화하는 mRNA 백신을 사용한 부스터 접종 후 관찰된 중화 항체 반응의 크기는, 실시예 5의 mRNA 백신을 사용한 원래의 프라임-부스트 면역화에 반응하여 35일차에 유도된 중화 항체 반응과 유사하거나 보다 높았다.

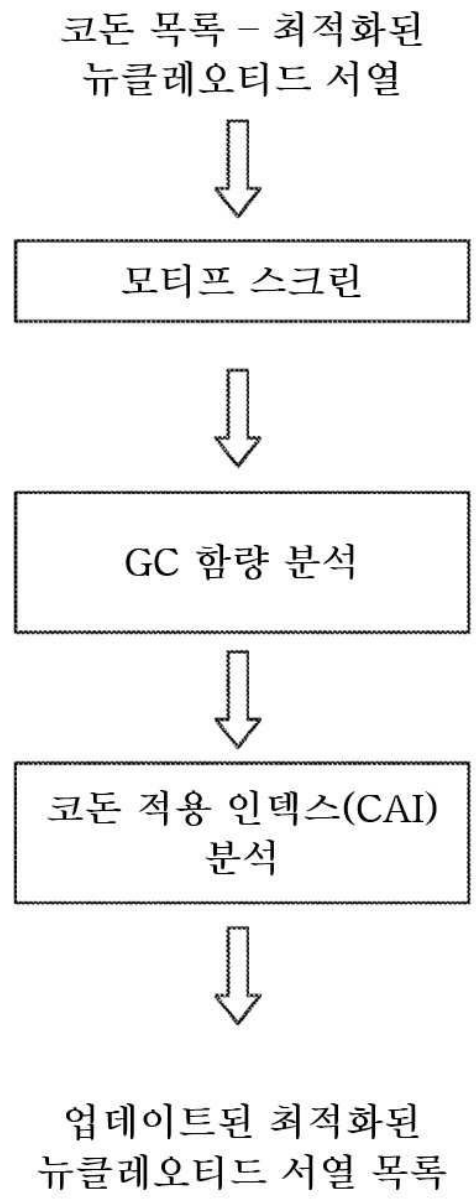
[1087] 이들 데이터는, SARS-CoV-2 우한의 S 단백질에 대한 중화 항체를 유도하는 백신으로 이전에 면역화된 대상체 및 SARS-CoV-2 S 단백질의 융합-전 안정화된 남아공 변이체를 암호화하는 본 발명의 최적화된 뉴클레오티드 서열을 포함하는 mRNA 백신을 후속적으로 투여받은 대상체는 광범위한 S 단백질 변이체에 대해 효과적이고 광범위 중화 항체 반응을 마운트할 수 있으며, 따라서 원래의 SARS-CoV-2 우한 계열 뿐만 아니라 다른 β-코로나바이러스, 특히 SARS-CoV-1과 같은 안지오텐신-전환 효소 2(ACE 2)에 결합하는 스파이크 단백질을 발현하는 β-코로나바이러스에 의해 야기되는 COVID-19 감염을 효과적으로 보호한다.

도면

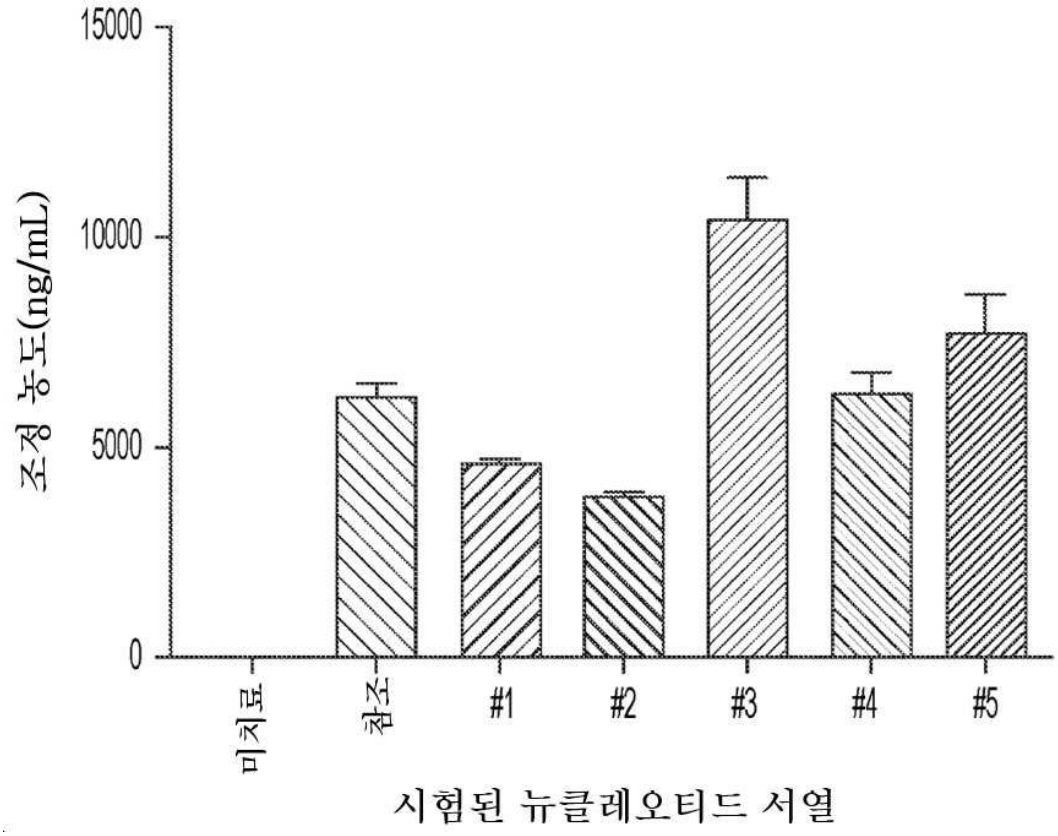
도면1a


아미노산 서열을 수령하는 단계
제 1 코돈 사용 테이블을 수령하는 단계
임계값 빈도 미만의 코돈을 수령하는 단계
정규화된 코돈 사용 테이블을 생성하는 단계
정규화된 코돈 사용 테이블을 사용하여
최적화된 뉴클레오티드 서열의 목록을
생성하는 단계

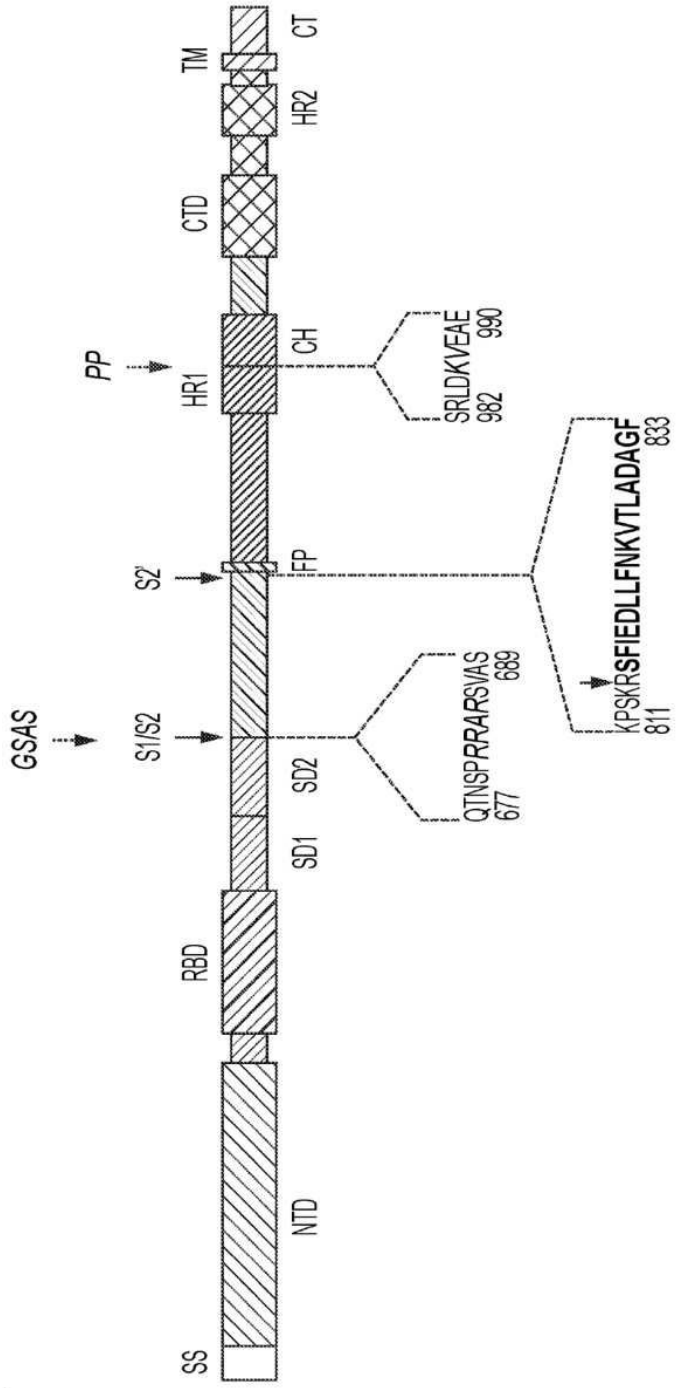
도면1b


코돈 목록 – 최적화된 뉴클레오티드 서열
모티프 스크린
GC 함량 분석
코돈 적용 인덱스(CAI) 분석
업데이트된 최적화된 뉴클레오티드 서열 목록

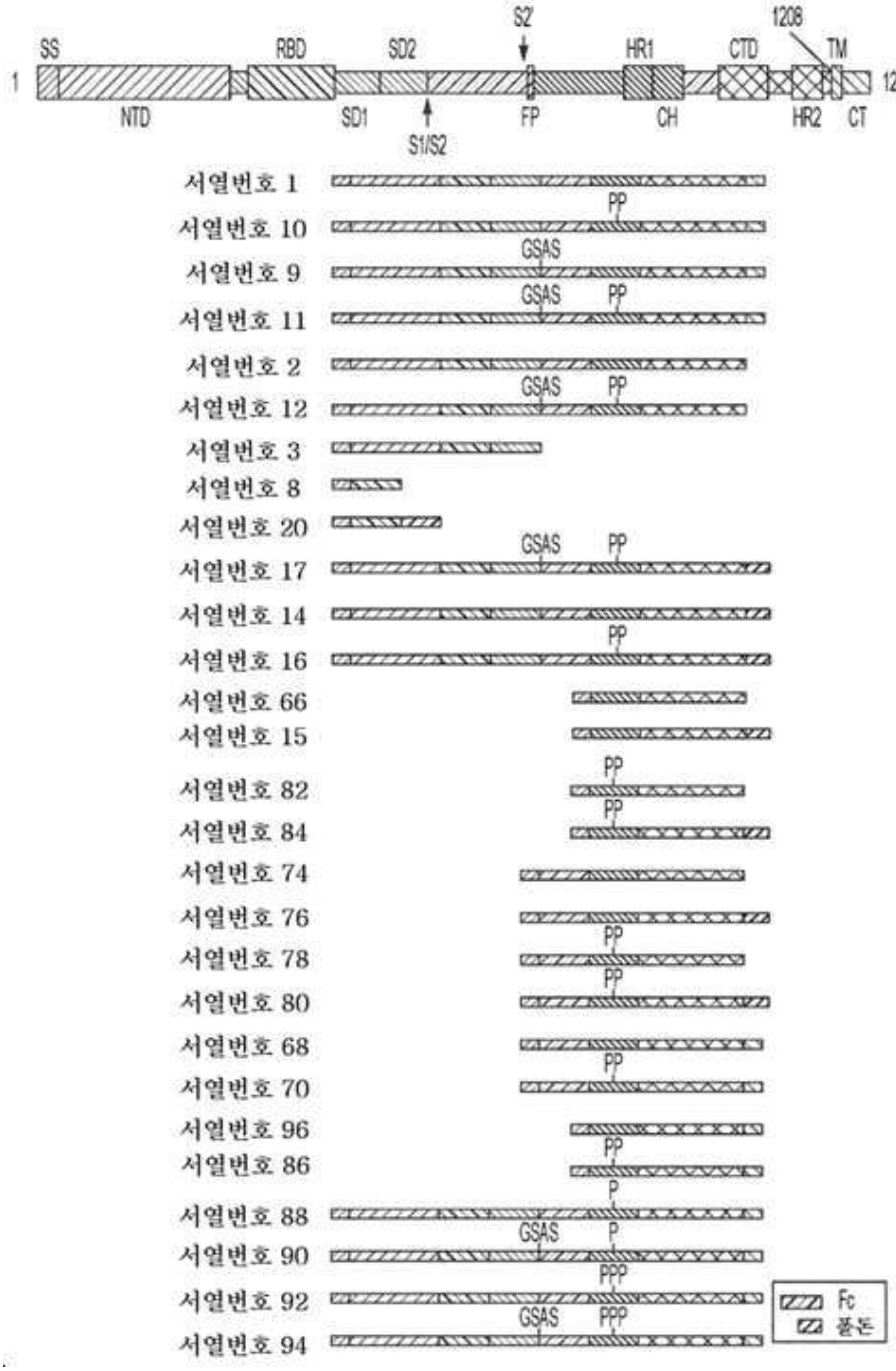
도면2



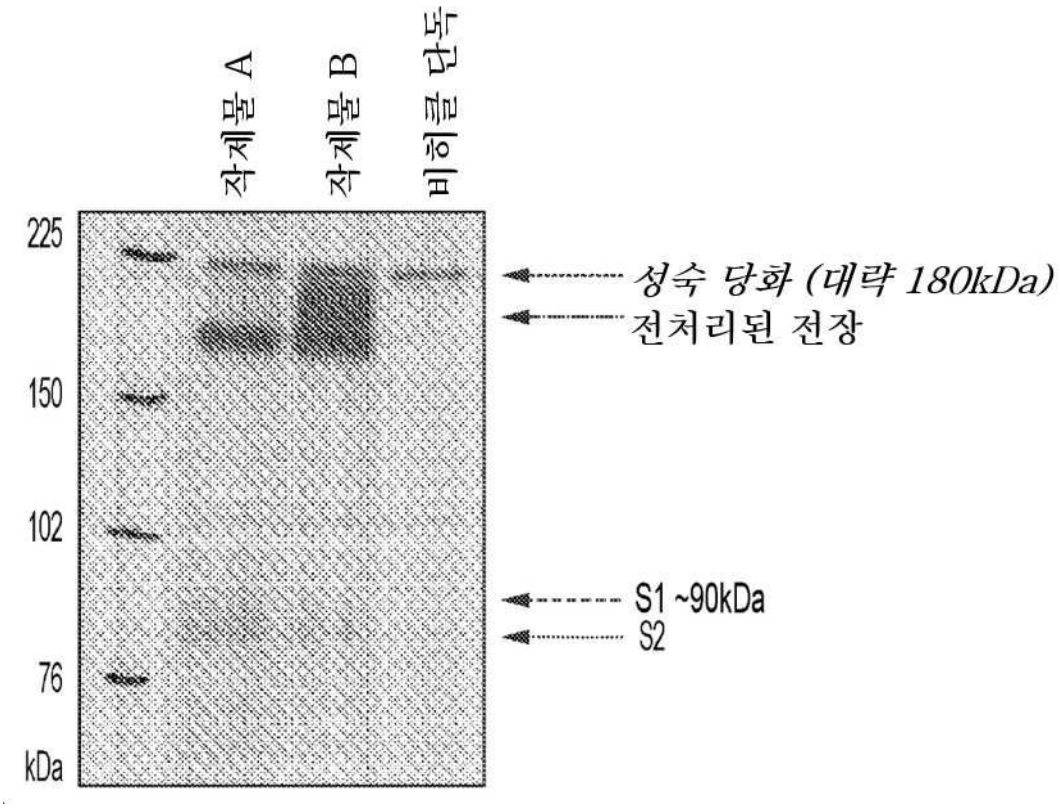
도면3


GSAS
PP
S1/S2
S2'
SS
NTD
RBD
SD1
SD2
HR1
CH
CTD
HR2
TM
CT
FP
SRLDKVEAE
982
990
QTNSPRRARSVAS
677
689
KPSKRSFIEDLLFNKVTLADAGF
811
833

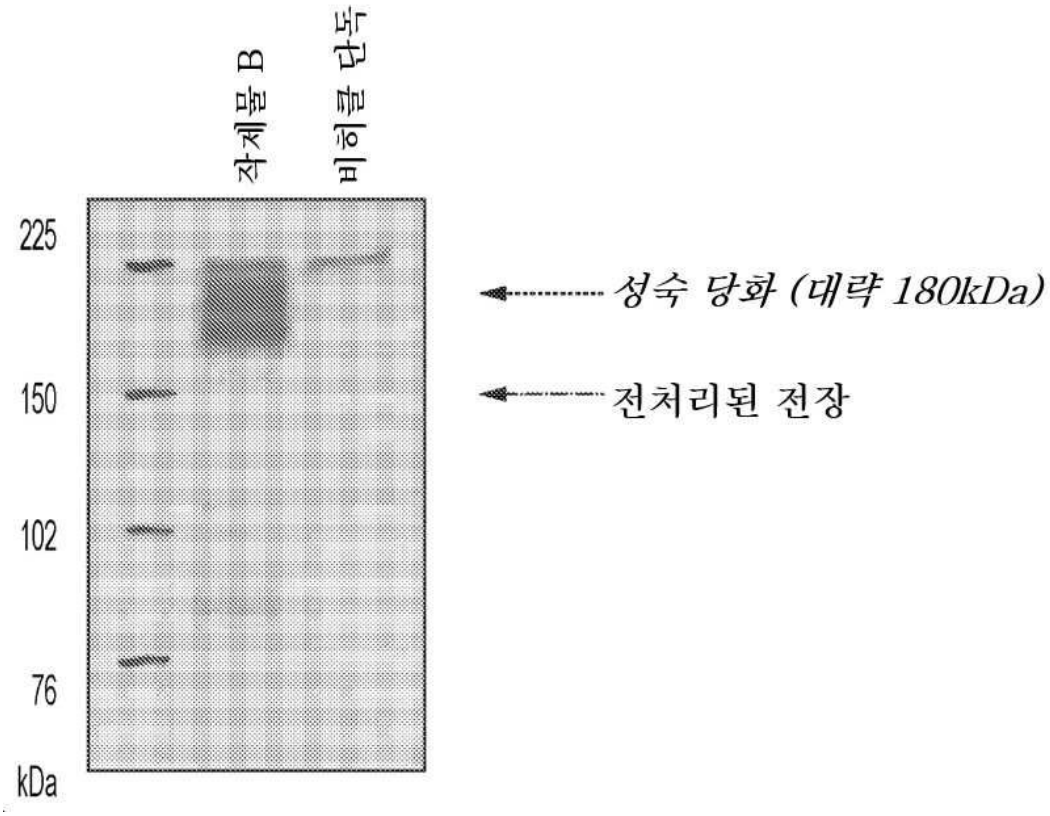
도면4


SS
RBD
SD2
S2'
HR1
CTD
1208
TM
1
1273
NTD
SD1
S1/S2
FP
CH
HR2
CT
서열번호 1
서열번호 10
서열번호 9
서열번호 11
서열번호 2
서열번호 12
서열번호 3
서열번호 8
서열번호 20
서열번호 17
서열번호 14
서열번호 16
서열번호 66
서열번호 15
서열번호 82
서열번호 84
서열번호 74
서열번호 76
서열번호 78
서열번호 80
서열번호 68
서열번호 70
서열번호 96
서열번호 86
서열번호 88
서열번호 90
서열번호 92
서열번호 94
PP
GSAS
P
PPP
Fc
폴돈

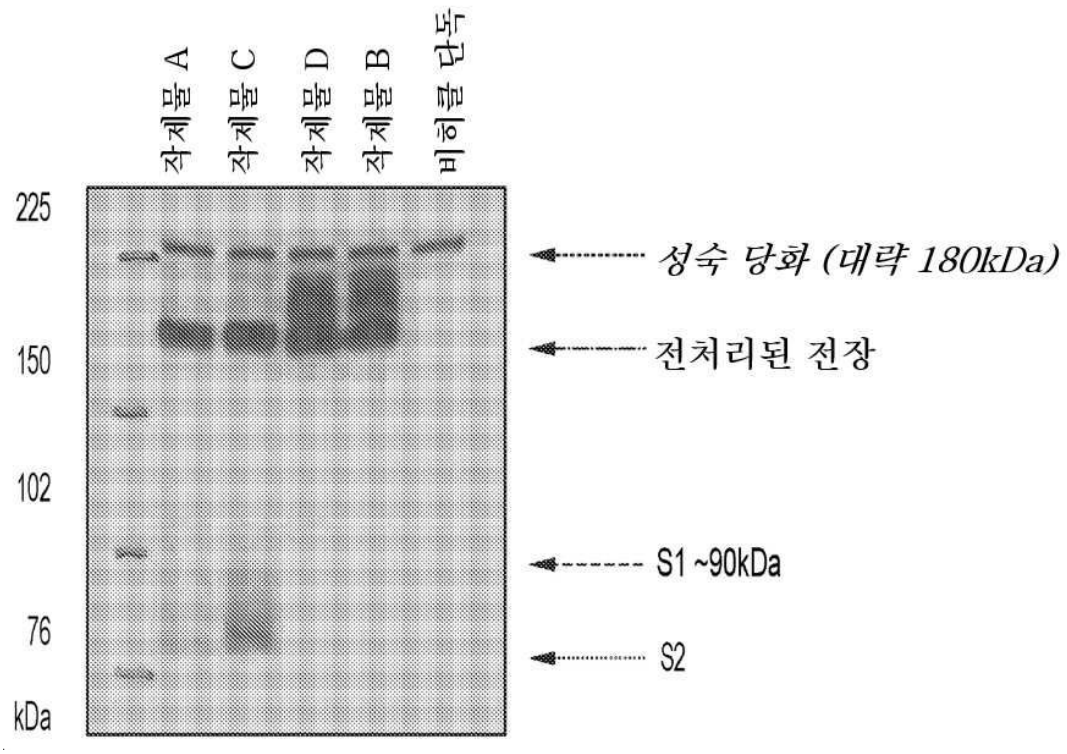
도면5


작제물 A
작제물 B
비히클 단독
225
150
102
76
kDa
성숙 당화 (대략 180kDa)
전처리된 전장
S1 ~90kDa
S2

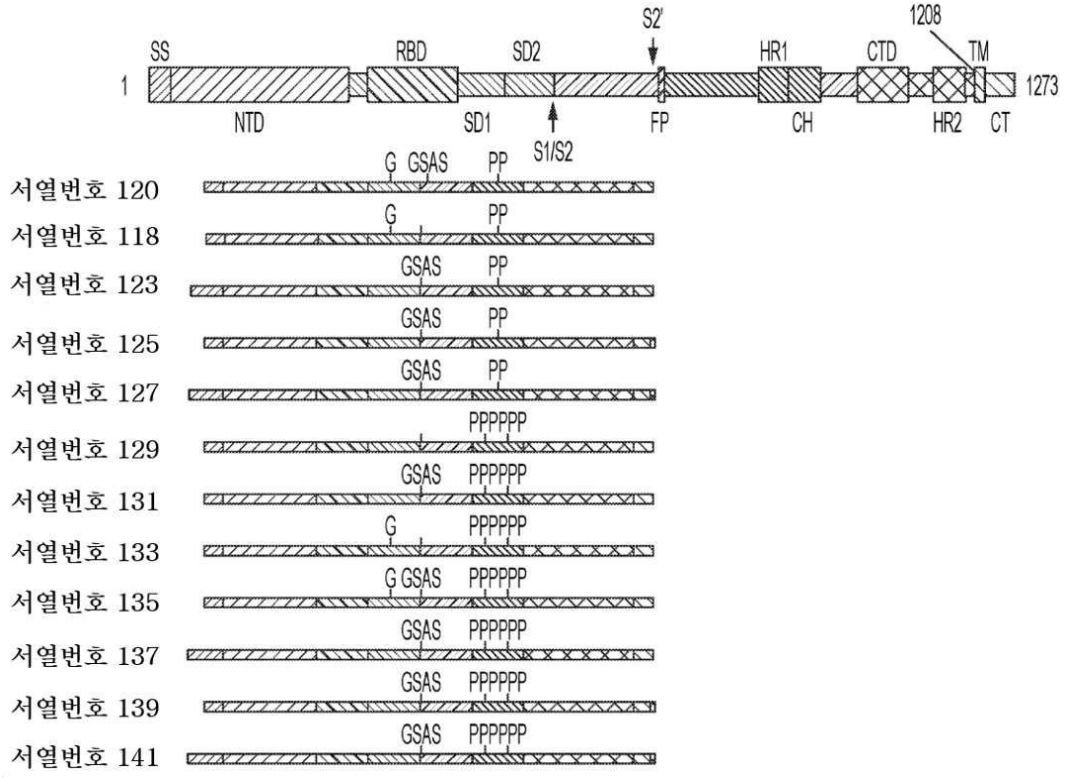
도면6


작제물 B
비히클 단독
225
150
102
76
kDa
성숙 당화 (대략 180kDa)
전처리된 전장

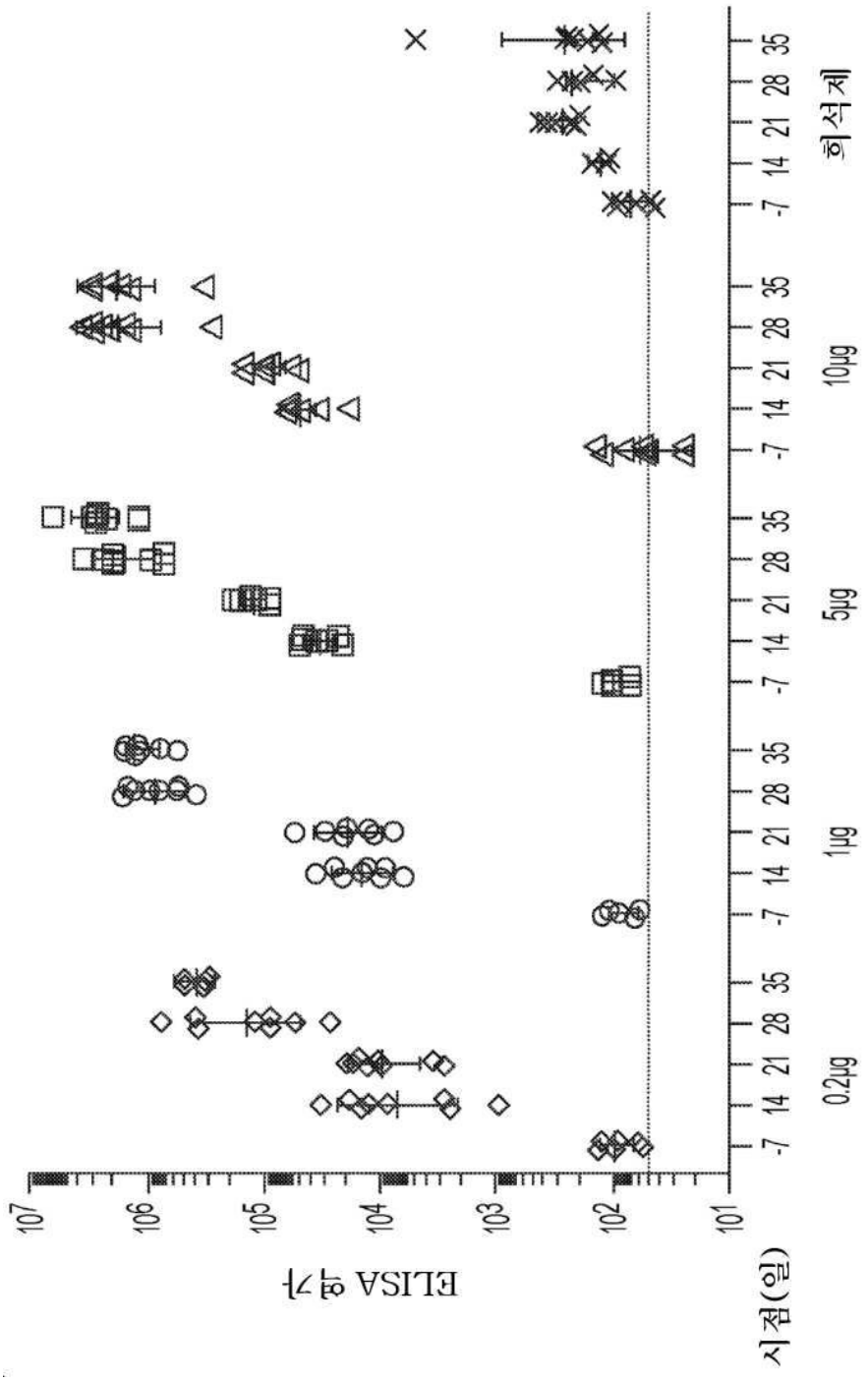
도면7


작제물 A
작제물 C
작제물 D
작제물 B
비히클 단독
225
150
102
76
kDa
성숙 당화 (대략 180kDa)
전처리된 전장
S1 ~90kDa
S2

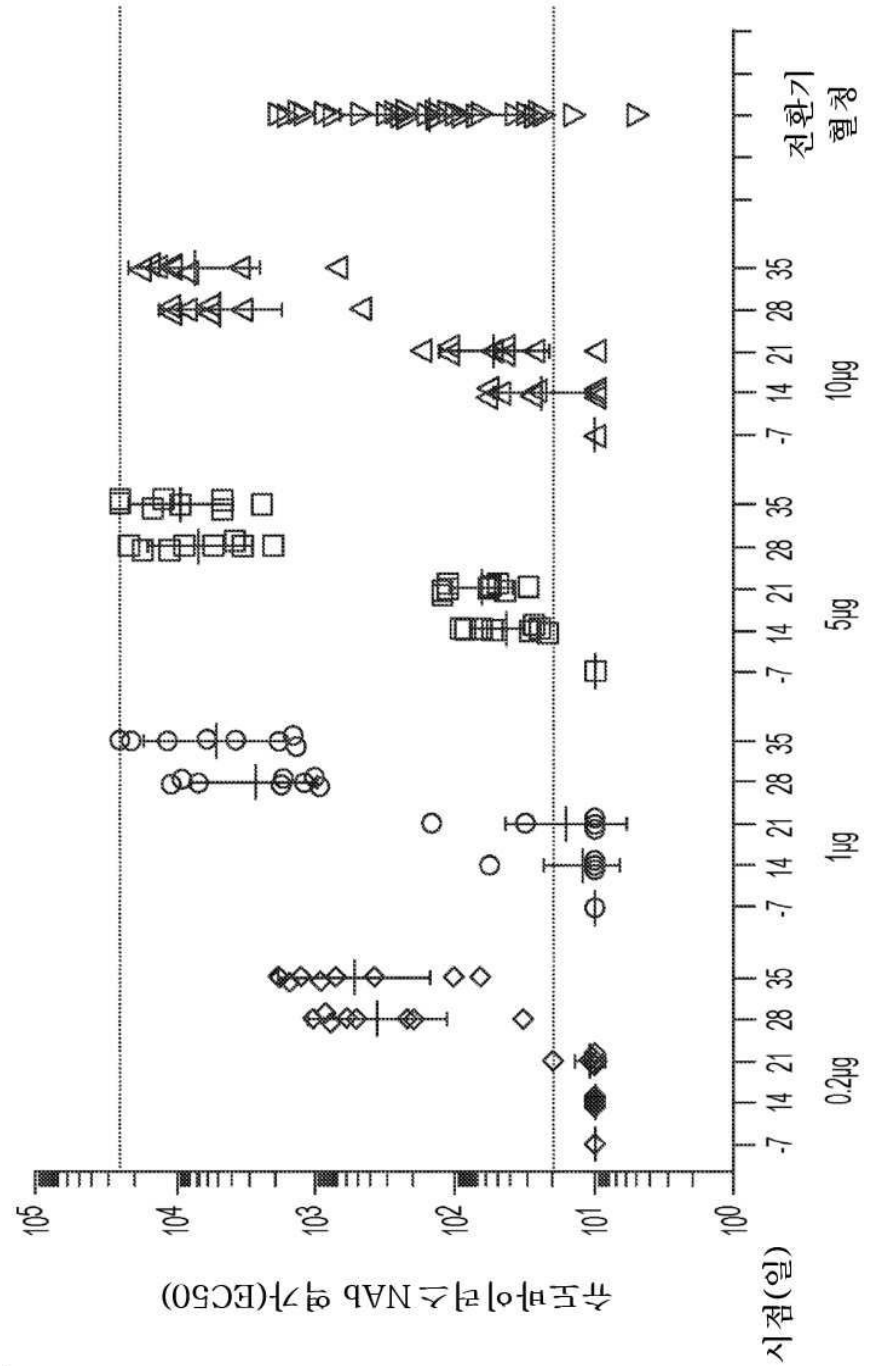
도면8


SS
RBD
SD2
S2'
HR1
CTD
1208
TM
1
1273
NTD
SD1
S1/S2
FP
CH
HR2
CT
서열번호 120
서열번호 118
서열번호 123
서열번호 125
서열번호 127
서열번호 129
서열번호 131
서열번호 133
서열번호 135
서열번호 137
서열번호 139
서열번호 141
G GSAS PP
G PP
GSAS PP
GSAS PP
GSAS PP
PPPPPP
GSAS PPPPPP
G PPPPPP
G GSAS PPPPPP
GSAS PPPPPP
GSAS PPPPPP
GSAS PPPPPP

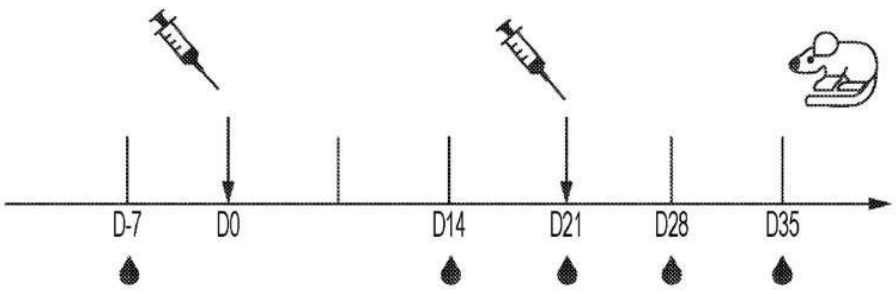
도면9a


ELISA 역가
$10^7$
$10^6$
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
시점(일)
-7 14 21 28 35
0.2μg
1μg
5μg
10μg
희석제

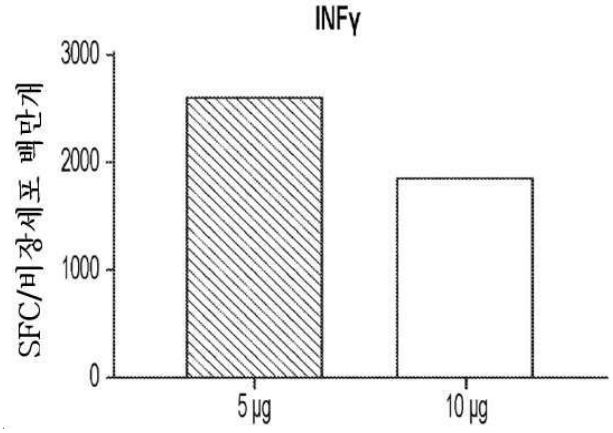
도면9b


슈도바이러스 NAb 역가(EC50)
시점(일)
전환기
혈청
0.2μg
1μg
5μg
10μg

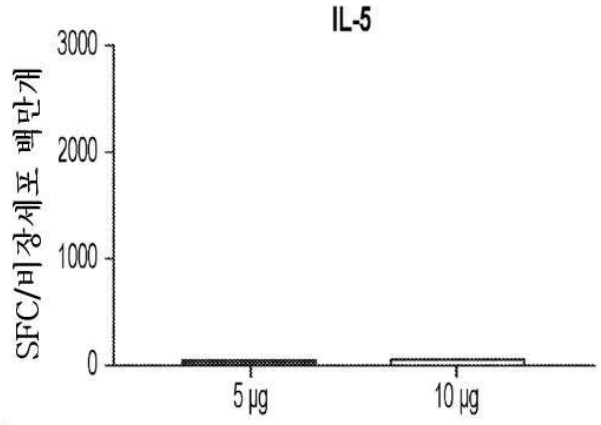
도면9c


D-7
D0
D14
D21
D28
D35

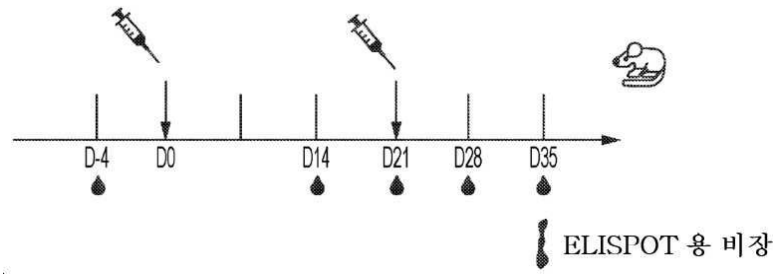
도면10a



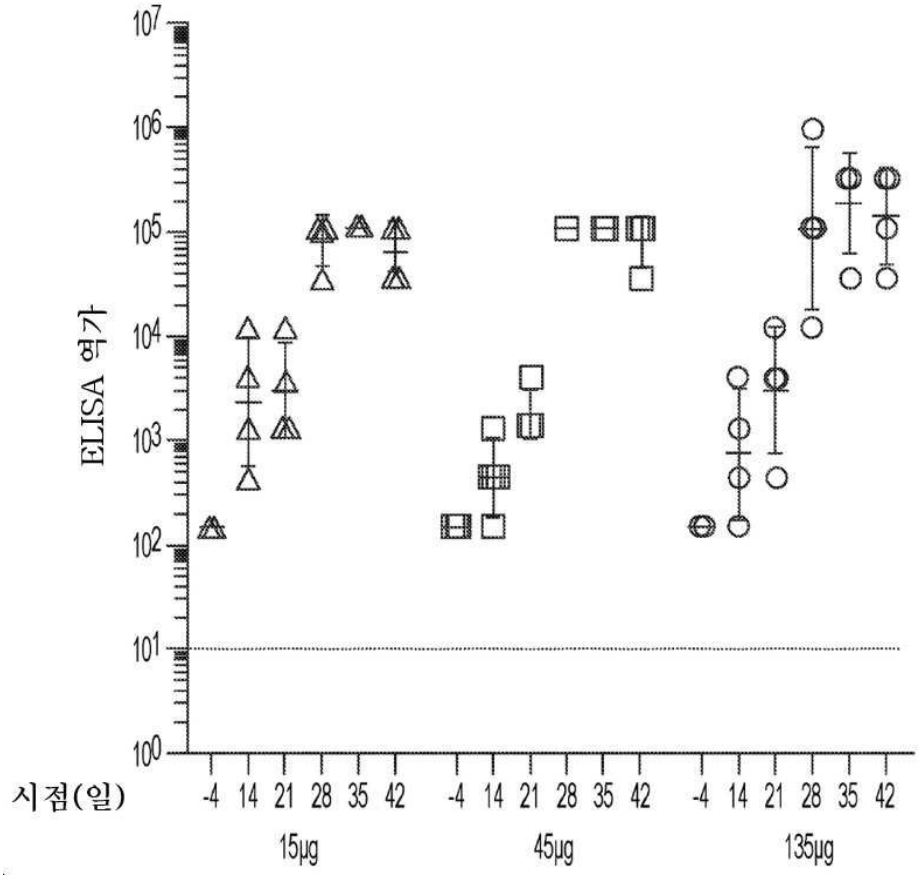

도면10b



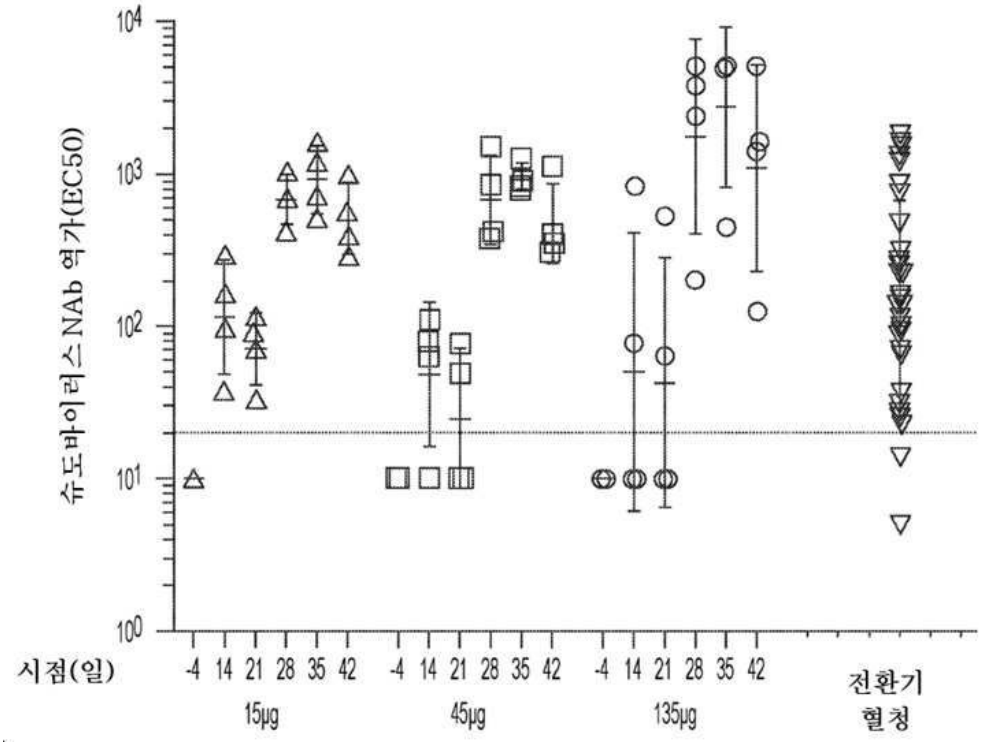

도면10c



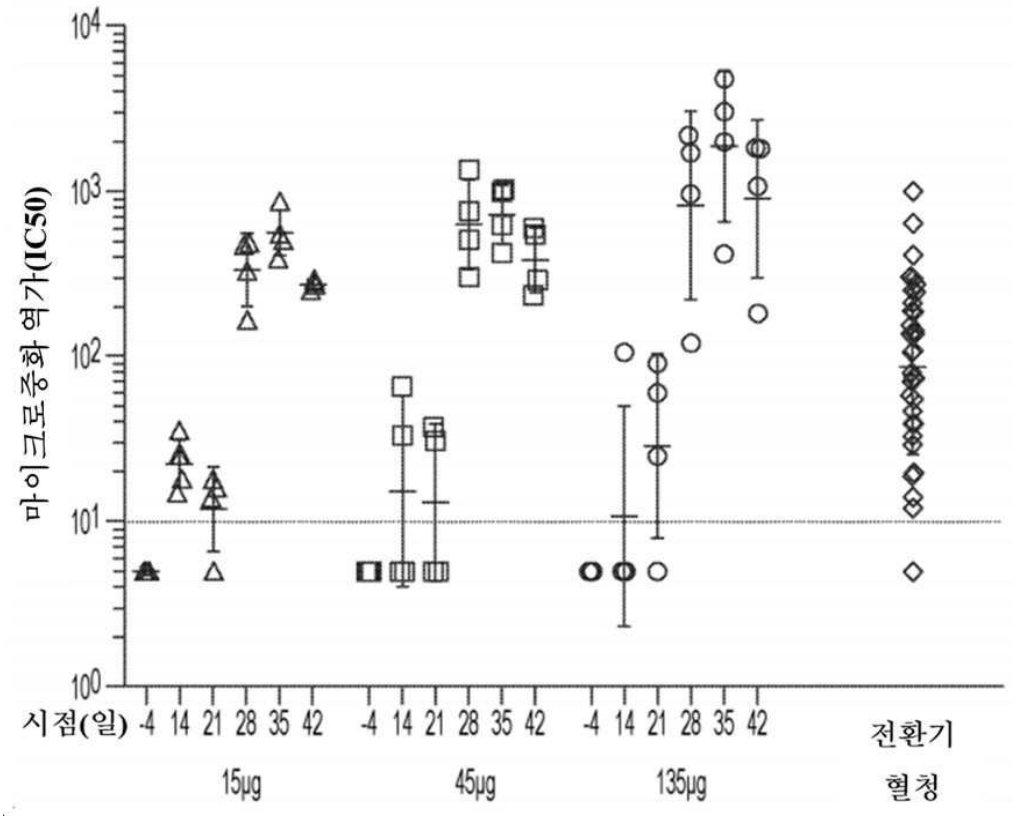
도면11a


ELISA 역가
$10^7$
$10^6$
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
시점(일)
-4 14 21 28 35 42 -4 14 21 28 35 42 -4 14 21 28 35 42
15μg
45μg
135μg

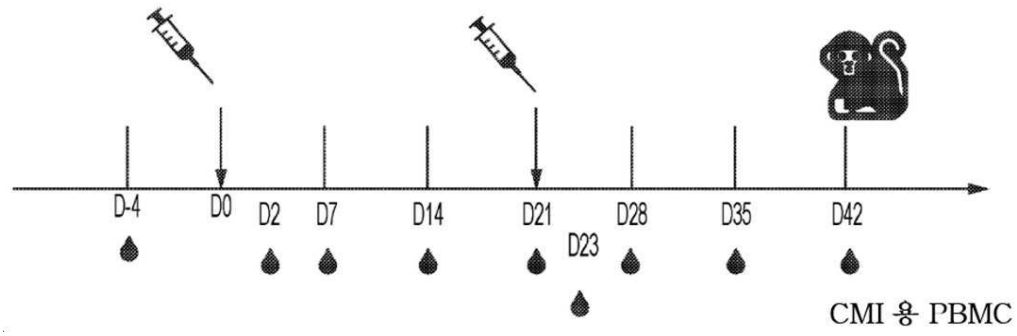
도면11b


슈도바이러스 NAb 역가(EC50)
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
시점(일)
-4 14 21 28 35 42 -4 14 21 28 35 42 -4 14 21 28 35 42
15μg
45μg
135μg
전환기
혈청

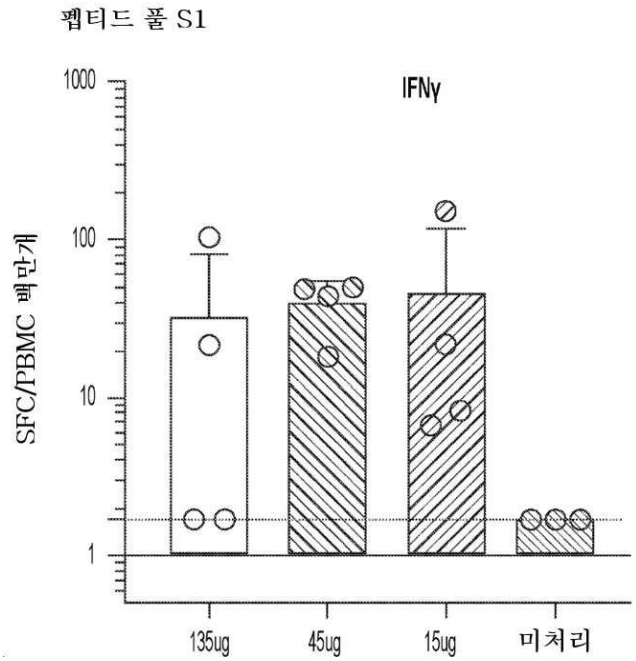
도면11c


마이크로중화 역가(IC50)
10^4
10^3
10^2
10^1
10^0
시점(일) -4 14 21 28 35 42 -4 14 21 28 35 42 -4 14 21 28 35 42
15μg
45μg
135μg
전환기
혈청

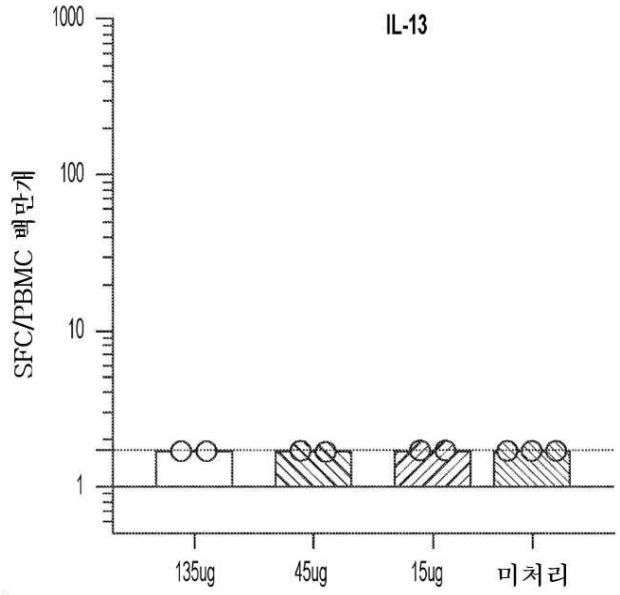
도면11d


D-4
D0
D2
D7
D14
D21
D23
D28
D35
D42
CMI 용 PBMC

도면12a

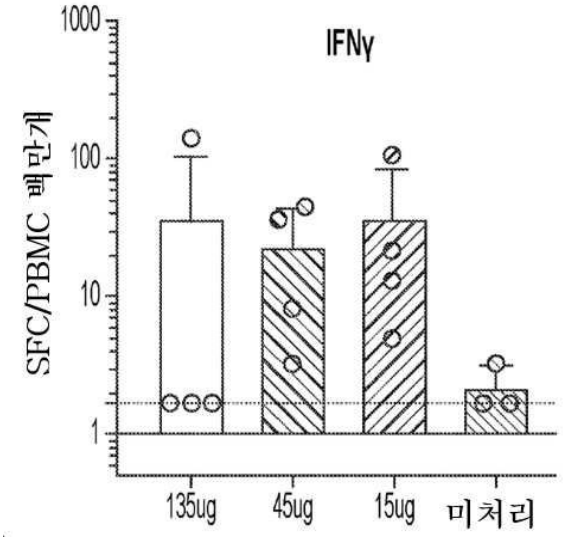

펩티드 풀 S1
IFNγ
SFC/PBMC 백만개
1000
100
10
1
135ug
45ug
15ug
미처리

도면12b

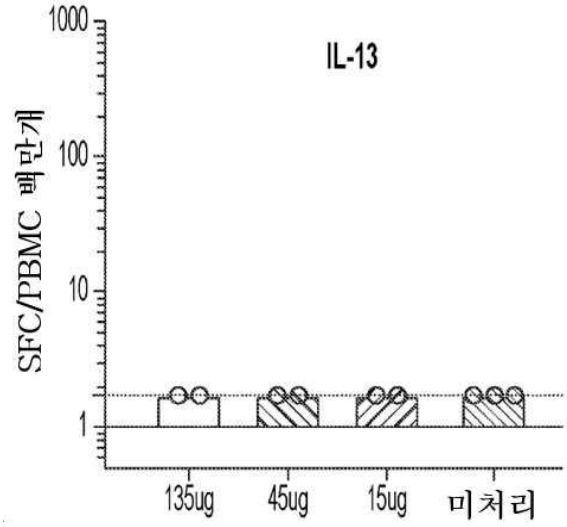

IL-13
SFC/PBMC 백만개
1000
100
10
1
135ug
45ug
15ug
미처리

도면12c

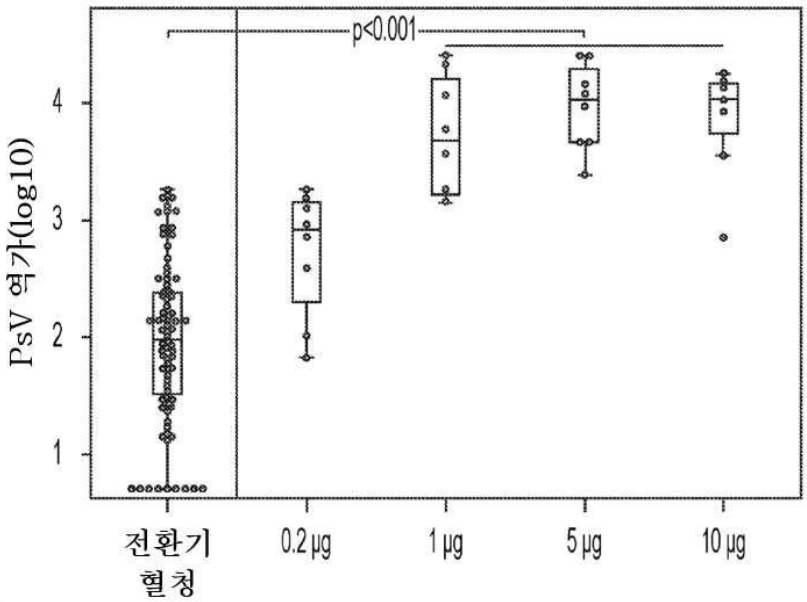
펩티드 풀 S2




도면12d



도면13a

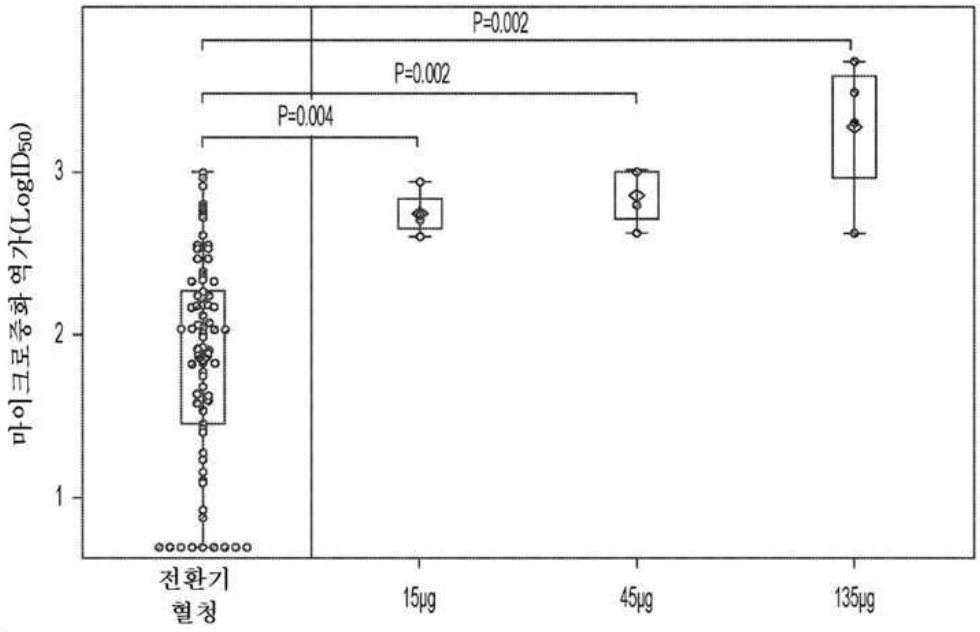



도면13b

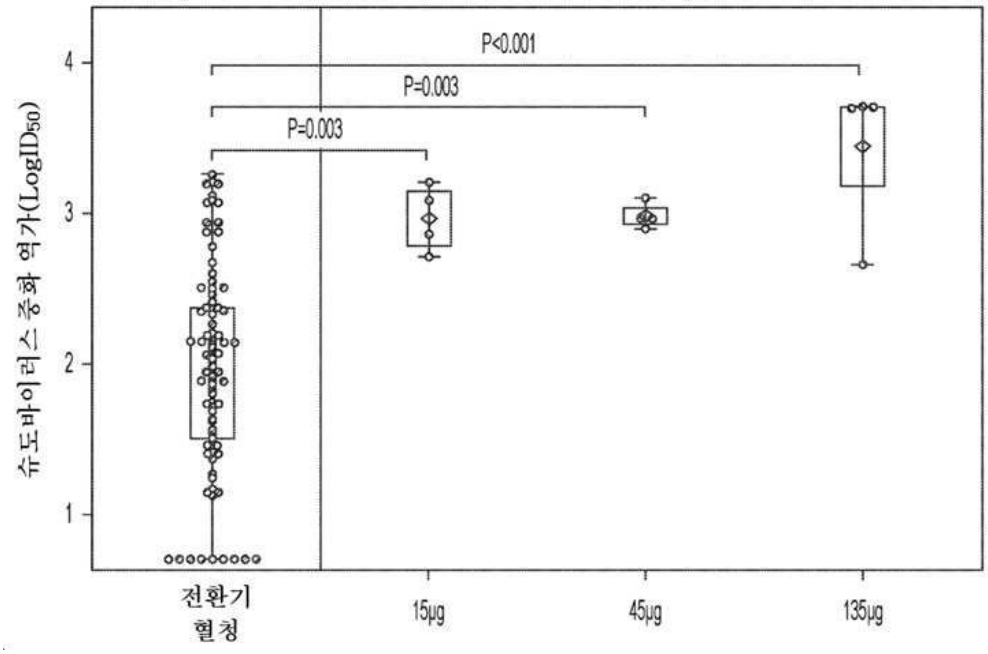
| | | Spearman 상관 계수* | | |
|---|---|---|---|---|
| 투여량 | N | IgG/MN | IgG/PsV | MN/PsV |
| 15 μg | 4 | 0.94 (± 0.01) | 0.95 (± 0.02) | 0.95 (± 0.02) |
| 45 μg | 4 | 0.88 (± 0.02) | 0.90 (± 0.02) | 0.92 (± 0.03) |
| 135 μg | 4 | 0.93 (± 0.03) | 0.88 (± 0.03) | 0.91 (± 0.02) |
| 합계 | 12 | 0.92 (± 0.01) | 0.91 (± 0.02) | 0.93 (± 0.01) |

*평균(± 표준 오차)

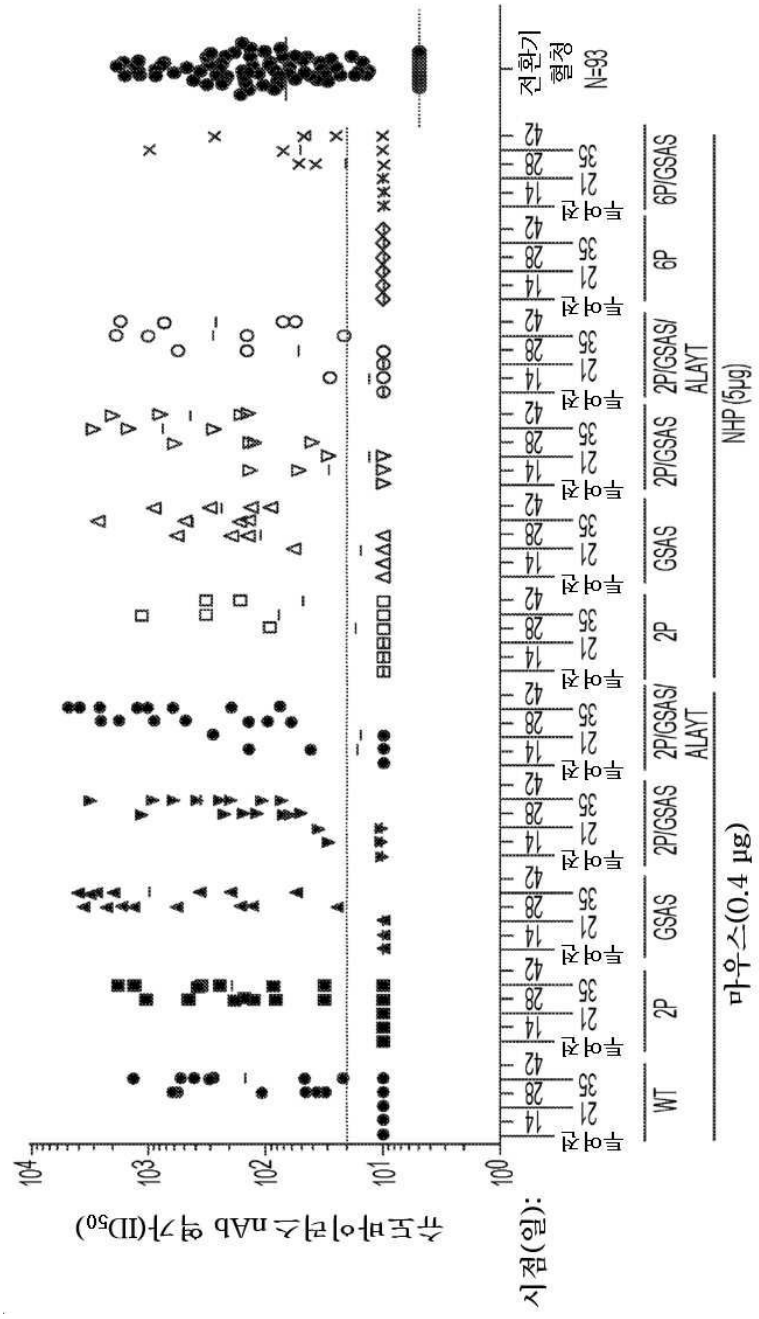
도면13c



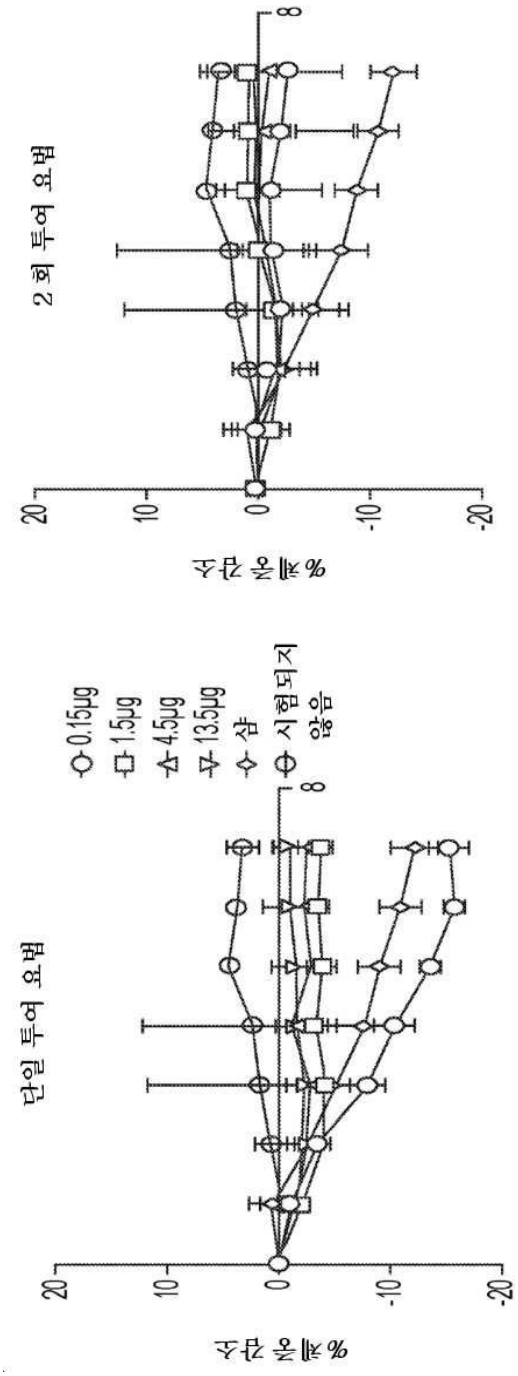
도면13d


P<0.001
P=0.003
P=0.003
슈도바이러스 중화 역가(LogID50)
4
3
2
1
전환기 혈청
15μg
45μg
135μg

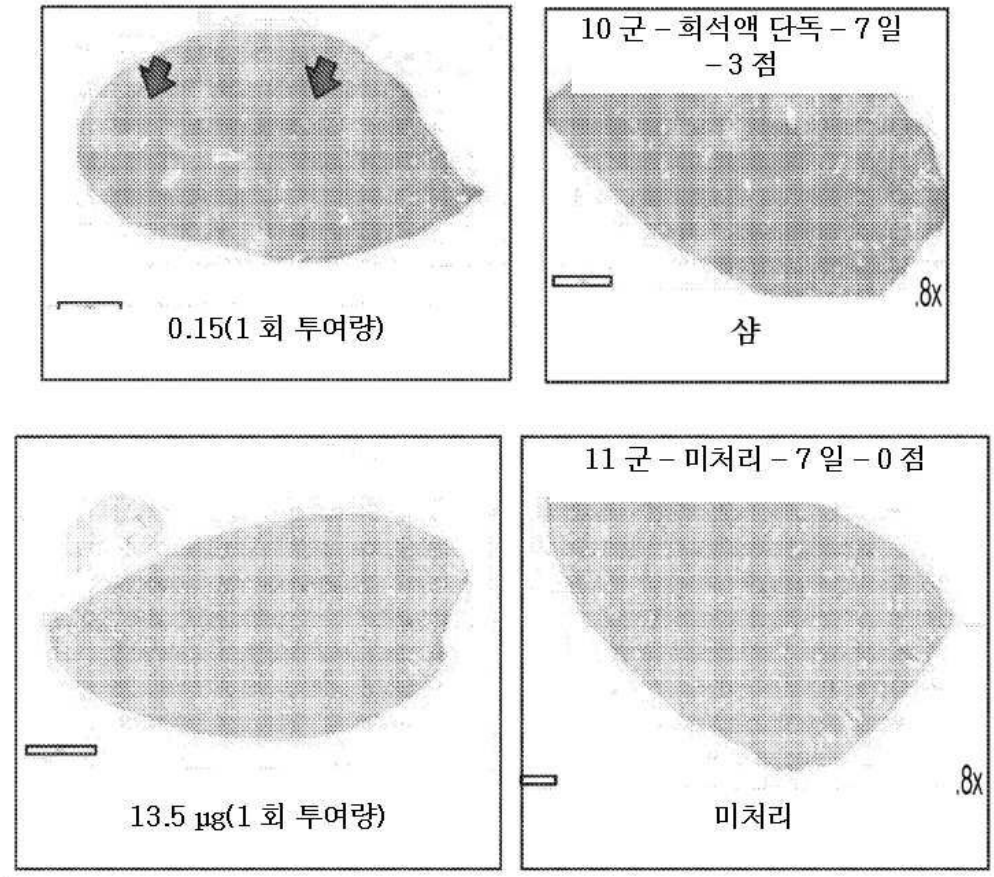
도면14


슈도바이러스 nAb 역가(ID50)
시점(일):
투여전
WT
2P
GSAS
2P/GSAS
2P/GSAS/ ALAYT
6P
6P/GSAS
전환기 혈청 N=93
마우스(0.4 μg)
NHP (5μg)

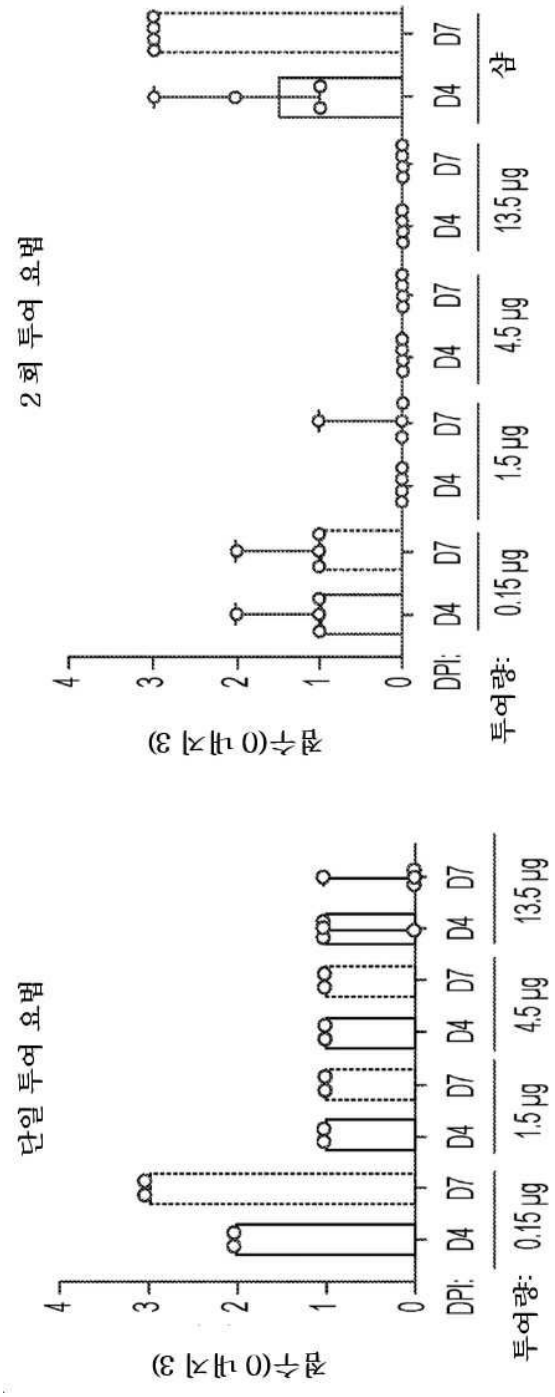
도면15a


2회 투여 요법
% 체중 감소
20
10
0
-10
-20
8
단일 투여 요법
0.15μg
1.5μg
4.5μg
13.5μg
샴
시험되지 않음

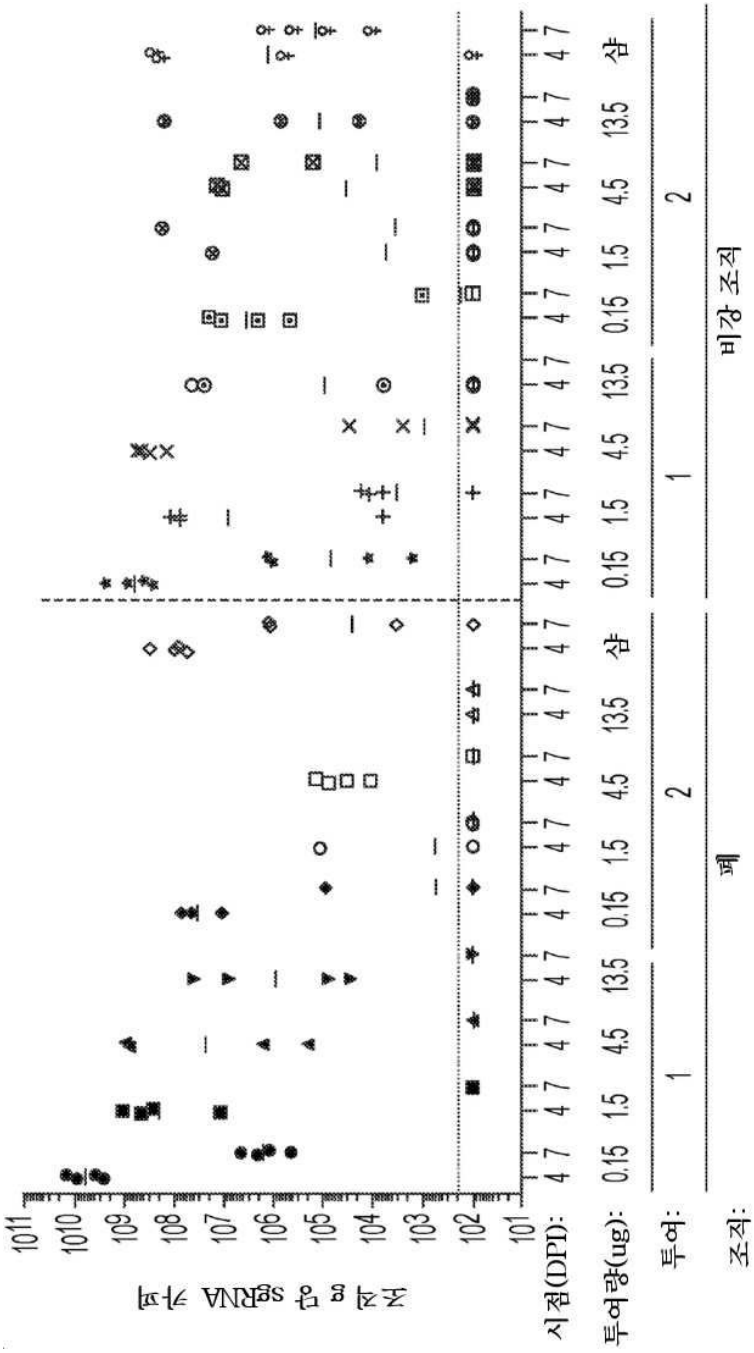
도면15b



도면15c


2 회 투여 요법
점수(0 내지 3)
0
1
2
3
4
DPI: D4 D7 D4 D7 D4 D7 D4 D7 D4 D7
투여량: 0.15 μg 1.5 μg 4.5 μg 13.5 μg
샴
단일 투여 요법
점수(0 내지 3)
0
1
2
3
4
DPI: D4 D7 D4 D7 D4 D7 D4 D7
투여량: 0.15 μg 1.5 μg 4.5 μg 13.5 μg

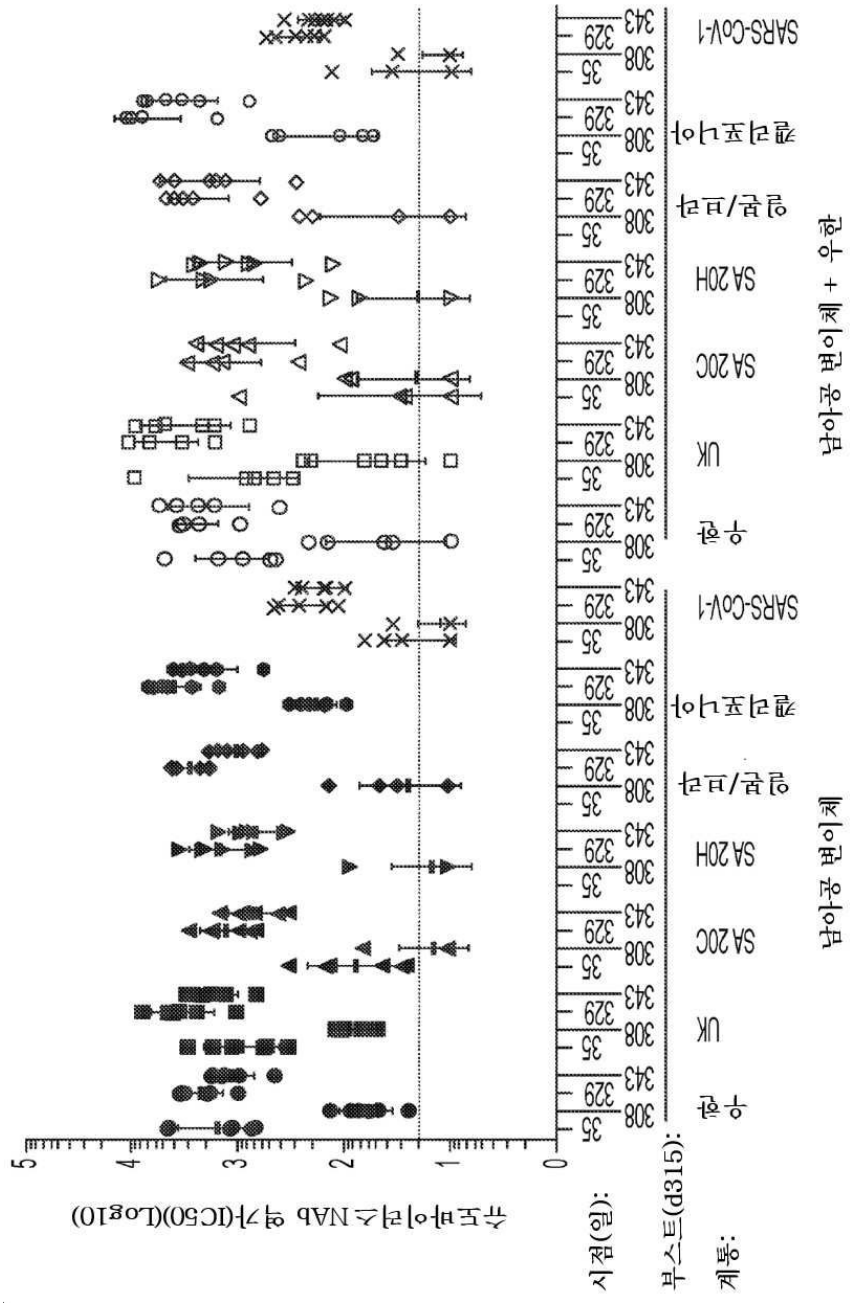
도면15d



도면16

| 계통 | | 우한에 대한 돌연변이 | D614G | GenBank | 카탈로그 # | 로트 |
|---|---|---|---|---|---|---|
| WIV04 | 우한-Hu-1 | NA | 아니오 | QHD43416.1 | RVP-701G | 113A |
| B.1.351 | 남아공 (20C/501Y.V2) | L18F, D80A, D215G, L242H, R246I, K417N, N501Y, E484K, A701V | 예 | Tegally 등, 2020 | RVP-707G | CG-136A |
| B.1.351 | 남아공 (20H/501Y.V2) | L18F, D80A, D215G, ΔL242/A243, R246I, K417N, N501Y, E484K, A701V | 예 | Tegally 등, 2020 | RVP-714G | CG-162A |
| P.1 | 일본/브라질 | L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I, V1176 | 예 | QQX12069.1 | RVP-708 | CG-160A |
| 452R | 캘리포니아 | S13I, W152C, L452R | 예 | QPJ72086.1 | RVP-713 | CG-161A |
| B.1.1.7 | UK (VUI202012/01) | ΔH69/V70, ΔY144, N501Y, A570D, P681H, T716I, S982A, D1118H | 예 | QQH18545.1 | RVP-706G | CG-135A |
| Urbani | SARS-CoV1 | NA | NA | P59594.1 | RVP-801G | SG-115B |

도면17




## 서 열 목 록

SEQUENCE LISTING

<110> Translate Bio, Inc.

<120> OPTIMIZED NUCLEOTIDE SEQUENCES ENCODING SARS-COV-2 ANTIGENS

<130> MRT-2161WO

<141> 2021-05-07

<150> 63/021,319

<151> 2020-05-07

<150> 63/032,825

<151> 2020-06-01

<150> 63/076,718

<151> 2020-09-10

<150> 63/076,729

<151> 2020-09-10

<150> 63/088,739

<151> 2020-10-07

<150> 63/143,604

<151> 2021-01-29

<150> 63/143,612

<151> 2021-01-29

<150> 63/146,807

<151> 2021-02-08

<160> 173

<170

> SeqWin2010, version 1.0

<210> 1

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 1

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

```
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355                 360                 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

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

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

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

                565                 570                 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580                 585                 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
```

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

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

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

1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 2

<211> 1208

<212> PRT

<213> SARS-CoV-2

<400> 2

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100
Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu

1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165
Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln
1205

<210> 3
<211> 681
<212> PRT
<213> SARS-CoV-2
<400> 3
Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15
Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285
Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300
Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335
Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

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

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

500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro

675 680

<210> 4

<211> 589

<212> PRT

<213> SARS-CoV-2

<400> 4

Met Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly

1 5 10 15

Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr

20 25 30

Asn Phe Thr Ile Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr

35 40 45

Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu

50 55 60

Cys Ser Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn

65 70 75 80

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

85 90 95

Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp

100 105 110

Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro

115 120 125

Ser Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu

130 135 140

Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile

145 150 155 160

Ala Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val

165 170 175

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

180 185 190

Leu Leu Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala

195 200 205

Ala Leu Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly

210 215 220

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

225 230 235 240

Asn Gln Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser

245 250 255

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

260 265 270

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

275 280 285

Ile Ser Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu

290 295 300

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

305 310 315 320

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

325 330 335

Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln

340 345 350

Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe

355 360 365

Pro Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val

370 375 380

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp

385 390 395 400

Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr

405 410 415

His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr

420 425 430

Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile

435 440 445

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

450 455 460

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

465 470 475 480

Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln

485 490 495

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

500 505 510

Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp

515 520 525

Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val

530 535 540

Met Val Thr Ile Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu
545 550 555 560
Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp
565 570 575
Ser Glu Pro Val Leu Lys Gly Val Lys Leu His Tyr Thr

580 585

<210> 5
<211> 394
<212> PRT
<213> SARS-CoV-2
<400> 5
Met Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp
1 5 10 15
Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
20 25 30
Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro
35 40 45

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

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala
115 120 125
Ser Ala Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala
130 135 140
Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser
145 150 155 160
Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu

165 170 175

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

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

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn
305 310 315 320
Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu
325 330 335
Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu
340 345 350
Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu
355 360 365

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
370 375 380
Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln
385 390
<210> 6
<211> 193

<212> PRT

<213> SARS-CoV-2

<400> 6

Met Pro Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala
1 5 10 15

Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn
20 25 30

Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr
35 40 45

Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe
50 55 60

Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg
65 70 75 80

Gln Ile Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys
85 90 95

Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn
100 105 110

Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe
115 120 125

Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile
130 135 140

Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys
145 150 155 160

Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly
165 170 175

Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His Ala
180 185 190

Pro

<210> 7

<211> 15

<212> PRT

<213> SARS-CoV-2

<400> 7

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys
1 5 10 15

<210> 8

<211> 207

<212> PRT

<213> Artificial Sequence

<400> 8

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro
1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg
20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val
35 40 45

Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys
50 55 60

Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn
65 70 75 80

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

Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro
100 105 110

Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn Leu Asp
115 120 125

Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys
130 135 140

Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln
145 150 155 160

Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe
165 170 175

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln

180 185 190

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

195 200 205

<210> 9

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 9

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

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

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

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

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 10

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 10

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly

865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala

930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val

995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu

1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 11

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 11

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu

1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100
Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165
Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu

1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215
Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230
Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260
Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 12

<211> 1208

<212> PRT

<213> Artificial Sequence

<400> 12

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

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

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

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

420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720
Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

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

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

865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala

930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln
1205

<210> 13

<211> 27

<212> PRT

<213> Artificial Sequence

<400> 13

Gly Tyr Ile Pro Glu Ala Pro Arg Asp Gly Gln Ala Tyr Val Arg Lys
1 5 10 15

Asp Gly Glu Trp Val Leu Leu Ser Thr Phe Leu
20 25

<210> 14

<211> 1235

<212> PRT

<213> Artificial Sequence

<400> 14

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

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

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205
Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255

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

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

370 375 380

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

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605
Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165
Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala Pro Arg
1205 1210 1215

Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu Leu Ser
1220 1225 1230
Thr Phe Leu
1235

<210> 15
<211> 435
<212> PRT
<213> Artificial Sequence
<400> 15

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15
Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30
Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
35 40 45
Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
50 55 60
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
65 70 75 80
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

85 90 95

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

100 105 110

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

115 120 125

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

130 135 140

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

145 150 155 160

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

225 230 235 240

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

245 250 255

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

260 265 270

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

275 280 285

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

290 295 300

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

305 310 315 320

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

325 330 335

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

340 345 350

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

355 360 365

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

370 375 380

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

385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala Pro Arg

405 410 415

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

420 425 430

Thr Phe Leu

435

<210> 16

<211> 1235

<212> PRT

<213> Artificial Sequence

<400> 16

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

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

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

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

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365

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

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

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

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

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

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

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala Pro Arg

1205 1210 1215

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

1220 1225 1230

Thr Phe Leu

1235

<210> 17

<211> 1235

<212> PRT

<213> Artificial Sequence

<400> 17

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590

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

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

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750
Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765
Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800
Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala Pro Arg
1205 1210 1215
Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu Leu Ser
1220 1225 1230

Thr Phe Leu
1235
<210> 18
<211> 231
<212> PRT
<213> Artificial Sequence
<400> 18
Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro
1 5 10 15
Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys

20 25 30

Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val

35 40 45

Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp

50 55 60

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

65 70 75 80

Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp

85 90 95

Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu

100 105 110

Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg

115 120 125

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

130 135 140

Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp

145 150 155 160

Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys

165 170 175

Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser

180 185 190

Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser

195 200 205

Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser

210 215 220

Leu Ser Leu Ser Pro Gly Lys

225 230

<210> 19

<211> 423

<212> PRT

<213> Artificial Sequence

<400> 19

Met Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr

1 5 10 15

Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys

20 25 30

Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe

35 40 45

Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr

50 55 60

Asn Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln

65 70 75 80

Ile Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu

85 90 95

Pro Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn Leu

100 105 110

Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg

115 120 125

Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr

130 135 140

Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr

145 150 155 160

Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr

165 170 175

Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His Ala Pro

180 185 190

Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro

195 200 205

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

210 215 220

Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val

225 230 235 240

Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp

245 250 255

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

260 265 270

Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp

275 280 285

Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu

290 295 300

Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg

305 310 315 320

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

325 330 335

Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp

340 345 350

Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys

355 360 365

Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser

370 375 380

Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser

385 390 395 400

Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser

405 410 415

Leu Ser Leu Ser Pro Gly Lys

420

<210> 20

<211> 438

<212> PRT

<213> Artificial Sequence

<400> 20

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro

1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg

20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val

35 40 45

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

50 55 60

Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn

65 70 75 80

Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln Ile

85 90 95

Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro

100 105 110

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

115 120 125

Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys

130 135 140

Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln

145 150 155 160

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

165 170 175

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln

180 185 190

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

195 200 205

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu

210 215 220

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

225 230 235 240

Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp

245 250 255

Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly

260 265 270

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

275 280 285

Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp

290 295 300

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

305 310 315 320

Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu

325 330 335

Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn

340 345 350

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

355 360 365

Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr

370 375 380

Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys

385 390 395 400

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys

405 410 415

Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu

420 425 430

Ser Leu Ser Pro Gly Lys

435

<210> 21

<211> 18

<212> PRT

<213> Artificial Sequence

<400> 21

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

1 5 10 15

Gly Phe

<210> 22

<211> 26

<212> PRT

<213> Artificial Sequence

<400> 22

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

Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
20 25

<210> 23

<211> 33

<212> PRT

<213> Artificial Sequence

<400> 23

Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val
1 5 10 15

Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr
20 25 30

Glu

<210> 24

<211> 32

<212> PRT

<213> Artificial Sequence

<400> 24

Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser
1 5 10 15

Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala
20 25 30

<210> 25

<211> 142

<212> PRT

<213> Artificial Sequence

<400> 25

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

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

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
130 135 140

<210> 26
<211> 1137
<212> DNA
<213> Artificial Sequence
<400> 26

atgctgctgt ttaacaaagt gactctggca gacgcaggct ttatcaagca gtacggagac 60
tgtctcgggg acattgcagc cggcggcgga ggctcatctt tcattgagga cctgctgttc 120
aacaaggtca ctctggcaga tgccggattc ggaggagggg gatctccagc tatctgccat 180
gacggaaagg ctcattttcc tcgggagggt gtgtttgtgt ccaacggaac ccattggttc 240
gtcacacagc gcaacttcta tgaaggaggg gggggctcca gcttcatcga ggacctgctc 300

tttaacaaag tgaccctggc cgatgctgga tttgggggag ggggatccct gctgttcaac 360
aaagttacac tggccgacgc aggcttcatc aaacagtacg gcgattgttt aggggacatc 420

gccgctggcg gcggaggatc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 480

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 540

accctcatga tctctcgcac accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 600

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 660

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtgctg 720

caccaggatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 780

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 840

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 900

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 960

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1020

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgatgcac 1080

gaagctctgc acaatcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1137

<210> 27

<211> 104

<212> PRT

<213> Artificial Sequence

<

400> 27

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

Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
20 25 30

Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val
35 40 45

Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly
50 55 60

Asp Ile Ala Ala Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp Leu Leu
65 70 75 80

Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly
85 90 95

Asp Cys Leu Gly Asp Ile Ala Ala
100

<210> 28

<211> 315

<212> DNA

<213> Artificial Sequence

<400> 28

atgtccttca ttgaggacct gctgtttaat aaggtgaccc tggccgacgc tgggttcatc 60

aaacagtatg gagattgtct gggagatatt gcagcaggcg ggggcggcag cagctttatt 120

gaggacctcc tgttcaacaa ggtgaccctt gccgacgcag ggtttattaa gcagtatggc 180

gactgtctgg gagacattgc agccggcggc ggcgggtctt cttttatcga ggacctgctg 240

ttcaacaagg tgacactggc cgacgcaggc tttattaagc agtacgggga ctgcctggga 300

gacattgccg cctga 315

<210> 29

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 29

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140
ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200
gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260
tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320
ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380
ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440
aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560
ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620
ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680
ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740
accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800
ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860
ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980
gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040
cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100
gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160
tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220
tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280
acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 30

<211> 3627

<212> DNA

<213> Artificial Sequence

<400> 30

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtga 3627

<210> 31
<211> 2046
<212> DNA
<213> Artificial Sequence
<400> 31

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60
aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120
aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

ccttga 2046

<210> 32

<211> 1770

<212> DNA

<213> Artificial Sequence

<400> 32

atgtccgttg ctagtcagtc tattattgcc tataccatga gcctcggagc tgagaatagc 60

gtggcctact ccaataattc catcgcaatc cctactaact tcactatttc tgtgaccacc 120

gagatcctgc ctgtgtctat gactaagact agcgttgatt gtaccatgta tatttgtggc 180

gactctaccg aatgttctaa cctgctgctt cagtacggct cattttgcac acagctgaac 240

agagccctga ctgggatcgc tgtggagcag gacaagaaca cacaggaggt gtttgcacag 300

gtgaagcaga tctataagac ccctcctatt aaggatttcg gcggattcaa tttctcacag 360

attctgccag accccagtaa gccttccaag aggagcttca tcgaggatct cctgtttaac 420

aaggtgaccc tggcagacgc cggctttatt aagcaatatg gggattgcct gggcgacatt 480

gctgccagag acctgatttg cgcccagaaa ttcaatggcc tcacagtgct gccacctctg 540

ctgaccgacg agatgatcgc tcaatacact agcgcactgc tggccggaac catcacatca 600

ggctggacct tcggggccgg agcagcactg cagattccat tcgccatgca gatggcctat 660

agattcaacg gcattggcgt cacacagaac gtgctgtacg aaaaccagaa gctcatcgct 720

aaccagttta attccgcaat tggaaagatc caagattcac tcagctcaac cgcctctgca 780

ctcggaaagc tgcaggacgt ggtcaaccag aatgctcagg ccctgaacac actcgtcaag 840

cagctgtcct ctaactttgg cgctatcagc tccgttctga acgacattct gagccgcctg 900

gataaggtgg aggctgaagt ccagattgac cgcctgatta ccggccggct gcagtctctg 960

caaacatacg tgacccagca gctgatcaga gcagccgaga tccgggcatc cgcaaatctg 1020

gcagcaacta agatgagcga atgcgtgctg ggccagtcca agcgggtgga cttttgtggc 1080

aagggctacc acctgatgag cttcccccag agcgccccac atggcgttgt ttttctgcac 1140

gtgacctatg tccctgctca ggaaaagaac tttacaactg ctcctgctat ctgccatgac 1200

ggcaaggccc acttcccacg ggagggagtg tttgtgtcca atggcacaca ctggttcgtg 1260

acccagagga acttctatga accccagatc atcaccactg acaatacctt cgtgtctgga 1320

aattgcgacg tcgtgatcgg catcgttaac aacaccgtgt acgaccctct ccagccagag 1380

ctggactcct ttaaggagga actggataag tattttaaga accacacaag cccagatgtg 1440

gatctcgggg acatctccgg aattaacgcc tccgtggtga atatccagaa ggagattgac 1500

cgcctaaatg aagttgccaa gaacctcaat gagtctctga ttgatctgca ggaactgggc 1560

aagtatgagc agtatatcaa atggccctgg tacatttggc tggggtttat cgccggactg 1620

attgccatcg tcatggtgac catcatgctg tgttgcatga cctcctgttg ttcctgtctg 1680

aagggctgct gtagttgcgg ctcttgctgt aaattcgacg aagatgatag cgagcccgtg 1740

ctgaagggcg tgaagctgca ttatacctga 1770

<210> 33

<211> 1185

<212> DNA

<213> Artificial Sequence

<400> 33

atgagcttca tcgaggatct cctgtttaac aaggtgaccc tggcagacgc cggctttatt 60

aagcaatatg gggattgcct gggcgacatt gctgccagag acctgatttg cgcccagaaa 120

ttcaatggcc tcacagtgct gccacctctg ctgaccgacg agatgatcgc tcaatacact 180

agcgcactgc tggccggaac catcacatca ggctggacct tcggggccgg agcagcactg 240

cagattccat tcgccatgca gatggcctat agattcaacg gcattggcgt cacacagaac 300

gtgctgtacg aaaaccagaa gctcatcgct aaccagttta attccgcaat tggaaagatc 360

caagattcac tcagctcaac cgcctctgca ctcggaaagc tgcaggacgt ggtcaaccag 420

aatgctcagg ccctgaacac actcgtcaag cagctgtcct ctaactttgg cgctatcagc 480

tccgttctga acgacattct gagccgcctg gataaggtgg aggctgaagt ccagattgac 540

cgcctgatta ccggccggct gcagtctctg caaacatacg tgacccagca gctgatcaga 600

gcagccgaga tccgggcatc cgcaaatctg gcagcaacta agatgagcga atgcgtgctg 660

ggccagtcca agcgggtgga cttttgtggc aagggctacc acctgatgag cttcccccag 720

agcgccccac atggcgttgt ttttctgcac gtgacctatg tccctgctca ggaaaagaac 780

tttacaactg ctcctgctat ctgccatgac ggcaaggccc acttcccacg ggagggagtg 840

tttgtgtcca atggcacaca ctggttcgtg acccagagga acttctatga accccagatc 900

atcaccactg acaatacctt cgtgtctgga aattgcgacg tcgtgatcgg catcgttaac 960

aacaccgtgt acgaccctct ccagccagag ctggactcct ttaaggagga actggataag 1020

tattttaaga accacacaag cccagatgtg gatctcgggg acatctccgg aattaacgcc 1080

tccgtggtga atatccagaa ggagattgac cgcctaaatg aagttgccaa gaacctcaat 1140

gagtctctga ttgatctgca ggaactgggc aagtatgagc agtaa 1185

<210> 34

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 34

atgcctaata tcactaacct gtgtcctttc ggtgaagtgt tcaacgccac caggtttgct 60

agcgtgtatg cctggaacag gaagaggatc tctaactgcg tcgccgacta ttccgtgctg 120

tataacagcg cctccttctc cacattcaaa tgctatggag tgagcccgac aaaactgaac 180

gatctctgct ttacaaatgt ctacgccgac tcttttgtga tcagagggga cgaggtccgg 240

cagatcgcac caggacagac aggcaagatt gctgactaca actataagct gcctgacgac 300

ttcacaggat gtgtgatcgc atggaactca aacaatctgg actccaaagt cgggggcaac 360

tataattacc tgtatcgcct gttccggaag tccaacctga agcccttcga gagggacatc 420

agtacagaga tctatcaggc tggctccacc ccttgcaatg gcgtcgaagg ctttaattgt 480

tattttcccc tgcagtctta cgggtttcag cctactaatg gagttgggta ccagccatac 540

agagtggtcg tgctcagctt cgagctcctg catgctccat aa 582

<210> 35

<211> 624

<212> DNA

<213> Artificial Sequence

<400

> 35

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60

ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180

tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240

gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300

acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360

gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc ataa 624

<210> 36

<211> 1317

<212> DNA

<213> Artificial Sequence

<400> 36

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60

ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180

tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240

gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300

acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360

gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 660

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 720

accctcatga tctctcgcac accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 780

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 840

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtgctg 900

caccaggatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 960

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 1020

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 1080

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 1140

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1200

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgatgcac 1260

gaagctctgc acaatcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1317

<210> 37

<211> 45

<212> DNA

<213> Artificial Sequence

<400> 37

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgt 45

<210> 38

<211> 33

<212> PRT

<213> Artificial Sequence

<400> 38

Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
1 5 10 15

Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
20 25 30

Ser

<210> 39

<211> 99

<212> RNA

<213> Artificial Sequence

<400> 39

auggccacug gaucaagaac cucacugcug cucgcuuuug gacugcuuug ccugcccugg 60

uugcaagaag gaucggcuuu cccgaccauc ccacucucc 99

<210> 40

<211> 33

<212> PRT

<213> Artificial Sequence

<400> 40

Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
1 5 10 15

Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
20 25 30

Ser

<210> 41

<211> 99

<212> RNA

<213> Artificial Sequence

<400> 41

auggcaacug gaucaagaac cucccuccug cucgcauucg gccugcucug ucucccaugg 60

cuccaagaag gaagcgcguu ccccacuauc ccccucucg 99

<210> 42

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 42

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 43

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 43

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180
gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600
caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 44
<211> 3822
<212> DNA
<213> Artificial Sequence
<400> 44

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60
aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120
aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240
aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300
attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360
aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420
ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 45
<211> 3627
<212> DNA
<213> Artificial Sequence
<400> 45

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980
gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040
cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100
gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160
tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220
tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280
acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtga 3627

<210> 46

<211> 3708

<212> DNA

<213> Artificial Sequence

<400> 46

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680
ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740
accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800
ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860
ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920
aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980
gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100
gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160
tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220
tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280
acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340
gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760
aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180
gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcaggggtac attcccgagg ctcctaggga cggccaggca 3660

tacgtgcgca aagacggcga gtgggtgctg ctgtccacat tcctgtaa 3708

<210> 47

<211> 1308

<212> DNA

<213> Artificial Sequence

<400> 47

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540

ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 600

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200

caggaactgg gcaagtatga gcaggggtac attcccgagg ctcctaggga cggccaggca 1260

tacgtgcgca aagacggcga gtgggtgctg ctgtccacat tcctgtaa 1308

<210> 48

<211> 3708

<212> DNA

<213> Artificial Sequence

<400> 48

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600
caggaactgg gcaagtatga gcaggggtac attcccgagg ctcctaggga cggccaggca 3660
tacgtgcgca aagacggcga gtgggtgctg ctgtccacat tcctgtaa 3708

<210> 49
<211> 3708
<212> DNA
<213> Artificial Sequence
<400> 49
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60
aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120
aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180
aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240
aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300
attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420
ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480
agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540
ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600
ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660
agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720
ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780
ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900
tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960
cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020
gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcaggggtac attcccgagg ctcctaggga cggccaggca 3660

tacgtgcgca aagacggcga gtgggtgctg ctgtccacat tcctgtaa 3708

<210> 50

<211> 1275

<212> DNA

<213> Artificial Sequence

<400> 50

atgcctaata tcactaacct gtgtcctttc ggtgaagtgt tcaacgccac caggtttgct 60

agcgtgtatg cctggaacag gaagaggatc tctaactgcg tcgccgacta ttccgtgctg 120

tataacagcg cctccttctc cacattcaaa tgctatggag tgagcccgac aaaactgaac 180

gatctctgct ttacaaatgt ctacgccgac tcttttgtga tcagagggga cgaggtccgg 240

cagatcgcac caggacagac aggcaagatt gctgactaca actataagct gcctgacgac 300

ttcacaggat gtgtgatcgc atggaactca aacaatctgg actccaaagt cgggggcaac 360

tataattacc tgtatcgcct gttccggaag tccaacctga agcccttcga gagggacatc 420

agtacagaga tctatcaggc tggctccacc ccttgcaatg gcgtcgaagg ctttaattgt 480

tattttcccc tgcagtctta cgggtttcag cctactaatg gagttgggta ccagccatac 540

agagtggtcg tgctcagctt cgagctcctg catgctccac ctaagtcctg cgacaaaacc 600

catacatgtc caccatgccc agctcctgaa ctgctcggcg ggcctagtgt tttcctcttc 660

cctcctaagc ccaaggatac cctcatgatc tctcgcacac cagaagtgac ctgcgtggtc 720

gtggatgtct ctcacgagga tcctgaagtg aagtttaact ggtatgtcga cggagtggaa 780

gtgcacaacg ccaagacaaa gccaagagaa gaacaataca attctactta tagggtggtg 840

tctgtgctga cagtgctgca ccaggattgg ctgaatggaa aagaatataa gtgtaaggtc 900

tctaacaagg ccctgcccgc tccaattgag aagacaattt ccaaggccaa ggggcagcct 960

cgggaacctc aggtgtacac actgccccca tccagggatg aactgactaa aaatcaggtg 1020

tctctgacat gcctggtgaa agggttttat ccaagtgaca ttgctgtgga gtgggagtct 1080

aatgggcagc ctgaaaataa ctacaagacc acaccaccag tgctcgatag cgacgggtct 1140

ttctttctgt attctaaact gaccgtggat aaatctcggt ggcagcaggg aaacgtgttt 1200

tcttgctcag tgatgcacga agctctgcac aatcactata cacagaaatc cctgtccctg 1260

tctccaggca aataa 1275

<210> 51

<211> 156

<212> PRT

<213> Artificial Sequence

<400> 51

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Leu
1 5 10 15

Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr
20 25 30

Gly Asp Cys Leu Gly Asp Ile Ala Ala Gly Gly Gly Gly Ser Ser Phe
35 40 45

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

50 55 60

Gly Gly Gly Gly Ser Pro Ala Ile Cys His Asp Gly Lys Ala His Phe
65 70 75 80

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

Gln Arg Asn Phe Tyr Glu Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp
100 105 110

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

115 120 125

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

130 135 140

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

145 150 155

<210> 52

<211> 471

<212> DNA

<213> Artificial Sequence

<400> 52

atgttcgtgt tcctggtgct gctgccactg gtttcctccc agtgtctgct gtttaacaag 60

gttacactgg cagacgccgg cttcatcaag cagtatgggg actgtctggg cgatatcgcc 120

gctggcggcg gaggatctag cttcattgag gacctgctgt tcaacaaagt gactctggct 180

gacgccggat ttggcggagg agggtctcct gccatttgtc atgacgggaa ggctcatttc 240

cctagggagg gggtttttgt ctccaatgga actcactggt tcgtgaccca aagaaacttc 300

tatgagggag gtggcggatc ctcttttatc gaggacctgc tgtttaacaa ggtcactctg 360

gccgatgcag gcttcggagg aggagggtct ctgctgttca acaaagttac tctggcagat 420

gctgggttca ttaagcagta cggcgactgt ctgggcgata ttgccgcctg a 471

<210> 53

<211> 118

<212> PRT

<213> Artificial Sequence

<400

> 53

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser

1 5 10 15

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

20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Gly Gly

35 40 45

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

50 55 60

Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile

65 70 75 80

Ala Ala Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp Leu Leu Phe Asn
85 90 95
Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys
100 105 110
Leu Gly Asp Ile Ala Ala
115

<210> 54
<211> 357
<212> DNA
<213> Artificial Sequence

<400> 54
atgttcgtgt tcctggtcct gctacccctg gtgtcctctc agtgctcctt cattgaggac 60
ctgctgttta ataaggtgac cctggccgac gctgggttca tcaaacagta tggagattgt 120
ctgggagata ttgcagcagg cgggggcggc agcagcttta ttgaggacct cctgttcaac 180
aaggtgaccc ttgccgacgc agggtttatt aagcagtatg gcgactgtct gggagacatt 240
gcagccggcg gcggcgggtc ttcttttatc gaggacctgc tgttcaacaa ggtgacactg 300
gccgacgcag gctttattaa gcagtacggg gactgcctgg gagacattgc cgcctga 357

<210> 55
<211> 392
<212> PRT
<213> Artificial Sequence

<400> 55
Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Leu
1 5 10 15
Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr
20 25 30
Gly Asp Cys Leu Gly Asp Ile Ala Ala Gly Gly Gly Gly Ser Ser Phe
35 40 45
Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe
50 55 60

Gly Gly Gly Gly Ser Pro Ala Ile Cys His Asp Gly Lys Ala His Phe
65 70 75 80

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

Gln Arg Asn Phe Tyr Glu Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp
100 105 110

Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Gly Gly Gly
115 120 125

Gly Ser Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile
130 135 140

Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Gly Gly Gly Gly
145 150 155 160

Ser Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala
165 170 175

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
180 185 190

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
195 200 205

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
210 215 220

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
225 230 235 240

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
245 250 255

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
260 265 270

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

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr
290 295 300

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

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr

325 330 335

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr

340 345 350

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe

355 360 365

Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys

370 375 380

Ser Leu Ser Leu Ser Pro Gly Lys

385 390

<210> 56

<211> 1179

<212> DNA

<213> Artificial Sequence

<400> 56

atgtttgtgt tcctcgttct gctgcctctg gtgagctccc agtgtctgct gtttaacaaa 60

gtgactctgg cagacgcagg ctttatcaag cagtacggag actgtctcgg ggacattgca 120

gccggcggcg gaggctcatc tttcattgag gacctgctgt tcaacaaggt cactctggca 180

gatgccggat tcggaggagg gggatctcca gctatctgcc atgacggaaa ggctcatttt 240

cctcgggagg gtgtgtttgt gtccaacgga acccattggt tcgtcacaca gcgcaacttc 300

tatgaaggag gggggggctc cagcttcatc gaggacctgc tctttaacaa agtgaccctg 360

gccgatgctg gatttggggg agggggatcc ctgctgttca acaaagttac actggccgac 420

gcaggcttca tcaaacagta cggcgattgt ttaggggaca tcgccgctgg cggcggagga 480

tcacctaagt cctgcgacaa aacccataca tgtccaccat gcccagctcc tgaactgctc 540

ggcgggccta gtgttttcct cttccctcct aagcccaagg ataccctcat gatctctcgc 600

acaccagaag tgacctgcgt ggtcgtggat gtctctcacg aggatcctga agtgaagttt 660

aactggtatg tcgacggagt ggaagtgcac aacgccaaga caaagccaag agaagaacaa 720

tacaattcta cttatagggt ggtgtctgtg ctgacagtgc tgcaccagga ttggctgaat 780

ggaaaagaat ataagtgtaa ggtctctaac aaggccctgc ccgctccaat tgagaagaca 840

atttccaagg ccaaggggca gcctcgggaa cctcaggtgt acacactgcc cccatccagg 900

gatgaactga ctaaaaatca ggtgtctctg acatgcctgg tgaaagggtt ttatccaagt 960

gacattgctg tggagtggga gtctaatggg cagcctgaaa ataactacaa gaccacacca 1020

ccagtgctcg atagcgacgg gtctttcttt ctgtattcta aactgaccgt ggataaatct 1080

cggtggcagc agggaaacgt gttttcttgc tcagtgatgc acgaagctct gcacaatcac 1140

tatacacaga aatccctgtc cctgtctcca ggcaaataa 1179

<210> 57

<211> 354

<212> PRT

<213> Artificial Sequence

<400> 57

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Gly Gly
35 40 45

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

50 55 60

Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile
65 70 75 80

Ala Ala Gly Gly Gly Gly Ser Ser Phe Ile Glu Asp Leu Leu Phe Asn
85 90 95

Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys
100 105 110

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

115 120 125

Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly
130 135 140

Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile
145 150 155 160

Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu
165 170 175

Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His

180 185 190

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

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

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

<210> 58
<211> 1065
<212> DNA
<213> Artificial Sequence
<400> 58
atgttcgtgt tcctggtcct gctgcctctg gtgtcctctc agtgcagctt catcgaggac 60
ctgctcttta acaaggtgac tctcgcagat gctggcttca tcaagcagta cggagactgc 120

cttggagaca tcgctgcagg cggaggggc agcagtttca tcgaggacct gctgtttaac 180
aaggtgaccc tggccgacgc cgggttcatt aagcaatacg gcgattgtct gggagacatc 240
gcagctgggg gagggggggag ctcttttatt gaggacctgc tgttcaacaa ggtgactctg 300

gccgacgcag ggttcatcaa acagtatggg gactgtctgg gagatatcgc agccggggga 360

ggaggctccc ctaagtcctg cgacaaaacc catacatgtc caccatgccc agctcctgaa 420

ctgctcggcg ggcctagtgt tttcctcttc cctcctaagc ccaaggatac cctcatgatc 480

tctcgcacac cagaagtgac ctgcgtggtc gtggatgtct ctcacgagga tcctgaagtg 540

aagtttaact ggtatgtcga cggagtggaa gtgcacaacg ccaagacaaa gccaagagaa 600

gaacaataca attctactta tagggtggtg tctgtgctga cagtgctgca ccaggattgg 660

ctgaatggaa aagaatataa gtgtaaggtc tctaacaagg ccctgcccgc tccaattgag 720

aagacaattt ccaaggccaa ggggcagcct cgggaacctc aggtgtacac actgccccca 780

tccagggatg aactgactaa aaatcaggtg tctctgacat gcctggtgaa agggttttat 840

ccaagtgaca ttgctgtgga gtgggagtct aatgggcagc ctgaaaataa ctacaagacc 900

acaccaccag tgctcgatag cgacgggtct ttctttctgt attctaaact gaccgtggat 960

aaatctcggt ggcagcaggg aaacgtgttt tcttgctcag tgatgcacga agctctgcac 1020

aatcactata cacagaaatc cctgtccctg tctccaggca aataa 1065

<210> 59

<211> 666

<212> DNA

<213> Artificial Sequence

<400> 59

atggcagaca acggtactat taccgttgag gagcttaaac aactcctgga acaatggaac 60

ctagtaatag gtttcctatt cctagcctgg attatgttac tacaatttgc ctattctaat 120

cggaacaggt ttttgtacat aataaagctt gttttcctct ggctcttgtg gccagtaaca 180

cttgcttgtt ttgtgcttgc tgctgtctac agaattaatt gggtgactgg cgggattgcg 240

attgcaatgg cttgtattgt aggcttgatg tggcttagct acttcgttgc ttccttcagg 300

ctgtttgctc gtacccgctc aatgtggtca ttcaacccag aaacaaacat tcttctcaat 360

gtgcctctcc gggggacaat tgtgaccaga ccgctcatgg aaagtgaact tgtcattggt 420

gctgtgatca ttcgtggtca cttgcgaatg gccggacact ccctagggcg ctgtgacatt 480

aaggacctgc caaaagagat cactgtggct acatcacgaa cgctttctta ttacaaatta 540

ggagcgtcgc agcgtgtagg cactgattca ggttttgctg catacaaccg ctaccgtatt 600

ggaaactata aattaaatac agaccacgcc ggtagcaacg acaatattgc tttgctagta 660

cagtaa 666

<210> 60

<211> 222

<212> PRT

<213> Artificial Sequence

<400

> 60

Met Ala Asp Ser Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Lys Leu
1 5 10 15

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

Cys Leu Leu Gln Phe Ala Tyr Ala Asn Arg Asn Arg Phe Leu Tyr Ile
35 40 45

Ile Lys Leu Ile Phe Leu Trp Leu Leu Trp Pro Val Thr Leu Ala Cys
50 55 60

Phe Val Leu Ala Ala Val Tyr Arg Ile Asn Trp Ile Thr Gly Gly Ile
65 70 75 80

Ala Ile Ala Met Ala Cys Leu Val Gly Leu Met Trp Leu Ser Tyr Phe
85 90 95

Ile Ala Ser Phe Arg Leu Phe Ala Arg Thr Arg Ser Met Trp Ser Phe
100 105 110

Asn Pro Glu Thr Asn Ile Leu Leu Asn Val Pro Leu His Gly Thr Ile
115 120 125

Leu Thr Arg Pro Leu Leu Glu Ser Glu Leu Val Ile Gly Ala Val Ile
130 135 140

Leu Arg Gly His Leu Arg Ile Ala Gly His His Leu Gly Arg Cys Asp
145 150 155 160

Ile Lys Asp Leu Pro Lys Glu Ile Thr Val Ala Thr Ser Arg Thr Leu
165 170 175

Ser Tyr Tyr Lys Leu Gly Ala Ser Gln Arg Val Ala Gly Asp Ser Gly
180 185 190

Phe Ala Ala Tyr Ser Arg Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr
195 200 205

Asp His Ser Ser Ser Ser Asp Asn Ile Ala Leu Leu Val Gln
210 215 220

<210> 61

<211> 231

<212> DNA

<213> Artificial Sequence

<400> 61

atgtactcat tcgtttcgga agaaacaggt acgttaatag ttaatagcgt acttcttttt 60

cttgctttcg tggtattctt gctagtcaca ctagccatcc ttactgcgct tcgattgtgt 120

gcgtactgct gcaatattgt taacgtgagt ttagtaaaac caacggttta cgtctactcg 180

cgtgttaaaa atctgaactc ttctgaagga gttcctgatc ttctggtcta a 231

<210> 62

<211> 75

<212> PRT

<213> Artificial Sequence

<400> 62

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

1 5 10 15

Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala

20 25 30

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

35 40 45

Val Ser Leu Val Lys Pro Ser Phe Tyr Val Tyr Ser Arg Val Lys Asn

50 55 60

Leu Asn Ser Ser Arg Val Pro Asp Leu Leu Val

65 70 75

<210> 63

<211> 1260

<212> DNA

<213> Artificial Sequence

<400> 63

atgtctgata atggacccca aaatcagcga aatgcacccc gcattacgtt tggtggaccc 60

tcagattcaa ctggcagtaa ccagaatgga gaacgcagtg gggcgcgatc aaaacaacgt 120

cggccccaag gtttacccaa taatactgcg tcttggttca ccgctctcac tcaacatggc 180

aaggaagacc ttaaattccc tcgaggacaa ggcgttccaa ttaacaccaa tagcagtcca 240

gatgaccaaa ttggctacta ccgaagagct accagacgaa ttcgtggtgg tgacggtaaa 300

atgaaagatc tcagtccaag atggtatttc tactacctag gaactgggcc agaagctgga 360

cttccctatg gtgctaacaa agacggcatc atatgggttg caactgaggg agccttgaat 420

acaccaaaag atcacattgg cacccgcaat cctgctaaca atgctgcaat cgtgctacaa 480

cttcctcaag gaacaacatt gccaaaaggc ttctacgcag aagggagcag aggcggcagt 540

caagcctctt ctcgttcctc atcacgtagt cgcaacagtt caagaaattc aactccaggc 600

agcagtaggg gaacttctcc tgctagaatg gctggcaatg gcggtgatgc tgctcttgct 660

ttgctgctgc ttgacagatt gaaccagctt gagagcaaaa tgtctggtaa aggccaacaa 720

caacaaggcc aaactgtcac taagaaatct gctgctgagg cttctaagaa gcctcggcaa 780

aaacgtactg ccactaaagc atacaatgta acacaagctt tcggcagacg tggtccagaa 840

caaacccaag gaaattttgg ggaccaggaa ctaatcagac aaggaactga ttacaaacat 900

tggccgcaaa ttgcacaatt tgcccccagc gcttcagcgt tcttcggaat gtcgcgcatt 960

ggcatggaag tcacaccttc gggaacgtgg ttgacctaca caggtgccat caaattggat 1020

gacaaagatc caaatttcaa agatcaagtc attttgctga ataagcatat tgacgcatac 1080

aaaacattcc caccaacaga gcctaaaaag gacaaaaaga agaaggctga tgaaactcaa 1140

gccttaccgc agagacagaa gaaacagcaa actgtgactc ttcttcctgc tgcagatttg 1200

gatgatttct ccaaacaatt gcaacaatcc atgagcagtg ctgactcaac tcaggcctaa 1260

<210> 64

<211> 419

<212> PRT

<213> Artificial Sequence

<400> 64

Met Ser Asp Asn Gly Pro Gln Asn Gln Arg Asn Ala Pro Arg Ile Thr
1 5 10 15

Phe Gly Gly Pro Ser Asp Ser Thr Gly Ser Asn Gln Asn Gly Glu Arg
20 25 30

Ser Gly Ala Arg Ser Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn Asn
35 40 45

Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Asp Leu
50 55 60

Lys Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Ser Pro

65 70 75 80

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

85 90 95

Gly Asp Gly Lys Met Lys Asp Leu Ser Pro Arg Trp Tyr Phe Tyr Tyr

100 105 110

Leu Gly Thr Gly Pro Glu Ala Gly Leu Pro Tyr Gly Ala Asn Lys Asp

115 120 125

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

130 135 140

His Ile Gly Thr Arg Asn Pro Ala Asn Asn Ala Ala Ile Val Leu Gln

145 150 155 160

Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly Ser

165 170 175

Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg Ser Arg Asn

180 185 190

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

195 200 205

Arg Met Ala Gly Asn Gly Gly Asp Ala Ala Leu Ala Leu Leu Leu Leu

210 215 220

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

225 230 235 240

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

245 250 255

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

260 265 270

Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly Asp

275 280 285

Gln Glu Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln Ile

290 295 300

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

305 310 315 320

Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr Thr Gly Ala
325 330 335
Ile Lys Leu Asp Asp Lys Asp Pro Asn Phe Lys Asp Gln Val Ile Leu
340 345 350

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

Thr Gln Ala

<210> 65
<211> 1227
<212> DNA
<213> Artificial Sequence
<400> 65
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540
ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 600
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200
caggaactgg gcaagtatga gcagtaa 1227

<210> 66
<211> 803
<212> PRT
<213> Artificial Sequence
<400> 66

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

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
145 150 155 160
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
165 170 175
Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln

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

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

305 310 315 320
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
325 330 335
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
340 345 350
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
355 360 365
Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

370 375 380
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu

385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln Gln Cys Ser Phe Ile Glu Asp Leu

405 410 415

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

420 425 430

Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys Ala Gln

435 440 445

Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp Glu Met

450 455 460

Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr Ser Gly

465 470 475 480

Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala Met Gln

485 490 495

Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val Leu Tyr

500 505 510

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

515 520 525

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

530 535 540

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

545 550 555 560

Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp Ile Leu

565 570 575

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

580 585 590

Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile

595 600 605

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

610 615 620

Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys

625 630 635 640

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

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

755 760 765
Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val
770 775 780
Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys
785 790 795 800
Tyr Glu Gln

<210> 67
<211> 1812
<212> DNA
<213> Artificial Sequence
<400> 67
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60

tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120
tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180
atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240
aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300

gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360

acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420

aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480

gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540

tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600

gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660

ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720

gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780

attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840

gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900

ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggataaggt ggaggctgaa 960

gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020

cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140

agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200

caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260

cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320

gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380

ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440

gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560

aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcagtatatc 1620

aaatggccct ggtacatttg gctggggttt atcgccggac tgattgccat cgtcatggtg 1680

accatcatgc tgtgttgcat gacctcctgt tgttcctgtc tgaagggctg ctgtagttgc 1740

ggctcttgct gtaaattcga cgaagatgat agcgagcccg tgctgaaggg cgtgaagctg 1800

cattatacct ga 1812

<210> 68

<211> 603

<212> PRT

<213> Artificial Sequence

<400> 68

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

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

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

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

245 250 255

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala

260 265 270

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

275 280 285

Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys

355 360 365

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln

370 375 380

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala

385 390 395 400

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

405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp

420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp

435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn

450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu

465 470 475 480

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

485 490 495

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu

500 505 510

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

515 520 525

Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp

530 535 540

Tyr Ile Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val

545 550 555 560

Thr Ile Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly

565 570 575

Cys Cys Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu

580 585 590

Pro Val Leu Lys Gly Val Lys Leu His Tyr Thr

595 600

<210> 69

<211> 1812

<212> DNA

<213> Artificial Sequence

<400> 69

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60

tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120

tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180

atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240

aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300

gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360

acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420

aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480

gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540

tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600

gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660

ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720

gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780

attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840

gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900

ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggatccccc agaggctgaa 960

gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020

cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140

agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200

caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260

cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320

gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380

ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440

gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560

aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcagtatatc 1620

aaatggccct ggtacatttg gctggggttt atcgccggac tgattgccat cgtcatggtg 1680

accatcatgc tgtgttgcat gacctcctgt tgttcctgtc tgaagggctg ctgtagttgc 1740

ggctcttgct gtaaattcga cgaagatgat agcgagcccg tgctgaaggg cgtgaagctg 1800

cattatacct ga 1812

<210> 70

<211> 603

<212> PRT

<213> Artificial Sequence

<400> 70

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu
20 25 30

Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe
35 40 45

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

50 55 60

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

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

Pro Leu Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu
195 200 205
Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu
210 215 220
Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly
225 230 235 240
Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln
245 250 255

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala
260 265 270
Ser Ala Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala
275 280 285
Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser
290 295 300

Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu
305 310 315 320

Val Gln Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr
325 330 335

Tyr Val Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala
340 345 350

Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
355 360 365

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln
370 375 380

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala
385 390 395 400

Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys
405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp
420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp
435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn
450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu
465 470 475 480

Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu
485 490 495

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu
500 505 510

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
515 520 525

Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp
530 535 540

Tyr Ile Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val

545 550 555 560

Thr Ile Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly

565 570 575

Cys Cys Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu

580 585 590

Pro Val Leu Lys Gly Val Lys Leu His Tyr Thr

595 600

<210> 71

<211> 1770

<212> DNA

<213> Artificial Sequence

<400> 71

atgtccgttg ctagtcagtc tattattgcc tataccatga gcctcggagc tgagaatagc 60

gtggcctact ccaataattc catcgcaatc cctactaact tcactatttc tgtgaccacc 120

gagatcctgc ctgtgtctat gactaagact agcgttgatt gtaccatgta tatttgtggc 180

gactctaccg aatgttctaa cctgctgctt cagtacggct cattttgcac acagctgaac 240

agagccctga ctgggatcgc tgtggagcag gacaagaaca cacaggaggt gtttgcacag 300

gtgaagcaga tctataagac ccctcctatt aaggatttcg gcggattcaa tttctcacag 360

attctgccag accccagtaa gccttccaag aggagcttca tcgaggatct cctgtttaac 420

aaggtgaccc tggcagacgc cggctttatt aagcaatatg gggattgcct gggcgacatt 480

gctgccagag acctgatttg cgcccagaaa ttcaatggcc tcacagtgct gccacctctg 540

ctgaccgacg agatgatcgc tcaatacact agcgcactgc tggccggaac catcacatca 600

ggctggacct tcggggccgg agcagcactg cagattccat tcgccatgca gatggcctat 660

agattcaacg gcattggcgt cacacagaac gtgctgtacg aaaaccagaa gctcatcgct 720

aaccagttta attccgcaat tggaaagatc caagattcac tcagctcaac cgcctctgca 780

ctcggaaagc tgcaggacgt ggtcaaccag aatgctcagg ccctgaacac actcgtcaag 840

cagctgtcct ctaactttgg cgctatcagc tccgttctga acgacattct gagccgcctg 900

gataaggtgg aggctgaagt ccagattgac cgcctgatta ccggccggct gcagtctctg 960

caaacatacg tgacccagca gctgatcaga gcagccgaga tccgggcatc cgcaaatctg 1020

gcagcaacta agatgagcga atgcgtgctg ggccagtcca agcgggtgga cttttgtggc 1080

aagggctacc acctgatgag cttcccccag agcgccccac atggcgttgt ttttctgcac 1140

gtgacctatg tccctgctca ggaaaagaac tttacaactg ctcctgctat ctgccatgac 1200

ggcaaggccc acttcccacg ggagggagtg tttgtgtcca atggcacaca ctggttcgtg 1260

acccagagga acttctatga accccagatc atcaccactg acaatacctt cgtgtctgga 1320

aattgcgacg tcgtgatcgg catcgttaac aacaccgtgt acgaccctct ccagccagag 1380

ctggactcct ttaaggagga actggataag tattttaaga accacacaag cccagatgtg 1440

gatctcgggg acatctccgg aattaacgcc tccgtggtga atatccagaa ggagattgac 1500

cgcctaaatg aagttgccaa gaacctcaat gagtctctga ttgatctgca ggaactgggc 1560

aagtatgagc agtatatcaa atggccctgg tacatttggc tggggtttat cgccggactg 1620

attgccatcg tcatggtgac catcatgctg tgttgcatga cctcctgttg ttcctgtctg 1680

aagggctgct gtagttgcgg ctcttgctgt aaattcgacg aagatgatag cgagcccgtg 1740

ctgaagggcg tgaagctgca ttatacctga 1770

<210> 72

<211> 589

<212> PRT

<213> Artificial Sequence

<400> 72

Met Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly

1 5 10 15

Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr

20 25 30

Asn Phe Thr Ile Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr

35 40 45

Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu

50 55 60

Cys Ser Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn

65 70 75 80

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

85 90 95

Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp

100 105 110

Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro

115 120 125

Ser Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu

130 135 140

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

195 200 205

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

260 265 270

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

325 330 335

Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln
340 345 350
Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe
355 360 365
Pro Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val

370 375 380

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp

385 390 395 400

Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr

405 410 415

His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr

420 425 430

Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile

435 440 445

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

450 455 460

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

465 470 475 480

Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln

485 490 495

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

500 505 510

Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp

515 520 525

Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val

530 535 540

Met Val Thr Ile Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu

545 550 555 560

Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp

565 570 575

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

580 585

<210> 73

<211> 1617

<212> DNA

<213> Artificial Sequence

<400> 73

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60

tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120

tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180

atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240

aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300

gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360

acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420

aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480

gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540

tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600

gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660

ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720

gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780

attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840

gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900

ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggataaggt ggaggctgaa 960

gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020

cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140

agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200

caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260

cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320

gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380

ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440

gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560

aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcagtga 1617

<210> 74

<211> 538

<212> PRT

<213> Artificial Sequence

<400> 74

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

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

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

Pro Leu Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu
195 200 205
Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu
210 215 220
Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly

225 230 235 240

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

245 250 255

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala

260 265 270

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

275 280 285

Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys

355 360 365

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln

370 375 380

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala

385 390 395 400

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

405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp

420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp

435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn

450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu

465 470 475 480

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

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
515 520 525
Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln
530 535

<210> 75
<211> 1698
<212> DNA
<213> Artificial Sequence
<400> 75

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60
tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120
tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180

atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240
aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300
gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360
acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420
aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480
gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540
tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600
gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660

ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720
gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780
attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840
gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900
ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggataaggt ggaggctgaa 960
gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020
cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140

agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200

caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260

cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320

gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380

ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440

gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560

aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcaggggtac 1620

attcccgagg ctcctaggga cggccaggca tacgtgcgca aagacggcga gtgggtgctg 1680

ctgtccacat tcctgtga 1698

<210> 76

<211> 565

<212> PRT

<213> Artificial Sequence

<400> 76

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu
20 25 30

Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe
35 40 45

Thr Ile Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr
50 55 60

Ser Val Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser
65 70 75 80

Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala
85 90 95

Leu Thr Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe
100 105 110

Ala Gln Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly
115 120 125

Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys
130 135 140

Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp
145 150 155 160
Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
165 170 175
Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro
180 185 190
Pro Leu Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu
195 200 205

Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu
210 215 220
Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly
225 230 235 240
Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln
245 250 255
Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala
260 265 270

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

Tyr Val Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala
340 345 350
Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
355 360 365
Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln

370 375 380

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala
385 390 395 400

Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys
405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp
420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp
435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn
450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu
465 470 475 480

Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu
485 490 495

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu
500 505 510

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
515 520 525

Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala
530 535 540

Pro Arg Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu
545 550 555 560

Leu Ser Thr Phe Leu
565

<210> 77

<211> 1617

<212> DNA

<213> Artificial Sequence

<400> 77

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60

tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120

tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180
atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240
aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300
gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360
acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420
aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480
gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540
tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600

gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660
ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720
gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780
attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840
gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900
ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggatccccc agaggctgaa 960
gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020
cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140
agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200
caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260
cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320
gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380
ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440
gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560
aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcagtga 1617

<210> 78

<211> 538

<212> PRT

<213> Artificial Sequence

<400> 78

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15
Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu
20 25 30
Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe

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

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

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

225 230 235 240
Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln

245 250 255

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala

260 265 270

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

275 280 285

Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys

355 360 365

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln

370 375 380

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala

385 390 395 400

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

405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp

420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp

435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn

450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu

465 470 475 480

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

485 490 495

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu
500 505 510

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
515 520 525

Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln
530 535

<210> 79

<211> 1698

<212> DNA

<213> Artificial Sequence

<400> 79

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgttccgt tgctagtcag 60

tctattattg cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat 120

tccatcgcaa tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct 180

atgactaaga ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct 240

aacctgctgc ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc 300

gctgtggagc aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag 360

acccctccta ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt 420

aagccttcca agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac 480

gccggcttta ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt 540

tgcgcccaga aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc 600

gctcaataca ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc 660

ggagcagcac tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc 720

gtcacacaga acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca 780

attggaaaga tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac 840

gtggtcaacc agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt 900

ggcgctatca gctccgttct gaacgacatt ctgagccgcc tggatccccc agaggctgaa 960

gtccagattg accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag 1020

cagctgatca gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc 1080

gaatgcgtgc tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg 1140

agcttccccc agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct 1200

caggaaaaga actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca 1260

cgggagggag tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat 1320

gaaccccaga tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc 1380

ggcatcgtta acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag 1440

gaactggata agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc 1500

ggaattaacg cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc 1560

aagaacctca atgagtctct gattgatctg caggaactgg gcaagtatga gcaggggtac 1620

attcccgagg ctcctaggga cggccaggca tacgtgcgca aagacggcga gtgggtgctg 1680

ctgtccacat tcctgtga 1698

<210> 80

<211> 565

<212> PRT

<213> Artificial Sequence

<400> 80

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser

1 5 10 15

Val Ala Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu

20 25 30

Asn Ser Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe

35 40 45

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

50 55 60

Ser Val Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser

65 70 75 80

Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala

85 90 95

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

100 105 110

Ala Gln Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly

115 120 125

Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys

130 135 140

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

195 200 205
Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu
210 215 220
Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly
225 230 235 240
Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln
245 250 255
Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala

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

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

385 390 395 400

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

405 410 415

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp

420 425 430

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp

435 440 445

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn

450 455 460

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu

465 470 475 480

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

485 490 495

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu

500 505 510

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

515 520 525

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

530 535 540

Pro Arg Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu

545 550 555 560

Leu Ser Thr Phe Leu

565

<210> 81

<211> 1227

<212> DNA

<213> Artificial Sequence

<400> 81

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 600

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200

caggaactgg gcaagtatga gcagtga 1227

<210> 82

<211> 408

<212> PRT

<213> Artificial Sequence

<400> 82

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
35 40 45

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
50 55 60

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

65 70 75 80

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
85 90 95
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
100 105 110
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
115 120 125
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
130 135 140

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

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
210 215 220
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
225 230 235 240
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
245 250 255
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
260 265 270

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
275 280 285
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
290 295 300
Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
305 310 315 320

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
325 330 335

Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
340 345 350

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
355 360 365

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
370 375 380

Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln
405

<210> 83

<211> 1308

<212> DNA

<213> Artificial Sequence

<400> 83

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 600

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200

caggaactgg gcaagtatga gcaggggtac attcccgagg ctcctaggga cggccaggca 1260

tacgtgcgca aagacggcga gtgggtgctg ctgtccacat tcctgtga 1308

<210> 84

<211> 435

<212> PRT

<213> Artificial Sequence

<400> 84

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
35 40 45

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
50 55 60

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
65 70 75 80

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
85 90 95

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
100 105 110

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
115 120 125

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
130 135 140

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

145 150 155 160

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

225 230 235 240

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

245 250 255

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

260 265 270

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

275 280 285

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

290 295 300

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

305 310 315 320

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

325 330 335

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

340 345 350

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

355 360 365

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

370 375 380

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

385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln Gly Tyr Ile Pro Glu Ala Pro Arg

405 410 415

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

420 425 430

Thr Phe Leu

435

<210> 85

<211> 1422

<212> DNA

<213> Artificial Sequence

<400> 85

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 600

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 1260

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 1320

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 1380

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 1422

<210> 86

<211> 473

<212> PRT

<213> Artificial Sequence

<400> 86

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
35 40 45

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
50 55 60

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
65 70 75 80

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
85 90 95

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
100 105 110

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
115 120 125

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
130 135 140

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
145 150 155 160

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

225 230 235 240

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

245 250 255

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

260 265 270

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

275 280 285

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

290 295 300

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

305 310 315 320

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

325 330 335

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

340 345 350

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

355 360 365

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

370 375 380

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

385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

405 410 415

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

420 425 430

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

435 440 445

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

450 455 460

Leu Lys Gly Val Lys Leu His Tyr Thr

465 470

<210> 87

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 87

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tgcccaaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 88

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 88

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly

865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala

930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Pro Lys Val Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val

995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu

1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 89

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 89

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040
cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100
gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160
tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220
tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280
acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340
gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760
aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tgcccaaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180
gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600
caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 90

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 90

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205
Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

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

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

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

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700
Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720
Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735

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

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

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

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

930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Pro Lys Val Glu Ala Glu Val Gln
980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135

Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180

Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val
1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 91

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 91

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tgcctccacc cgaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 92
<211> 1273
<212> PRT
<213> Artificial Sequence
<400> 92

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335

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

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Pro Pro Pro Glu Ala Glu Val Gln
980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 93

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 93

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tgcctccacc cgaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 94
<211> 1273
<212> PRT
<213> Artificial Sequence
<400> 94

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95
Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110
Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125
Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160
Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

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

Leu Asn Asp Ile Leu Ser Arg Leu Pro Pro Pro Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

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

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215
Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val
1250 1255 1260
Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 95
<211> 1422
<212> DNA
<213> Artificial Sequence
<400> 95

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtagctt catcgaggat 60
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 120

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 180
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 240
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 300
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 360
aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 420
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 480
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 540
ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 600

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 660

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 720

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 780

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 840

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 900

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 960

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 1020

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 1080

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 1140

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 1200

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 1260

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 1320

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 1380

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 1422

<210> 96

<211> 473

<212> PRT

<213> Artificial Sequence

<400> 96

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Ser
1 5 10 15

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
20 25 30

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
35 40 45

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
50 55 60

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
65 70 75 80

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
85 90 95

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
100 105 110

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
115 120 125

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

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
195 200 205
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
210 215 220
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
225 230 235 240
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
245 250 255

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

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
325 330 335
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
340 345 350
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

355 360 365

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

370 375 380

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

385 390 395 400

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

405 410 415

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

420 425 430

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

435 440 445

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

450 455 460

Leu Lys Gly Val Lys Leu His Tyr Thr

465 470

<210> 97

<211> 1380

<212> DNA

<213> Artificial Sequence

<400> 97

atgagcttca tcgaggatct cctgtttaac aaggtgaccc tggcagacgc cggctttatt 60

aagcaatatg gggattgcct gggcgacatt gctgccagag acctgatttg cgcccagaaa 120

ttcaatggcc tcacagtgct gccacctctg ctgaccgacg agatgatcgc tcaatacact 180

agcgcactgc tggccggaac catcacatca ggctggacct tcggggccgg agcagcactg 240

cagattccat tcgccatgca gatggcctat agattcaacg gcattggcgt cacacagaac 300

gtgctgtacg aaaaccagaa gctcatcgct aaccagttta attccgcaat tggaaagatc 360

caagattcac tcagctcaac cgcctctgca ctcggaaagc tgcaggacgt ggtcaaccag 420

aatgctcagg ccctgaacac actcgtcaag cagctgtcct ctaactttgg cgctatcagc 480

tccgttctga acgacattct gagccgcctg gataaggtgg aggctgaagt ccagattgac 540

cgcctgatta ccggccggct gcagtctctg caaacatacg tgacccagca gctgatcaga 600

gcagccgaga tccgggcatc cgcaaatctg gcagcaacta agatgagcga atgcgtgctg 660

ggccagtcca agcgggtgga cttttgtggc aagggctacc acctgatgag cttcccccag 720

agcgccccac atggcgttgt ttttctgcac gtgacctatg tccctgctca ggaaaagaac 780

tttacaactg ctcctgctat ctgccatgac ggcaaggccc acttcccacg ggagggagtg 840

tttgtgtcca atggcacaca ctggttcgtg acccagagga acttctatga accccagatc 900

atcaccactg acaatacctt cgtgtctgga aattgcgacg tcgtgatcgg catcgttaac 960

aacaccgtgt acgaccctct ccagccagag ctggactcct ttaaggagga actggataag 1020

tattttaaga accacacaag cccagatgtg gatctcgggg acatctccgg aattaacgcc 1080

tccgtggtga atatccagaa ggagattgac cgcctaaatg aagttgccaa gaacctcaat 1140

gagtctctga ttgatctgca ggaactgggc aagtatgagc agtatatcaa atggccctgg 1200

tacatttggc tggggtttat cgccggactg attgccatcg tcatggtgac catcatgctg 1260

tgttgcatga cctcctgttg ttcctgtctg aagggctgct gtagttgcgg ctcttgctgt 1320

aaattcgacg aagatgatag cgagcccgtg ctgaagggcg tgaagctgca ttatacctga 1380

<210> 98

<211> 459

<212> PRT

<213> SARS-CoV-2

<400> 98

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

Ala Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
20 25 30

Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro
35 40 45

Pro Leu Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu
50 55 60

Ala Gly Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu
65 70 75 80

Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly
85 90 95

Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln
100 105 110

Phe Asn Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala

115 120 125

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

130 135 140

Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys

210 215 220

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln

225 230 235 240

Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala

245 250 255

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

260 265 270

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp

275 280 285

Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp

290 295 300

Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn

305 310 315 320

Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu

325 330 335

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

340 345 350

Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu

355 360 365

Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile
370 375 380

Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp
385 390 395 400

Tyr Ile Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val
405 410 415

Thr Ile Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly
420 425 430

Cys Cys Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu
435 440 445

Pro Val Leu Lys Gly Val Lys Leu His Tyr Thr
450 455

<210> 99

<211> 31

<212> PRT

<213> Artificial Sequence

<400> 99

Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala
1 5 10 15

Gly Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala
20 25 30

<210> 100

<211> 231

<212> PRT

<213> Artificial Sequence

<400> 100

Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro
1 5 10 15

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

Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val
35 40 45

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

Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Asp His His Asp
85 90 95
Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu
100 105 110
Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg
115 120 125
Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys
130 135 140

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

Cys Ser Val Met His Glu Ala Leu His Ser His Tyr Thr Gln Lys Ser
210 215 220
Leu Ser Leu Ser Pro Gly Lys
225 230

<210> 101
<211> 231
<212> PRT
<213> Artificial Sequence
<400> 101

Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro
1 5 10 15
Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys

20 25 30

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

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

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

Leu Ser Leu Ser Pro Gly Lys
225 230

<210> 102
<211> 231
<212> PRT
<213> Artificial Sequence

<400> 102

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

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

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

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

<210> 103

<211> 231

<212> PRT

<213> Artificial Sequence

<400> 103

Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro
1 5 10 15

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

Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val
35 40 45

Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp
50 55 60

Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr
65 70 75 80

Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp
85 90 95

Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu
100 105 110

Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg
115 120 125

Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys
130 135 140

Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp
145 150 155 160

Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys
165 170 175

Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser
180 185 190

Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser
195 200 205

Cys Ser Val Met His Glu Ala Leu Lys Phe His Tyr Thr Gln Lys Ser

210 215 220

Leu Ser Leu Ser Pro Gly Lys

225 230

<210> 104

<211> 438

<212> PRT

<213> Artificial Sequence

<400> 104

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro

1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg

20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val

35 40 45

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

50 55 60

Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn

65 70 75 80

Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln Ile

85 90 95

Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro

100 105 110

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

115 120 125

Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys

130 135 140

Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln

145 150 155 160

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

165 170 175

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln

180 185 190

Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His Ala Pro Pro
195 200 205

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
210 215 220
Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
225 230 235 240
Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp
245 250 255
Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
260 265 270

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

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

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
405 410 415
Ser Val Met His Glu Ala Leu His Ser His Tyr Thr Gln Lys Ser Leu
420 425 430
Ser Leu Ser Pro Gly Lys

435

<210> 105

<211> 1317

<212> DNA

<213> Artificial Sequence

<400> 105

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60

ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180

tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240

gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300

acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360

gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 660

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 720

accctcatga tctctcgcac accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 780

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 840

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtggat 900

caccatgatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 960

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 1020

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 1080

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 1140

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1200

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgatgcac 1260

gaagctctgc actctcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1317

<210> 106

<211> 438

<212> PRT

<213> Artificial Sequence

<400> 106

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro
1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg
20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val
35 40 45

Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys
50 55 60

Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn
65 70 75 80

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

Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro
100 105 110

Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn Leu Asp
115 120 125

Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys
130 135 140

Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln
145 150 155 160

Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe
165 170 175

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln
180 185 190

Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His Ala Pro Pro
195 200 205

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
210 215 220

Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
225 230 235 240

Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp
245 250 255
Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
260 265 270

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

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

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
405 410 415
Ser Val Leu His Glu Ala Leu His Ser His Tyr Thr Gln Lys Ser Leu
420 425 430
Ser Leu Ser Pro Gly Lys
435

<210> 107
<211> 1317
<212> DNA
<213> Artificial Sequence
<400> 107
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60

ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180

tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240

gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300

acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360

gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 660

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 720

accctcatga tctctcgcac accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 780

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 840

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtgctg 900

caccaggatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 960

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 1020

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 1080

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 1140

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1200

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgttacac 1260

gaagctctgc actctcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1317

<210> 108

<211> 438

<212> PRT

<213> Artificial Sequence

<400> 108

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro
1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg
20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val

```
        35                  40                  45
Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys
    50                  55                  60
Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn
65                  70                  75                  80

Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln Ile
                85                  90                  95
Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro
            100                 105                 110
Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn Leu Asp
        115                 120                 125
Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys
    130                 135                 140

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

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
    210                 215                 220
Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
225                 230                 235                 240
Thr Leu Tyr Ile Thr Arg Glu Pro Glu Val Thr Cys Val Val Val Asp
                245                 250                 255
Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
            260                 265                 270

Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn
        275                 280                 285
```

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

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

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
405 410 415
Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu
420 425 430
Ser Leu Ser Pro Gly Lys
435

<210> 109
<211> 1317
<212> DNA
<213> Artificial Sequence
<400> 109
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60
ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180
tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240
gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300
acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360
gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 660

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 720

accctctata tcactcgcga accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 780

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 840

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtgctg 900

caccaggatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 960

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 1020

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 1080

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 1140

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1200

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgatgcac 1260

gaagctctgc acaatcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1317

<210> 110

<211> 438

<212> PRT

<213> Artificial Sequence

<400> 110

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Pro
1 5 10 15

Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg
20 25 30

Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val
35 40 45

Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys
50 55 60

Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn
65 70 75 80

Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln Ile

85 90 95

Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro

100 105 110

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

115 120 125

Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys

130 135 140

Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln

145 150 155 160

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

165 170 175

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln

180 185 190

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

195 200 205

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu

210 215 220

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

225 230 235 240

Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp

245 250 255

Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly

260 265 270

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

275 280 285

Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp

290 295 300

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

305 310 315 320

Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu

325 330 335

Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn
340 345 350

Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile
355 360 365

Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr
370 375 380

Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys
385 390 395 400

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
405 410 415

Ser Val Met His Glu Ala Leu Lys Phe His Tyr Thr Gln Lys Ser Leu
420 425 430

Ser Leu Ser Pro Gly Lys
435

<210> 111

<211> 1317

<212> DNA

<213> Artificial Sequence

<400> 111

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtcctaa tatcactaac 60

ctgtgtcctt tcggtgaagt gttcaacgcc accaggtttg ctagcgtgta tgcctggaac 120

aggaagagga tctctaactg cgtcgccgac tattccgtgc tgtataacag cgcctccttc 180

tccacattca aatgctatgg agtgagcccg acaaaactga acgatctctg ctttacaaat 240

gtctacgccg actcttttgt gatcagaggg gacgaggtcc ggcagatcgc accaggacag 300

acaggcaaga ttgctgacta caactataag ctgcctgacg acttcacagg atgtgtgatc 360

gcatggaact caaacaatct ggactccaaa gtcgggggca actataatta cctgtatcgc 420

ctgttccgga agtccaacct gaagcccttc gagagggaca tcagtacaga gatctatcag 480

gctggctcca ccccttgcaa tggcgtcgaa ggctttaatt gttattttcc cctgcagtct 540

tacgggtttc agcctactaa tggagttggg taccagccat acagagtggt cgtgctcagc 600

ttcgagctcc tgcatgctcc acctaagtcc tgcgacaaaa cccatacatg tccaccatgc 660

ccagctcctg aactgctcgg cgggcctagt gttttcctct tccctcctaa gcccaaggat 720

accctcatga tctctcgcac accagaagtg acctgcgtgg tcgtggatgt ctctcacgag 780

gatcctgaag tgaagtttaa ctggtatgtc gacggagtgg aagtgcacaa cgccaagaca 840

aagccaagag aagaacaata caattctact tatagggtgg tgtctgtgct gacagtgctg 900

caccaggatt ggctgaatgg aaaagaatat aagtgtaagg tctctaacaa ggccctgccc 960

gctccaattg agaagacaat ttccaaggcc aaggggcagc ctcgggaacc tcaggtgtac 1020

acactgcccc catccaggga tgaactgact aaaaatcagg tgtctctgac atgcctggtg 1080

aaagggtttt atccaagtga cattgctgtg gagtgggagt ctaatgggca gcctgaaaat 1140

aactacaaga ccacaccacc agtgctcgat agcgacgggt ctttctttct gtattctaaa 1200

ctgaccgtgg ataaatctcg gtggcagcag ggaaacgtgt tttcttgctc agtgatgcac 1260

gaagctctga aatttcacta tacacagaaa tccctgtccc tgtctccagg caaataa 1317

<210> 112

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 112

atgggtgtgc acgaatgtcc tgcttggctg tggctccttc tctccctgct gtccctgcct 60

cttggactcc cggtgcttgg agcacccccg agactgatct gcgacagcag ggtgctcgag 120

cgctacctcc tggaagccaa ggaagccgaa aacatcacta ctggctgcgc cgaacactgc 180

tccctgaacg agaacatcac cgtgccggac accaaggtca acttctacgc gtggaagaga 240

atggaggtcg gacagcaagc cgtggaagtg tggcagggac ttgcgctcct gtcggaagcc 300

gtgctgaggg gacaagccct gctcgtgaac agctcacagc cttgggagcc cctgcagctg 360

catgtcgaca aggccgtgtc cggactgcgc tcactgacca ctctgctgag ggccttgggt 420

gcccagaaag aggctatttc cccaccggat gcagcctcgg cagctcctct gcggaccatt 480

acggcggaca cctttcggaa gctgttccgc gtctacagca atttcctccg ggggaagttg 540

aaactgtata ccggcgaagc ctgtcggact ggcgatcgct ga 582

<210> 113

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 113

atgggggttc atgagtgccc agcttggctt tggctcctgc tcagcttgct tagtctccct 60

ttgggcctgc ccgtgctggg cgcccctcca cgcttgatct gtgacagcag ggtcttggaa 120

cggtatttgc ttgaagctaa agaagctgag aacataacaa cgggatgtgc tgaacattgc 180

tccttgaacg aaaacatcac agttcccgac acaaaagtca atttttacgc atggaagcgg 240

atggaggttg gccagcaagc tgtggaggtc tggcaagggc tggctcttct cagtgaagcc 300

gtgctgcgcg gacaagcact cttggtgaac tccagccagc cctgggagcc ccttcagctc 360

catgtcgata aagcagttag cggcctccga tcattgacta ccctccttag ggctttgggt 420

gcacaaaaag aggccatttc accaccggac gcggcaagtg ctgctccgtt gcgaactata 480

actgctgaca ccttccggaa actttttcgg gtatattcca actttctcag ggggaaactc 540

aagctctaca ccggcgaggc gtgccgaact ggagaccgct ga 582

<210> 114

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 114

atgggcgtac atgaatgccc ggcatggctt tggctgctgc tgtccctgct gagtttgccg 60

ctgggcctcc ccgtcctcgg cgctcccccg agactcattt gcgactctag ggtcctcgaa 120

cgctatctgc tggaagcaaa agaagctgag aacataacta caggatgcgc tgagcactgt 180

tccttgaatg agaatatcac agtacctgac actaaggtga atttttacgc atggaaacgc 240

atggaagtgg gtcagcaggc cgtggaagtg tggcagggcc tggcgctgct gtccgaggct 300

gttcttagag gccaagcctt gttggtcaat tcctctcaac cctgggagcc cctccagctg 360

catgttgata aagccgtctc tggtctccgg tcccttacca ccctgctcag ggcacttggc 420

gcacagaagg aagctatctc ccccccagac gctgccagtg ccgccccct ccggactatt 480

accgccgata ctttcaggaa actgtttcga gtctatagca attttctccg cgggaaactg 540

aagctgtata caggtgaggc ctgcaggaca ggagatcgct ga 582

<210> 115

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 115

atgggcgtgc acgaatgtcc tgcttggctg tggctgctgc tgagtctgct gtctctgcct 60

ctgggactgc ctgttcttgg agcccctcct agactgatct gcgacagcag agtgctggaa 120

agatacctgc tggaagccaa agaggccgag aacatcacaa caggctgtgc cgagcactgc 180

agcctgaacg agaatatcac cgtgcctgac accaaagtga acttctacgc ctggaagcgg 240

atggaagtgg gacagcaggc tgtggaagtt tggcaaggac tggccctgct gtctgaagct 300

gttctgagag gacaggctct gctggtcaat agctctcagc cttgggaacc tctccagctg 360

catgtggata aggccgtgtc tggcctgaga agcctgacaa cactgctgag agccctggga 420

gcccagaaag aggccatttc tccacctgat gctgccagcg ctgcccctct gagaacaatc 480

accgccgaca ccttcagaaa gctgttccgg gtgtacagca acttcctgcg gggcaagctg 540

aaactgtaca ccggcgaagc ctgcagaacc ggcgatagat aa 582

<210> 116

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 116

atgggggtgc acgagtgccc tgcctggctg tggttgctgc tgtccctgct gtctctgcca 60

ctgggactgc cagtgctggg agctccacct aggctgatct gcgacagccg ggtcctggag 120

aggtacctgc tcgaggccaa ggaggccgag aacattacca caggctgcgc cgagcactgc 180

agcctgaacg agaacattac agtgcccgat acaaaggtga acttctacgc ctggaagagg 240

atggaggtgg gccagcaggc cgtggaggtg tggcaggggc tggccctgct gagcgaggcc 300

gtgctgaggg gccaagccct gctggtcaac agcagccagc cttgggagcc cctgcagctc 360

cacgtggaca aggctgtgtc tggcttgagg tctctcacaa cattgctgag ggccctgggc 420

gcacagaaag aagctatcag cccacctgat gccgctagtg ccgctccact gcggacaatt 480

accgccgata cctttagaaa attgttcagg gtctactcca acttttgcg cgggaagctg 540

aagctctata ccggcgaggc ctgccggaca ggggacagat ga 582

<210> 117

<211> 582

<212> DNA

<213> Artificial Sequence

<400> 117

atgggagtgc acgaatgtcc tgcatggctc tggctcctgc tgtctctcct gagcctgcca 60

ctgggactcc cagtgctggg agcacccccct aggctgatct gcgattctcg ggtgctggag 120

cgctacctgc tcgaggctaa ggaggccgag aatatcacta ctgggtgtgc cgaacactgt 180

agcctcaatg aaaacattac agtcccagat accaaggtga acttttatgc atggaagagg 240

atggaggtcg ggcagcaggc agtggaggtg tggcagggac tggctctgct gtccgaagcc 300

gtgctcagag gtcaggccct gctggttaat tccagccagc cttgggaacc tctgcagctg 360

catgtggaca aggcagtgtc tggcctgaga tcccttacta cactgctgag agcactgggg 420

gctcagaaag aagctatttc cccaccagac gccgcctcag cagcacctct ccggaccatc 480

actgctgaca ccttccgcaa gctctttagg gtgtactcca acttcctgcg cgggaagctc 540

aagctgtaca ccggcgaagc ctgcaggacc ggggatcgct ga 582

<210> 118

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 118

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln

980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100
Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165
Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 119

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 119

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 120
<211> 1273
<212> PRT
<213> Artificial Sequence
<400> 120

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335

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

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 121

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 121

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 122
<211> 3849
<212> DNA
<213> Artificial Sequence
<400> 122

atgttcctgc tgacaacaaa aagaaccatg tttgtgttcc tggtgctgct gcctctggtg 60
tcctcacagt gtgtcaacct gacaacaaga actcagctgc caccagccta caccaactcc 120

ttcaccagag gcgtgtatta cccagacaag gtgtttagaa gcagcgtgct gcactctacc 180
caggacctct ttctgccctt tttcagcaac gtgacatggt ttcacgcaat tcacgtgtcc 240
ggcactaatg gcacaaagcg gttcgacaat ccagtcctgc ctttcaacga tggcgtctac 300
tttgcatcta ctgagaaatc caatatcatt aggggatgga tcttcggcac aaccctggat 360
tctaagaccc agagcctgct gatcgtcaac aacgccacaa acgtggtcat taaggtttgc 420
gagtttcagt tctgtaacga tccttttctg ggcgtgtatt atcataagaa caataagagc 480
tggatggagt ccgagtttag agtgtatagc tctgcaaata attgtacctt tgagtacgtg 540
agccagccct ttctgatgga cctggaggga aaacaaggaa acttcaaaaa cctgcgggaa 600

ttcgttttca aaaacatcga cggctatttc aagatctata gcaagcatac cccaatcaac 660
ctcgtgaggg acctccccca gggctttagc gcactggagc cactggttga cctgcctatc 720
ggcattaata tcacaagatt tcagaccctg ctggcactgc atagaagcta tctgaccccc 780
ggagactcct ctagtgggtg gactgccggc gccgctgcct actatgtggg ctatctgcag 840
ccacggacat tcctgctgaa atacaatgag aacgggacaa tcacagatgc tgttgattgc 900
gcactcgacc ccctgtccga gacaaagtgc actctcaaga gctttaccgt cgagaagggc 960
atctatcaga cctcaaactt cagggtgcag cccacagaat ctatcgtgcg cttccctaat 1020
atcactaacc tgtgtccttt cggtgaagtg ttcaacgcca ccaggtttgc tagcgtgtat 1080

gcctggaaca ggaagaggat ctctaactgc gtcgccgact attccgtgct gtataacagc 1140
gcctccttct ccacattcaa atgctatgga gtgagcccga caaaactgaa cgatctctgc 1200
tttacaaatg tctacgccga ctcttttgtg atcagagggg acgaggtccg gcagatcgca 1260

ccaggacaga caggcaagat tgctgactac aactataagc tgcctgacga cttcacagga 1320

tgtgtgatcg catggaactc aaacaatctg gactccaaag tcgggggcaa ctataattac 1380

ctgtatcgcc tgttccggaa gtccaacctg aagcccttcg agagggacat cagtacagag 1440

atctatcagg ctggctccac cccttgcaat ggcgtcgaag gctttaattg ttattttccc 1500

ctgcagtctt acgggtttca gcctactaat ggagttgggt accagccata cagagtggtc 1560

gtgctcagct tcgagctcct gcatgctcca gctacagttt gcgggccaaa gaagtccact 1620

aacctggtga agaataagtg cgtcaacttc aactttaacg ggctcaccgg caccggcgtg 1680

ctgactgaga gcaacaagaa gtttctgcca tttcaacagt ttggacggga cattgccgac 1740

accaccgatg ccgttcggga tccacagacc ctggaaattc tggacattac accgtgcagc 1800

ttcgggggcg tgagcgtgat cacacccgga accaatacaa gcaaccaggt tgccgtcctg 1860

tatcaggatg tcaattgcac agaagtgcca gttgctatcc acgcagacca gctgactccc 1920

acatggcggg tgtatagcac cggatccaac gtgtttcaga cccgcgccgg atgtctcatt 1980

ggggccgagc acgtgaataa cagctacgag tgcgacatcc ccattggcgc cggcatttgt 2040

gcgtcttacc agactcagac caactctcct ggctccgcct cttccgttgc tagtcagtct 2100

attattgcct ataccatgag cctcggagct gagaatagcg tggcctactc caataattcc 2160

atcgcaatcc ctactaactt cactatttct gtgaccaccg agatcctgcc tgtgtctatg 2220

actaagacta gcgttgattg taccatgtat atttgtggcg actctaccga atgttctaac 2280

ctgctgcttc agtacggctc attttgcaca cagctgaaca gagccctgac tgggatcgct 2340

gtggagcagg acaagaacac acaggaggtg tttgcacagg tgaagcagat ctataagacc 2400

cctcctatta aggatttcgg cggattcaat ttctcacaga ttctgccaga ccccagtaag 2460

ccttccaaga ggagcttcat cgaggatctc ctgtttaaca aggtgaccct ggcagacgcc 2520

ggctttatta agcaatatgg ggattgcctg ggcgacattg ctgccagaga cctgatttgc 2580

gcccagaaat tcaatggcct cacagtgctg ccacctctgc tgaccgacga gatgatcgct 2640

caatacacta gcgcactgct ggccggaacc atcacatcag gctggacctt cggggccgga 2700

gcagcactgc agattccatt cgccatgcag atggcctata gattcaacgg cattggcgtc 2760

acacagaacg tgctgtacga aaaccagaag ctcatcgcta accagtttaa ttccgcaatt 2820

ggaaagatcc aagattcact cagctcaacc gcctctgcac tcggaaagct gcaggacgtg 2880

gtcaaccaga atgctcaggc cctgaacaca ctcgtcaagc agctgtcctc taactttggc 2940

gctatcagct ccgttctgaa cgacattctg agccgcctgg atcccccaga ggctgaagtc 3000

cagattgacc gcctgattac cggccggctg cagtctctgc aaacatacgt gacccagcag 3060

ctgatcagag cagccgagat ccgggcatcc gcaaatctgg cagcaactaa gatgagcgaa 3120

tgcgtgctgg gccagtccaa gcgggtggac ttttgtggca agggctacca cctgatgagc 3180

ttccccccaga gcgccccaca tggcgttgtt tttctgcacg tgacctatgt ccctgctcag 3240

gaaaagaact ttacaactgc tcctgctatc tgccatgacg gcaaggccca cttcccacgg 3300

gagggagtgt ttgtgtccaa tggcacacac tggttcgtga cccagaggaa cttctatgaa 3360

ccccagatca tcaccactga caataccttc gtgtctggaa attgcgacgt cgtgatcggc 3420

atcgttaaca acaccgtgta cgaccctctc cagccagagc tggactcctt taaggaggaa 3480

ctggataagt attttaagaa ccacacaagc ccagatgtgg atctcgggga catctccgga 3540

attaacgcct ccgtggtgaa tatccagaag gagattgacc gcctaaatga agttgccaag 3600

aacctcaatg agtctctgat tgatctgcag gaactgggca agtatgagca gtatatcaaa 3660

tggccctggt acatttggct ggggtttatc gccggactga ttgccatcgt catggtgacc 3720

atcatgctgt gttgcatgac ctcctgttgt tcctgtctga agggctgctg tagttgcggc 3780

tcttgctgta aattcgacga agatgatagc gagcccgtgc tgaagggcgt gaagctgcat 3840

tatacctga 3849

<210> 123

<211> 1282

<212> PRT

<213> Artificial Sequence

<400> 123

Met Phe Leu Leu Thr Thr Lys Arg Thr Met Phe Val Phe Leu Val Leu

1 5 10 15

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

20 25 30

Leu Pro Pro Ala Tyr Thr Asn Ser Phe Thr Arg Gly Val Tyr Tyr Pro

35 40 45

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

50 55 60

Leu Pro Phe Phe Ser Asn Val Thr Trp Phe His Ala Ile His Val Ser

65 70 75 80

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

85 90 95

Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser Asn Ile Ile Arg Gly

100 105 110

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

115 120 125

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

130 135 140

Cys Asn Asp Pro Phe Leu Gly Val Tyr Tyr His Lys Asn Asn Lys Ser

145 150 155 160

Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala Asn Asn Cys Thr

165 170 175

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

180 185 190

Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys Asn Ile Asp Gly

195 200 205

Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn Leu Val Arg Asp

210 215 220

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

225 230 235 240

Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala Leu His Arg Ser

245 250 255

Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr Ala Gly Ala Ala

260 265 270

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

275 280 285

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

290 295 300

Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr Val Glu Lys Gly

305 310 315 320

Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr Glu Ser Ile Val

325 330 335

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

340 345 350

Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser
355 360 365
Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser
370 375 380
Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys

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

450 455 460
Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu
465 470 475 480
Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn
485 490 495
Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val
500 505 510
Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His

515 520 525
Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr Asn Leu Val Lys
530 535 540
Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr Gly Thr Gly Val
545 550 555 560
Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln Gln Phe Gly Arg
565 570 575
Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro Gln Thr Leu Glu

580 585 590
Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly Val Ser Val Ile Thr

595 600 605

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

610 615 620

Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp Gln Leu Thr Pro

625 630 635 640

Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe Gln Thr Arg Ala

645 650 655

Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser Tyr Glu Cys Asp

660 665 670

Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln Thr Gln Thr Asn

675 680 685

Ser Pro Gly Ser Ala Ser Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr

690 695 700

Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser

705 710 715 720

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

725 730 735

Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys

740 745 750

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

755 760 765

Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala Val Glu Gln Asp

770 775 780

Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr

785 790 795 800

Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro

805 810 815

Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe

820 825 830

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

835 840 845

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

900 905 910
Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val Leu Tyr Glu Asn
915 920 925
Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile Gly Lys Ile Gln
930 935 940
Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys Leu Gln Asp Val
945 950 955 960
Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser

965 970 975
Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp Ile Leu Ser Arg
980 985 990
Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp Arg Leu Ile Thr Gly
995 1000 1005
Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala
1010 1015 1020
Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu

1025 1030 1035 1040
Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr
1045 1050 1055
His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly Val Val Phe Leu
1060 1065 1070
His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro
1075 1080 1085
Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe

1090 1095 1100

Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu

1105 1110 1115 1120

Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp

1125 1130 1135

Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro

1140 1145 1150

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu

1205 1210 1215

Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly

1220 1225 1230

Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys Met Thr Ser

1235 1240 1245

Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys

1250 1255 1260

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

1265 1270 1275 1280

Tyr Thr

<210> 124

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 124

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtggccctg gcttatacct ga 3822

<210> 125

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 125

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

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

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

565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605
Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620
His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser

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

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Ala Leu Ala Tyr Thr

1265 1270

<210> 126

<211> 3849

<212> DNA

<213> Artificial Sequence

<400> 126

atgttcctgc tgacaacaaa aagaaccatg tttgtgttcc tggtgctgct gcctctggtg 60

tcctcacagt gtgtcaacct gacaacaaga actcagctgc caccagccta caccaactcc 120

ttcaccagag gcgtgtatta cccagacaag gtgtttagaa gcagcgtgct gcactctacc 180

caggacctct ttctgccctt tttcagcaac gtgacatggt ttcacgcaat tcacgtgtcc 240

ggcactaatg gcacaaagcg gttcgacaat ccagtcctgc ctttcaacga tggcgtctac 300

tttgcatcta ctgagaaatc caatatcatt aggggatgga tcttcggcac aaccctggat 360

tctaagaccc agagcctgct gatcgtcaac aacgccacaa acgtggtcat taaggtttgc 420

gagtttcagt tctgtaacga tccttttctg ggcgtgtatt atcataagaa caataagagc 480

tggatggagt ccgagtttag agtgtatagc tctgcaaata attgtacctt tgagtacgtg 540

agccagccct ttctgatgga cctggaggga aaacaaggaa acttcaaaaa cctgcgggaa 600

ttcgttttca aaaacatcga cggctatttc aagatctata gcaagcatac cccaatcaac 660

ctcgtgaggg acctccccca gggctttagc gcactggagc cactggttga cctgcctatc 720

ggcattaata tcacaagatt tcagaccctg ctggcactgc atagaagcta tctgaccccc 780

ggagactcct ctagtgggtg gactgccggc gccgctgcct actatgtggg ctatctgcag 840

ccacggacat tcctgctgaa atacaatgag aacgggacaa tcacagatgc tgttgattgc 900

gcactcgacc ccctgtccga gacaaagtgc actctcaaga gctttaccgt cgagaagggc 960

atctatcaga cctcaaactt cagggtgcag cccacagaat ctatcgtgcg cttccctaat 1020

atcactaacc tgtgtccttt cggtgaagtg ttcaacgcca ccaggtttgc tagcgtgtat 1080

gcctggaaca ggaagaggat ctctaactgc gtcgccgact attccgtgct gtataacagc 1140

gcctccttct ccacattcaa atgctatgga gtgagcccga caaaactgaa cgatctctgc 1200

tttacaaatg tctacgccga ctcttttgtg atcagagggg acgaggtccg gcagatcgca 1260

ccaggacaga caggcaagat tgctgactac aactataagc tgcctgacga cttcacagga 1320

tgtgtgatcg catggaactc aaacaatctg gactccaaag tcgggggcaa ctataattac 1380

ctgtatcgcc tgttccggaa gtccaacctg aagcccttcg agagggacat cagtacagag 1440

atctatcagg ctggctccac cccttgcaat ggcgtcgaag gctttaattg ttattttccc 1500

ctgcagtctt acgggtttca gcctactaat ggagttgggt accagccata cagagtggtc 1560

gtgctcagct tcgagctcct gcatgctcca gctacagttt gcgggccaaa gaagtccact 1620

aacctggtga agaataagtg cgtcaacttc aactttaacg ggctcaccgg caccggcgtg 1680

ctgactgaga gcaacaagaa gtttctgcca tttcaacagt ttggacggga cattgccgac 1740

accaccgatg ccgttcggga tccacagacc ctggaaattc tggacattac accgtgcagc 1800

ttcgggggcg tgagcgtgat cacacccgga accaatacaa gcaaccaggt tgccgtcctg 1860

tatcaggatg tcaattgcac agaagtgcca gttgctatcc acgcagacca gctgactccc 1920

acatggcggg tgtatagcac cggatccaac gtgtttcaga cccgcgccgg atgtctcatt 1980

ggggccgagc acgtgaataa cagctacgag tgcgacatcc ccattggcgc cggcatttgt 2040

gcgtcttacc agactcagac caactctcct ggctccgcct cttccgttgc tagtcagtct 2100

attattgcct ataccatgag cctcggagct gagaatagcg tggcctactc caataattcc 2160

atcgcaatcc ctactaactt cactatttct gtgaccaccg agatcctgcc tgtgtctatg 2220

actaagacta gcgttgattg taccatgtat atttgtggcg actctaccga atgttctaac 2280

ctgctgcttc agtacggctc attttgcaca cagctgaaca gagccctgac tgggatcgct 2340

gtggagcagg acaagaacac acaggaggtg tttgcacagg tgaagcagat ctataagacc 2400

cctcctatta aggatttcgg cggattcaat ttctcacaga ttctgccaga ccccagtaag 2460

ccttccaaga ggagcttcat cgaggatctc ctgtttaaca aggtgaccct ggcagacgcc 2520

ggctttatta agcaatatgg ggattgcctg ggcgacattg ctgccagaga cctgatttgc 2580

gcccagaaat tcaatggcct cacagtgctg ccacctctgc tgaccgacga gatgatcgct 2640

caatacacta gcgcactgct ggccggaacc atcacatcag gctggacctt cggggccgga 2700

gcagcactgc agattccatt cgccatgcag atggcctata gattcaacgg cattggcgtc 2760

acacagaacg tgctgtacga aaaccagaag ctcatcgcta accagtttaa ttccgcaatt 2820

ggaaagatcc aagattcact cagctcaacc gcctctgcac tcggaaagct gcaggacgtg 2880

gtcaaccaga atgctcaggc cctgaacaca ctcgtcaagc agctgtcctc taactttggc 2940

gctatcagct ccgttctgaa cgacattctg agccgcctgg atcccccaga ggctgaagtc 3000

cagattgacc gcctgattac cggccggctg cagtctctgc aaacatacgt gacccagcag 3060

ctgatcagag cagccgagat ccgggcatcc gcaaatctgg cagcaactaa gatgagcgaa 3120

tgcgtgctgg gccagtccaa gcgggtggac ttttgtggca agggctacca cctgatgagc 3180

ttcccccaga gcgccccaca tggcgttgtt tttctgcacg tgacctatgt ccctgctcag 3240

gaaaagaact ttacaactgc tcctgctatc tgccatgacg gcaaggccca cttcccacgg 3300

gagggagtgt ttgtgtccaa tggcacacac tggttcgtga cccagaggaa cttctatgaa 3360

ccccagatca tcaccactga caataccttc gtgtctggaa attgcgacgt cgtgatcggc 3420

atcgttaaca acaccgtgta cgaccctctc cagccagagc tggactcctt taaggaggaa 3480

ctggataagt attttaagaa ccacacaagc ccagatgtgg atctcgggga catctccgga 3540

attaacgcct ccgtggtgaa tatccagaag gagattgacc gcctaaatga agttgccaag 3600

aacctcaatg agtctctgat tgatctgcag gaactgggca agtatgagca gtatatcaaa 3660

tggccctggt acatttggct ggggtttatc gccggactga ttgccatcgt catggtgacc 3720

atcatgctgt gttgcatgac ctcctgttgt tcctgtctga agggctgctg tagttgcggc 3780

tcttgctgta aattcgacga agatgatagc gagcccgtgc tgaagggcgt ggccctggct 3840

tatacctga 3849

<210> 127

<211> 1282

<212> PRT

<213> Artificial Sequence

<400> 127

Met Phe Leu Leu Thr Thr Lys Arg Thr Met Phe Val Phe Leu Val Leu

1 5 10 15

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

20 25 30

Leu Pro Pro Ala Tyr Thr Asn Ser Phe Thr Arg Gly Val Tyr Tyr Pro

35 40 45

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

50 55 60

Leu Pro Phe Phe Ser Asn Val Thr Trp Phe His Ala Ile His Val Ser

65 70 75 80

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

85 90 95

Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser Asn Ile Ile Arg Gly

100 105 110

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

115 120 125

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

130 135 140

Cys Asn Asp Pro Phe Leu Gly Val Tyr Tyr His Lys Asn Asn Lys Ser

145 150 155 160

Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala Asn Asn Cys Thr

165 170 175

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

180 185 190

Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys Asn Ile Asp Gly

195 200 205

Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn Leu Val Arg Asp

210 215 220

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

225 230 235 240

Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala Leu His Arg Ser

245 250 255

Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr Ala Gly Ala Ala

260 265 270

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

275 280 285

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

290 295 300

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

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

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

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp Gln Leu Thr Pro

625 630 635 640

Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe Gln Thr Arg Ala

645 650 655

Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser Tyr Glu Cys Asp

660 665 670

Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln Thr Gln Thr Asn

675 680 685

Ser Pro Gly Ser Ala Ser Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr

690 695 700

Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser

705 710 715 720

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

725 730 735

Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys

740 745 750

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

755 760 765

Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala Val Glu Gln Asp

770 775 780

Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr

785 790 795 800

Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro

805 810 815

Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe

820 825 830

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

835 840 845

Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe

850 855 860

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

865 870 875 880

Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr Ser Gly Trp Thr

885 890 895

Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala Met Gln Met Ala

900 905 910

Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val Leu Tyr Glu Asn

915 920 925

Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile Gly Lys Ile Gln

930 935 940

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

945 950 955 960

Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser

965 970 975

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

980 985 990

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

995 1000 1005

Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala

1010 1015 1020

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

1025 1030 1035 1040

Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr

1045 1050 1055

His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly Val Val Phe Leu

1060 1065 1070

His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro

1075 1080 1085

Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe

1090 1095 1100

Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu

1105 1110 1115 1120

Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp

1125 1130 1135

Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro

1140 1145 1150

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu

1205 1210 1215

Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly

1220 1225 1230

Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys Met Thr Ser

1235 1240 1245

Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys

1250 1255 1260

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

1265 1270 1275 1280

Tyr Thr

<210> 128

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 128

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagccc tatcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggaccagcac tgcagattcc attccctatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

acccccctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 129

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 129

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

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

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

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Pro Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile
885 890 895

Pro Phe Pro Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

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

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 130

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 130

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagccc tatcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggaccagcac tgcagattcc attccctatg 2700
cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
acccccctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120
gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 131

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 131

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

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

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

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

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile

885 890 895

Pro Phe Pro Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135

Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180

Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 132

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 132

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagccc tatcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggaccagcac tgcagattcc attccctatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

acccccctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540
aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660
atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720
tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780
agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 133
<211> 1273
<212> PRT
<213> Artificial Sequence
<400> 133

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95
Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125
Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

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

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255
Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

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

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

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

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750
Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765
Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

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

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile
885 890 895

Pro Phe Pro Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Pro Ser Ala
930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215
Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val
1250 1255 1260
Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 134
<211> 3822
<212> DNA
<213> Artificial Sequence
<400> 134
atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60
aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180
aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240
aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagccc tatcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggaccagcac tgcagattcc attccctatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

acccccctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 135

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 135

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

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

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

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

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile

885 890 895

Pro Phe Pro Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

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

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 136

<211> 3849

<212> DNA

<213> Artificial Sequence

<400> 136

atgttcctgc tgacaacaaa aagaaccatg tttgtgttcc tggtgctgct gcctctggtg 60

tcctcacagt gtgtcaacct gacaacaaga actcagctgc caccagccta caccaactcc 120

ttcaccagag gcgtgtatta cccagacaag gtgtttagaa gcagcgtgct gcactctacc 180

caggacctct ttctgccctt tttcagcaac gtgacatggt ttcacgcaat tcacgtgtcc 240

ggcactaatg gcacaaagcg gttcgacaat ccagtcctgc ctttcaacga tggcgtctac 300

tttgcatcta ctgagaaatc caatatcatt aggggatgga tcttcggcac aaccctggat 360

tctaagaccc agagcctgct gatcgtcaac aacgccacaa acgtggtcat taaggtttgc 420

gagtttcagt tctgtaacga tccttttctg ggcgtgtatt atcataagaa caataagagc 480

tggatggagt ccgagtttag agtgtatagc tctgcaaata attgtacctt tgagtacgtg 540

agccagccct ttctgatgga cctggaggga aaacaaggaa acttcaaaaa cctgcgggaa 600

ttcgttttca aaaacatcga cggctatttc aagatctata gcaagcatac cccaatcaac 660

ctcgtgaggg acctcccca gggctttagc gcactggagc cactggttga cctgcctatc 720

ggcattaata tcacaagatt tcagaccctg ctggcactgc atagaagcta tctgacccct 780

ggagactcct ctagtgggtg gactgccggc gccgctgcct actatgtggg ctatctgcag 840

ccacggacat tcctgctgaa atacaatgag aacgggacaa tcacagatgc tgttgattgc 900

gcactcgacc ccctgtccga gacaaagtgc actctcaaga gctttaccgt cgagaagggc 960

atctatcaga cctcaaactt cagggtgcag cccacagaat ctatcgtgcg cttccctaat 1020

atcactaacc tgtgtccttt cggtgaagtg ttcaacgcca ccaggtttgc tagcgtgtat 1080

gcctggaaca ggaagaggat ctctaactgc gtcgccgact attccgtgct gtataacagc 1140

gcctccttct ccacattcaa atgctatgga gtgagcccga caaaactgaa cgatctctgc 1200

tttacaaatg tctacgccga ctcttttgtg atcagagggg acgaggtccg gcagatcgca 1260

ccaggacaga caggcaagat tgctgactac aactataagc tgcctgacga cttcacagga 1320

tgtgtgatcg catggaactc aaacaatctg gactccaaag tcgggggcaa ctataattac 1380

ctgtatcgcc tgttccggaa gtccaacctg aagcccttcg agagggacat cagtacagag 1440

atctatcagg ctggctccac cccttgcaat ggcgtcgaag gctttaattg ttattttccc 1500

ctgcagtctt acgggtttca gcctactaat ggagttgggt accagccata cagagtggtc 1560

gtgctcagct tcgagctcct gcatgctcca gctacagttt gcgggccaaa gaagtccact 1620

aacctggtga agaataagtg cgtcaacttc aactttaacg ggctcaccgg caccggcgtg 1680

ctgactgaga gcaacaagaa gtttctgcca tttcaacagt ttggacggga cattgccgac 1740

accaccgatg ccgttcggga tccacagacc ctggaaattc tggacattac accgtgcagc 1800

ttcgggggcg tgagcgtgat cacacccgga accaatacaa gcaaccaggt tgccgtcctg 1860

tatcaggatg tcaattgcac agaagtgcca gttgctatcc acgcagacca gctgactccc 1920

acatggcggg tgtatagcac cggatccaac gtgtttcaga cccgcgccgg atgtctcatt 1980

ggggccgagc acgtgaataa cagctacgag tgcgacatcc ccattggcgc cggcatttgt 2040

gcgtcttacc agactcagac caactctcct ggctccgcct cttccgttgc tagtcagtct 2100

attattgcct ataccatgag cctcggagct gagaatagcg tggcctactc caataattcc 2160

atcgcaatcc ctactaactt cactatttct gtgaccaccg agatcctgcc tgtgtctatg 2220

actaagacta gcgttgattg taccatgtat atttgtggcg actctaccga atgttctaac 2280

ctgctgcttc agtacggctc attttgcaca cagctgaaca gagccctgac tgggatcgct 2340

gtggagcagg acaagaacac acaggaggtg tttgcacagg tgaagcagat ctataagacc 2400

cctcctatta aggatttcgg cggattcaat ttctcacaga ttctgccaga ccccagtaag 2460

ccttccaaga ggagccctat cgaggatctc ctgtttaaca aggtgaccct ggcagacgcc 2520

ggctttatta agcaatatgg ggattgcctg ggcgacattg ctgccagaga cctgatttgc 2580

gcccagaaat tcaatggcct cacagtgctg ccacctctgc tgaccgacga gatgatcgct 2640

caatacacta gcgcactgct ggccggaacc atcacatcag gctggacctt cggggccgga 2700

ccagcactgc agattccatt ccctatgcag atggcctata gattcaacgg cattggcgtc 2760

acacagaacg tgctgtacga aaaccagaag ctcatcgcta accagtttaa ttccgcaatt 2820

ggaaagatcc aagattcact cagctcaacc ccctctgcac tcggaaagct gcaggacgtg 2880

gtcaaccaga atgctcaggc cctgaacaca ctcgtcaagc agctgtcctc taactttggc 2940

gctatcagct ccgttctgaa cgacattctg agccgcctgg atcccccaga ggctgaagtc 3000

cagattgacc gcctgattac cggccggctg cagtctctgc aaacatacgt gacccagcag 3060

ctgatcagag cagccgagat ccgggcatcc gcaaatctgg cagcaactaa gatgagcgaa 3120

tgcgtgctgg gccagtccaa gcgggtggac ttttgtggca agggctacca cctgatgagc 3180

ttcccccaga gcgccccaca tggcgttgtt tttctgcacg tgacctatgt ccctgctcag 3240

gaaaagaact ttacaactgc tcctgctatc tgccatgacg gcaaggccca cttcccacgg 3300

gagggagtgt ttgtgtccaa tggcacacac tggttcgtga cccagaggaa cttctatgaa 3360

ccccagatca tcaccactga caataccttc gtgtctggaa attgcgacgt cgtgatcggc 3420

atcgttaaca acaccgtgta cgaccctctc cagccagagc tggactcctt taaggaggaa 3480

ctggataagt attttaagaa ccacacaagc ccagatgtgg atctcgggga catctccgga 3540

attaacgcct ccgtggtgaa tatccagaag gagattgacc gcctaaatga agttgccaag 3600

aacctcaatg agtctctgat tgatctgcag gaactgggca agtatgagca gtatatcaaa 3660

tggccctggt acatttggct ggggtttatc gccggactga ttgccatcgt catggtgacc 3720

atcatgctgt gttgcatgac ctcctgttgt tcctgtctga agggctgctg tagttgcggc 3780

tcttgctgta aattcgacga agatgatagc gagcccgtgc tgaagggcgt gaagctgcat 3840

tatacctga 3849

<210> 137

<211> 1282

<212> PRT

<213> Artificial Sequence

<400> 137

Met Phe Leu Leu Thr Thr Lys Arg Thr Met Phe Val Phe Leu Val Leu

1 5 10 15

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

20 25 30

Leu Pro Pro Ala Tyr Thr Asn Ser Phe Thr Arg Gly Val Tyr Tyr Pro

35 40 45

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

50 55 60

Leu Pro Phe Phe Ser Asn Val Thr Trp Phe His Ala Ile His Val Ser

65 70 75 80

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

85 90 95

Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser Asn Ile Ile Arg Gly

100 105 110

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

115 120 125

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

130 135 140

Cys Asn Asp Pro Phe Leu Gly Val Tyr Tyr His Lys Asn Asn Lys Ser

145 150 155 160

Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala Asn Asn Cys Thr

165 170 175

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

180 185 190

Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys Asn Ile Asp Gly

195 200 205

Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn Leu Val Arg Asp

210 215 220

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

225 230 235 240

Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala Leu His Arg Ser

245 250 255

Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr Ala Gly Ala Ala

260 265 270

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

275 280 285

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

290 295 300

Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr Val Glu Lys Gly

305 310 315 320

Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr Glu Ser Ile Val

325 330 335

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

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

450 455 460
Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu
465 470 475 480
Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn
485 490 495
Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val
500 505 510
Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His

515 520 525
Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr Asn Leu Val Lys
530 535 540
Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr Gly Thr Gly Val
545 550 555 560
Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln Gln Phe Gly Arg
565 570 575
Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro Gln Thr Leu Glu

580 585 590

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

595 600 605

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

610 615 620

Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp Gln Leu Thr Pro

625 630 635 640

Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe Gln Thr Arg Ala

645 650 655

Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser Tyr Glu Cys Asp

660 665 670

Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln Thr Gln Thr Asn

675 680 685

Ser Pro Gly Ser Ala Ser Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr

690 695 700

Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser

705 710 715 720

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

725 730 735

Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys

740 745 750

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

755 760 765

Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala Val Glu Gln Asp

770 775 780

Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr

785 790 795 800

Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro

805 810 815

Asp Pro Ser Lys Pro Ser Lys Arg Ser Pro Ile Glu Asp Leu Leu Phe

820 825 830

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

835 840 845

Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe

850 855 860

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

865 870 875 880

Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr Ser Gly Trp Thr

885 890 895

Phe Gly Ala Gly Pro Ala Leu Gln Ile Pro Phe Pro Met Gln Met Ala

900 905 910

Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val Leu Tyr Glu Asn

915 920 925

Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile Gly Lys Ile Gln

930 935 940

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

945 950 955 960

Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser

965 970 975

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

980 985 990

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

995 1000 1005

Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala

1010 1015 1020

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

1025 1030 1035 1040

Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr

1045 1050 1055

His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly Val Val Phe Leu

1060 1065 1070

His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro

1075 1080 1085

Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe

1090 1095 1100

Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu

1105 1110 1115 1120

Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp

1125 1130 1135

Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro

1140 1145 1150

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu

1205 1210 1215

Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly

1220 1225 1230

Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys Met Thr Ser

1235 1240 1245

Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys

1250 1255 1260

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

1265 1270 1275 1280

Tyr Thr

<210> 138

<211> 3822

<212> DNA

<213> Artificial Sequence

<400> 138

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcaa gattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtcg aaggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

aatggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg atgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agcacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagccc tatcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggaccagcac tgcagattcc attccctatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

acccccctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaacccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccgtggt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtggccctg gcttatacct ga 3822

<210> 139

<211> 1273

<212> PRT

<213> Artificial Sequence

<400> 139

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

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

115 120 125

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

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

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

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

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

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

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

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys

420 425 430

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

435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys

465 470 475 480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val

500 505 510

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

515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn

530 535 540

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

545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val

565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

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

610 615 620

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

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

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

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

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

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

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

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

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

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile

885 890 895

Pro Phe Pro Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

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

930 935 940

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

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

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

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His

1075 1080 1085

Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr

1105 1110 1115 1120

Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr

1125 1130 1135

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

1140 1145 1150

Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

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

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Ala Leu Ala Tyr Thr

1265 1270

<210> 140

<211> 3849

<212> DNA

<213> Artificial Sequence

<400> 140

atgttcctgc tgacaacaaa aagaaccatg tttgtgttcc tggtgctgct gcctctggtg 60

tcctcacagt gtgtcaacct gacaacaaga actcagctgc caccagccta caccaactcc 120

ttcaccagag gcgtgtatta cccagacaag gtgtttagaa gcagcgtgct gcactctacc 180

caggacctct ttctgccctt tttcagcaac gtgacatggt ttcacgcaat tcacgtgtcc 240

ggcactaatg gcacaaagcg gttcgacaat ccagtcctgc ctttcaacga tggcgtctac 300

tttgcatcta ctgagaaatc caatatcatt aggggatgga tcttcggcac aaccctggat 360

tctaagaccc agagcctgct gatcgtcaac aacgccacaa acgtggtcat taaggtttgc 420

gagtttcagt tctgtaacga tccttttctg ggcgtgtatt atcataagaa caataagagc 480

tggatggagt ccgagtttag agtgtatagc tctgcaaata attgtacctt tgagtacgtg 540

agccagccct ttctgatgga cctggaggga aaacaaggaa acttcaaaaa cctgcgggaa 600

ttcgttttca aaaacatcga cggctatttc aagatctata gcaagcatac cccaatcaac 660

ctcgtgaggg acctccccca gggctttagc gcactggagc cactggttga cctgcctatc 720

ggcattaata tcacaagatt tcagaccctg ctggcactgc atagaagcta tctgacccct 780

ggagactcct ctagtgggtg gactgccggc gccgctgcct actatgtggg ctatctgcag 840

ccacggacat tcctgctgaa atacaatgag aacgggacaa tcacagatgc tgttgattgc 900

gcactcgacc ccctgtccga gacaaagtgc actctcaaga gctttaccgt cgagaagggc 960

atctatcaga cctcaaactt cagggtgcag cccacagaat ctatcgtgcg cttccctaat 1020

atcactaacc tgtgtccttt cggtgaagtg ttcaacgcca ccaggtttgc tagcgtgtat 1080

gcctggaaca ggaagaggat ctctaactgc gtcgccgact attccgtgct gtataacagc 1140

gcctccttct ccacattcaa atgctatgga gtgagcccga caaaactgaa cgatctctgc 1200

tttacaaatg tctacgccga ctcttttgtg atcagagggg acgaggtccg gcagatcgca 1260

ccaggacaga caggcaagat tgctgactac aactataagc tgcctgacga cttcacagga 1320

tgtgtgatcg catggaactc aaacaatctg gactccaaag tcgggggcaa ctataattac 1380

ctgtatcgcc tgttccggaa gtccaacctg aagcccttcg agagggacat cagtacagag 1440

atctatcagg ctggctccac cccttgcaat ggcgtcgaag gctttaattg ttattttccc 1500

ctgcagtctt acgggtttca gcctactaat ggagttgggt accagccata cagagtggtc 1560

gtgctcagct tcgagctcct gcatgctcca gctacagttt gcgggccaaa gaagtccact 1620

aacctggtga agaataagtg cgtcaacttc aactttaacg ggctcaccgg caccggcgtg 1680

ctgactgaga gcaacaagaa gtttctgcca tttcaacagt ttggacggga cattgccgac 1740

accaccgatg ccgttcggga tccacagacc ctggaaattc tggacattac accgtgcagc 1800

ttcgggggcg tgagcgtgat cacacccgga accaatacaa gcaaccaggt tgccgtcctg 1860

tatcaggatg tcaattgcac agaagtgcca gttgctatcc acgcagacca gctgactccc 1920

acatggcggg tgtatagcac cggatccaac gtgtttcaga cccgcgccgg atgtctcatt 1980

ggggccgagc acgtgaataa cagctacgag tgcgacatcc ccattggcgc cggcatttgt 2040

gcgtcttacc agactcagac caactctcct ggctccgcct cttccgttgc tagtcagtct 2100

attattgcct ataccatgag cctcggagct gagaatagcg tggcctactc caataattcc 2160

atcgcaatcc ctactaactt cactatttct gtgaccaccg agatcctgcc tgtgtctatg 2220

actaagacta gcgttgattg taccatgtat atttgtggcg actctaccga atgttctaac 2280

ctgctgcttc agtacggctc attttgcaca cagctgaaca gagccctgac tgggatcgct 2340

gtggagcagg acaagaacac acaggaggtg tttgcacagg tgaagcagat ctataagacc 2400

cctcctatta aggatttcgg cggattcaat ttctcacaga ttctgccaga ccccagtaag 2460

ccttccaaga ggagccctat cgaggatctc ctgtttaaca aggtgaccct ggcagacgcc 2520

ggctttatta agcaatatgg ggattgcctg ggcgacattg ctgccagaga cctgatttgc 2580

gcccagaaat tcaatggcct cacagtgctg ccacctctgc tgaccgacga gatgatcgct 2640

caatacacta gcgcactgct ggccggaacc atcacatcag gctggacctt cggggccgga 2700

ccagcactgc agattccatt ccctatgcag atggcctata gattcaacgg cattggcgtc 2760

acacagaacg tgctgtacga aaaccagaag ctcatcgcta accagtttaa ttccgcaatt 2820

ggaaagatcc aagattcact cagctcaacc ccctctgcac tcggaaagct gcaggacgtg 2880

gtcaaccaga atgctcaggc cctgaacaca ctcgtcaagc agctgtcctc taactttggc 2940

gctatcagct ccgttctgaa cgacattctg agccgcctgg atcccccaga ggctgaagtc 3000

cagattgacc gcctgattac cggccggctg cagtctctgc aaacatacgt gacccagcag 3060

ctgatcagag cagccgagat ccgggcatcc gcaaatctgg cagcaactaa gatgagcgaa 3120

tgcgtgctgg gccagtccaa gcgggtggac ttttgtggca agggctacca cctgatgagc 3180

ttcccccaga gcgccccaca tggcgttgtt tttctgcacg tgacctatgt ccctgctcag 3240

gaaaagaact ttacaactgc tcctgctatc tgccatgacg gcaaggccca cttcccacgg 3300

gagggagtgt ttgtgtccaa tggcacacac tggttcgtga cccagaggaa cttctatgaa 3360

ccccagatca tcaccactga caataccttc gtgtctggaa attgcgacgt cgtgatcggc 3420

atcgttaaca acaccgtgta cgaccctctc cagccagagc tggactcctt taaggaggaa 3480

ctggataagt attttaagaa ccacacaagc ccagatgtgg atctcgggga catctccgga 3540

attaacgcct ccgtggtgaa tatccagaag gagattgacc gcctaaatga agttgccaag 3600

aacctcaatg agtctctgat tgatctgcag gaactgggca agtatgagca gtatatcaaa 3660

tggccctggt acatttggct ggggtttatc gccggactga ttgccatcgt catggtgacc 3720

atcatgctgt gttgcatgac ctcctgttgt tcctgtctga agggctgctg tagttgcggc 3780

tcttgctgta aattcgacga agatgatagc gagcccgtgc tgaagggcgt ggccctggct 3840

tatacctga 3849

<210> 141

<211> 1282

<212> PRT

<213> Artificial Sequence

<400> 141

Met Phe Leu Leu Thr Thr Lys Arg Thr Met Phe Val Phe Leu Val Leu

1 5 10 15

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

20 25 30

Leu Pro Pro Ala Tyr Thr Asn Ser Phe Thr Arg Gly Val Tyr Tyr Pro

35 40 45

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

50 55 60

Leu Pro Phe Phe Ser Asn Val Thr Trp Phe His Ala Ile His Val Ser

65 70 75 80

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

85 90 95

Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser Asn Ile Ile Arg Gly

100 105 110

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

115 120 125

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

130 135 140

Cys Asn Asp Pro Phe Leu Gly Val Tyr Tyr His Lys Asn Asn Lys Ser

145 150 155 160

Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala Asn Asn Cys Thr

165 170 175

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

180 185 190

Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys Asn Ile Asp Gly

195 200 205

Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn Leu Val Arg Asp

210 215 220

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

225 230 235 240

Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala Leu His Arg Ser

245 250 255

Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr Ala Gly Ala Ala

260 265 270

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

275 280 285

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

290 295 300

Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr Val Glu Lys Gly

305 310 315 320

Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr Glu Ser Ile Val

325 330 335

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

340 345 350

Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser

355 360 365

Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser

370 375 380

Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys

385 390 395 400

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

405 410 415

Arg Gln Ile Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr

420 425 430

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

435 440 445

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

450 455 460

Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu

465 470 475 480

Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn

485 490 495

Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val

500 505 510

Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His

515 520 525

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

530 535 540

Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr Gly Thr Gly Val

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp Gln Leu Thr Pro

625 630 635 640

Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe Gln Thr Arg Ala

645 650 655

Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser Tyr Glu Cys Asp

660 665 670

Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln Thr Gln Thr Asn

675 680 685

Ser Pro Gly Ser Ala Ser Ser Val Ala Ser Gln Ser Ile Ile Ala Tyr

690 695 700

Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr Ser Asn Asn Ser

705 710 715 720

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

725 730 735

Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr Met Tyr Ile Cys

740 745 750

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

755 760 765

Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala Val Glu Gln Asp

770 775 780

Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr

785 790 795 800

Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro

805 810 815

Asp Pro Ser Lys Pro Ser Lys Arg Ser Pro Ile Glu Asp Leu Leu Phe

820 825 830

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

835 840 845

Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe

850 855 860

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

865 870 875 880

Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr Ser Gly Trp Thr

885 890 895

Phe Gly Ala Gly Pro Ala Leu Gln Ile Pro Phe Pro Met Gln Met Ala

900 905 910

Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val Leu Tyr Glu Asn

915 920 925

Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile Gly Lys Ile Gln

930 935 940

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

945 950 955 960

Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser

965 970 975

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

980 985 990

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

995 1000 1005

Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala

1010 1015 1020

Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu
1025 1030 1035 1040
Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr
1045 1050 1055
His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly Val Val Phe Leu

1060 1065 1070
His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro
1075 1080 1085
Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe
1090 1095 1100
Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu
1105 1110 1115 1120
Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp

1125 1130 1135
Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1140 1145 1150
Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His
1155 1160 1165
Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser
1170 1175 1180
Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys

1185 1190 1195 1200
Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu
1205 1210 1215
Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile Ala Gly
1220 1225 1230
Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys Met Thr Ser
1235 1240 1245
Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly Ser Cys Cys Lys

1250 1255 1260
Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys Gly Val Ala Leu Ala

1265 1270 1275 1280

Tyr Thr

<210> 142

<211> 24

<212> PRT

<213> SARS-COV-2

<400> 142

Met Phe Leu Leu Thr Thr Lys Arg Thr Met Phe Val Phe Leu Val Leu

1 5 10 15

Leu Pro Leu Val Ser Ser Gln Cys

20

<210> 143

<211> 72

<212> DNA

<213> Artificial Sequence

<400> 143

atgttcctgc tgacaacaaa aagaaccatg tttgtgttcc tggtgctgct gcctctggtg 60

tcctcacagt gt 72

<210> 144

<211> 140

<212> RNA

<213> Artificial Sequence

<400> 144

ggacagaucg ccuggagacg ccauccacgc uguuuugacc uccauagaag acaccgggac 60

cgauccagcc uccgcggccg ggaacggugc auuggaacgc ggauuccccg ugccaagagu 120

gacucaccgu ccuugacacg 140

<210> 145

<211> 105

<212> RNA

<213> Artificial Sequence

<400> 145

cggguggcau cccugugacc ccucccccagu gccucuccug gcccuggaag uugccacucc 60

agugcccacc agccuugucc uaauaaaauu aaguugcauc aagcu 105

<210> 146

<211> 105

<212> RNA

<213> Artificial Sequence

<400> 146

ggguggcauc ccugugaccc cucccagug ccucuccugg cccuggaagu ugccacucca 60

gugcccacca gccuuguccu aauaaaauua aguugcauca aagcu 105

<210> 147

<211> 4067

<212> RNA

<213> Artificial Sequence

<400> 147

ggacagaucg ccuggagacg ccauccacgc uguuuugacc uccauagaag acaccgggac 60

cgauccagcc uccgcggccg ggaacggugc auuggaacgc ggauuccccg ugccaagagu 120

gacucaccgu ccuugacacg auguucgucu uccucgugcu gcucccacuc guuucuuccc 180

agugugucaa ccugacaacu aggacucagc ugccaccagc cuacaccaac uccuucacca 240

gaggcgugua uuacccagac aagguguuua gaagcagcgu gcugcacucu acccaggacc 300

ucuuucugcc cuuuuucagc aacgugacau gguuucacgc aauucacgug uccggcacua 360

auggcacaaa gcgguucgac aauccagucc ugccuuucaa cgauggcguc uacuuugcau 420

cuacugagaa auccaauauc auuaggggau ggaucuucgg cacaacccug gauucuaaga 480

cccagagccu gcugaucguc aacaacgcca caaacguggu cauuaagguu ugcgaguuuc 540

aguucuguaa cgauccuuuu cugggcgugu auuaucauaa gaacaauaag agcuggaugg 600

aguccgaguu uagaguguau agcucugcaa auaauuguac cuuugaguac gugagccagc 660

ccuuucugau ggaccuggag ggaaaacaag gaaacuucaa aaaccugcgg gaauucguuu 720

ucaaaaacau cgacggcuau uucaagaucu auagcaagca uaccccaauc aaccucguga 780

gggaccuccc ccagggcuuu agcgcacugg agccacuggu ugaccugccu aucggcauua 840

auaucacaag auuucagacc cugcuggcac ugcauagaag cuaucugacc ccuggagacu 900

ccucuagugg guggacugcc ggcgccgcug ccuacuaugu gggcuaucug cagccacgga 960

cauuccugcu gaaauacaau gagaacggga caaucacaga ugcuguugau ugcgcacucg 1020

acccccuguc cgagacaaag ugcacucuca agagcuuuac cgucgagaag ggcaucuauc 1080

agaccucaaa cuucagggug cagcccacag aaucuaucgu gcgcuucccu aauaucacua 1140

accugugucc uuucggugaa guguucaacg ccaccagguu ugcuagcgug uaugccugga 1200

acaggaagag gaucucuaac ugcgucgccg acuauuccgu gcuguauaac agcgccuccu 1260

ucuccacauu caaaugcuau ggagugagcc cgacaaaacu gaacgaucuc ugcuuuacaa 1320

augucuacgc cgacucuuuu gugaucagag gggacgaggu ccggcagauc gcaccaggac 1380

agacaggcaa gauugcugac uacaacuaua agcugccuga cgacuucaca ggauguguga 1440

ucgcauggaa cucaaacaau cuggacucca aagucggggg caacuauaau uaccuguauc 1500

gccuguuccg gaaguccaac cugaagcccu ucgagaggga caucaguaca gagaucuauc 1560

aggcuggcuc caccccuugc aauggcgucg aaggcuuuaa uuguuauuuu ccccugcagu 1620

cuuacggguu ucagccuacu aauggaguug gguaccagcc auacagagug gucgugcuca 1680

gcuucgagcu ccugcaugcu ccagcuacag uuugcgggcc aaagaagucc acuaaccugg 1740

ugaagaauaa gugcgucaac uucaacuuua acgggcucac cggcaccggc gugcugacug 1800

agagcaacaa gaaguuucug ccauuucaac aguuuggacg ggacauugcc gacaccaccg 1860

augccguucg ggauccacag acccuggaaa uucuggacau uacaccgugc agcuucgggg 1920

gcgugagcgu gaucacaccc ggaaccaaua caagcaacca gguugccguc cuguaucagg 1980

augucaauug cacagaagug ccaguugcua uccacgcaga ccagcugacu cccacauggc 2040

ggguguauag caccggaucc aacguguuuc agacccgcgc cggaugucuc auuggggccg 2100

agcacgugaa uaacagcuac gagugcgaca uccccauugg cgccggcauu ugugcgucuu 2160

accagacuca gaccaacucu ccuggcuccg ccucuuccgu ugcuagucag ucuauuauug 2220

ccuauaccau gagccucgga gcugagaaua gcguggccua cuccaauaau uccaucgcaa 2280

ucccuacuaa cuucacuauu ucugugacca ccgagauccu gccugugucu augacuaaga 2340

cuagcguuga uuguaccaug uauauuugug gcgacucuac cgaauguucu aaccugcugc 2400

uucaguacgg cucauuuugc acacagcuga acagagcccu gacugggauc gcuguggagc 2460

aggacaagaa cacacaggag guguuugcac aggugaagca gaucuauaag accccuccua 2520

uuaaggauuu cggcggauuc aauuucucac agauucugcc agaccccagu aagccuucca 2580

agaggagcuu caucgaggau cuccuguuua acaaggugac ccuggcagac gccggcuuua 2640

uuaagcaaua uggggauugc cugggcgaca uugcugccag agaccugauu ugcgcccaga 2700

aauucaaugg ccucacagug cugccaccuc ugcugaccga cgagaugauc gcucaauaca 2760

cuagcgcacu gcuggccgga accaucacau caggcuggac cuucggggcc ggagcagcac 2820

ugcagauucc auucgccaug cagauggccu auagauucaa cggcauuggc gucacacaga 2880

acgugcugua cgaaaaccag aagcucaucg cuaaccaguu uaauuccgca auuggaaaga 2940

uccaagauuc acucagcuca accgccucug cacucggaaa gcugcaggac guggucaacc 3000

agaaugcuca ggcccugaac acacucguca agcagcuguc cucuaacuuu ggcgcuauca 3060

gcuccguucu gaacgacauu cugagccgcc uggaucccccc agaggcugaa guccagauug 3120

accgccugau uaccggccgg cugcagucuc ugcaaacaua cgugacccag cagcugauca 3180

gagcagccga gauccgggca uccgcaaauc uggcagcaac uaagaugagc gaaugcgugc 3240

ugggccaguc caagcgggug gacuuuugug gcaagggcua ccaccugaug agcuuccccc 3300

agagcgcccc acauggcguu guuuuucugc acgugaccua ugucccugcu caggaaaaga 3360

acuuuacaac ugcuccugcu aucugccaug acggcaaggc ccacuuccca cgggagggag 3420

uguuuguguc caauggcaca cacugguucg ugacccagag gaacuucuau gaaccccaga 3480

ucaucaccac ugacaauacc uucgugucug gaaauugcga cgucgugauc ggcaucguua 3540

acaacaccgu guacgacccu cuccagccag agcuggacuc cuuuaaggag gaacuggaua 3600

aguauuuuaa gaaccacaca agcccagaug uggaucucgg ggacaucucc ggaauuaacg 3660

ccuccguggu gaauauccag aaggagauug accgccuaaa ugaaguugcc aagaaccuca 3720

augagucucu gauugaucug caggaacugg gcaaguauga gcaguauauc aaauggcccu 3780

gguacauuug gcugggguuu aucgccggac ugauugccau cgucauggug accaucaugc 3840

uguguugcau gaccuccugu uguuccuguc ugaagggcug cuguaguugc ggcucuugcu 3900

guaaauucga cgaagaugau agcgagcccg ugcugaaggg cgugaagcug cauuauaccu 3960

gacggguggc aucccuguga ccccucccca gugccucucc uggcccugga aguugccacu 4020

ccagugccca ccagccuugu ccuaauaaaa uuaaguugca ucaagcu 4067

<210> 148

<211> 3822

<212> RNA

<213> Artificial Sequence

<400> 148

auguucgucu uccucgugcu gcucccacuc guuucuuccc agugugucaa ccugacaacu 60

aggacucagc ugccaccagc cuacaccaac uccuucacca gaggcgugua uuacccagac 120

aagguguuua gaagcagcgu gcugcacucu acccaggacc ucuuucugcc cuuuuucagc 180

aacgugacau gguuucacgc aauucacgug uccggcacua auggcacaaa gcgguucgac 240

aauccagucc ugccuuucaa cgauggcguc uacuuugcau cuacugagaa auccaauauc 300

auuaggggau ggaucuucgg cacaacccug gauucuaaga cccagagccu gcugaucguc 360

aacaacgcca caaacguggu cauuaagguu ugcgaguuuc aguucuguaa cgauccuuuu 420

cugggcgugu auuaucauaa gaacaauaag agcuggaugg aguccgaguu uagaguguau 480

agcucugcaa auaauuguac cuuugaguac gugagccagc ccuuucugau ggaccuggag 540

ggaaaacaag gaaacuucaa aaaccugcgg gaauucguuu ucaaaaacau cgacggcuau 600

uucaagaucu auagcaagca uaccccaauc aaccucguga gggaccuccc ccagggcuuu 660

agcgcacugg agccacuggu ugaccugccu aucggcauua auaucacaag auuucagacc 720

cugcuggcac ugcauagaag cuaucugacc ccuggagacu ccucuagugg guggacugcc 780

ggcgccgcug ccuacuaugu gggcuaucug cagccacgga cauuccugcu gaaauacaau 840

gagaacggga caaucacaga ugcuguugau ugcgcacucg acccccuguc cgagacaaag 900

ugcacucuca agagcuuuac cgucgagaag ggcaucuauc agaccucaaa cuucagggug 960

cagcccacag aaucuaucgu gcgcuucccu aauaucacua accugugucc uuucggugaa 1020

guguucaacg ccaccagguu ugcuagcgug uaugccugga acaggaagag gaucucuaac 1080

ugcgucgccg acuauuccgu gcuguauaac agcgccuccu ucuccacauu caaaugcuau 1140

ggagugagcc cgacaaaacu gaacgaucuc ugcuuuacaa augucuacgc cgacucuuuu 1200

gugaucagag gggacgaggu ccggcagauc gcaccaggac agacaggcaa gauugcugac 1260

uacaacuaua agcugccuga cgacuucaca ggauguguga ucgcauggaa cucaaacaau 1320

cuggacucca aagucggggg caacuauaau uaccuguauc gccuguuccg gaaguccaac 1380

cugaagcccu ucgagaggga caucaguaca gagaucuauc aggcuggcuc cacccuugc 1440

aauggcgucg aaggcuuuaa uuguuauuuu ccccugcagu cuuacggguu ucagccuacu 1500

aauggaguug gguaccagcc auacagagug gucgugcuca gcuucgagcu ccugcaugcu 1560

ccagcuacag uuugcgggcc aaagaagucc acuaaccugg ugaagaauaa gugcgucaac 1620

uucaacuuua acgggcucac cggcaccggc gugcugacug agagcaacaa gaaguuucug 1680

ccauuucaac aguuuggacg ggacauugcc gacaccaccg augccguucg ggauccacag 1740

acccuggaaa uucuggacau uacaccgugc agcuucgggg gcgugagcgu gaucacaccc 1800

ggaaccaaua caagcaacca gguugccguc cuguaucagg augucaauug cacagaagug 1860

ccaguugcua uccacgcaga ccagcugacu cccacauggc ggguguauag caccggaucc 1920

aacguguuuc agacccgcgc cggaugucuc auuggggccg agcacgugaa uaacagcuac 1980

gagugcgaca uccccauugg cgccggcauu ugugcgucuu accagacuca gaccaacucu 2040

ccuggcuccg ccucuuccgu ugcuagucag ucuauuauug ccuauaccau gagccucgga 2100

gcugagaaua gcguggccua cuccaauaau uccaucgcaa ucccuacuaa cuucacuauu 2160

ucugugacca ccgagauccu gccugugucu augacuaaga cuagcguuga uuguaccaug 2220

uauauuugug gcgacucuac cgaauguucu aaccugcugc uucaguacgg cucauuuugc 2280

acacagcuga acagagcccu gacugggauc gcuguggagc aggacaagaa cacacaggag 2340

guguuugcac aggugaagca gaucuauaag accccuccua uuaaggauuu cggcggauuc 2400
aauuucucac agauucugcc agaccccagu aagccuucca agaggagcuu caucgaggau 2460
cuccuguuua acaaggugac ccuggcagac gccggcuuua uuaagcaaua uggggauugc 2520
cugggcgaca uugcugccag agaccugauu ugcgcccaga aauucaaugg ccucacagug 2580
cugccaccuc ugcugaccga cgagaugauc gcucaauaca cuagcgcacu gcuggccgga 2640
accaucacau caggcuggac cuucggggcc ggagcagcac ugcagauucc auucgccaug 2700
cagauggccu auagauucaa cggcauuggc gucacacaga acgugcugua cgaaaaccag 2760

aagcucaucg cuaaccaguu uaauuccgca auuggaaaga uccaagauuc acucagcuca 2820
accgccucug cacucggaaa gcugcaggac guggucaacc agaaugcuca ggcccugaac 2880
acacucguca agcagcuguc cucuaacuuu ggcgcuauca gcuccguucu gaacgacauu 2940
cugagccgcc uggaucccc agaggcugaa guccagauug accgccugau uaccggccgg 3000
cugcagucuc ugcaaacaua cgugacccag cagcugauca gagcagccga gauccgggca 3060
uccgcaaauc uggcagcaac uaagaugagc gaaugcgugc ugggccaguc caagcgggug 3120
gacuuuugug gcaagggcua ccaccugaug agcuuccccc agagcgcccc acauggcguu 3180

guuuuucugc acgugaccua ugucccugcu caggaaaaga acuuuacaac ugcuccugcu 3240
aucugccaug acggcaaggc ccacuuccca cgggagggag uguuuguguc caauggcaca 3300
cacugguucg ugacccagag gaacuucuau gaacccccaga ucaucaccac ugacaauacc 3360
uucgugucug gaaauugcga cgucgugauc ggcaucguua acaacaccgu guacgacccu 3420
cuccagccag agcuggacuc cuuuaaggag gaacuggaua aguauuuuaa gaaccacaca 3480
agcccagaug uggaucucgg ggacaucucc ggaauuaacg ccuccguggu gaauauccag 3540
aaggagauug accgccuaaa ugaaguugcc aagaaccuca augagucucu gauugaucug 3600

caggaacugg gcaaguauga gcaguauauc aaauggcccu gguacauuug gcugggguuu 3660
aucgccggac ugauugccau cgucauggug accaucaugc uguguugcau gaccuccugu 3720
uguuccuguc ugaagggcug cuguaguugc ggcucuugcu guaaauucga cgaagaugau 3780
agcgagcccg ugcugaaggg cgugaagcug cauuauaccu ga 3822

<210> 149

<211> 4067

<212> DNA

<213> Artificial Sequence

<400> 149

ggacagatcg cctggagacg ccatccacgc tgttttgacc tccatagaag acaccgggac 60

cgatccagcc tccgcggccg ggaacggtgc attggaacgc ggattccccg tgccaagagt 120

gactcaccgt ccttgacacg atgttcgtct tcctcgtgct gctcccactc gtttcttccc 180

agtgtgtcaa cctgacaact aggactcagc tgccaccagc ctacaccaac tccttcacca 240

gaggcgtgta ttacccagac aaggtgttta gaagcagcgt gctgcactct acccaggacc 300

tctttctgcc ctttttcagc aacgtgacat ggtttcacgc aattcacgtg tccggcacta 360

atggcacaaa gcggttcgac aatccagtcc tgcctttcaa cgatggcgtc tactttgcat 420

ctactgagaa atccaatatc attaggggat ggatcttcgg cacaaccctg gattctaaga 480

cccagagcct gctgatcgtc aacaacgcca caaacgtggt cattaaggtt tgcgagtttc 540

agttctgtaa cgatcctttt ctgggcgtgt attatcataa gaacaataag agctggatgg 600

agtccgagtt tagagtgtat agctctgcaa ataattgtac ctttgagtac gtgagccagc 660

cctttctgat ggacctggag ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt 720

tcaaaaacat cgacggctat ttcaagatct atagcaagca taccccaatc aacctcgtga 780

gggacctccc ccagggcttt agcgcactgg agccactggt tgacctgcct atcggcatta 840

atatcacaag atttcagacc ctgctggcac tgcatagaag ctatctgacc cctggagact 900

cctctagtgg gtggactgcc ggcgccgctg cctactatgt gggctatctg cagccacgga 960

cattcctgct gaaatacaat gagaacggga caatcacaga tgctgttgat tgcgcactcg 1020

acccccctgtc cgagacaaag tgcactctca agagctttac cgtcgagaag ggcatctatc 1080

agacctcaaa cttcagggtg cagcccacag aatctatcgt gcgcttccct aatatcacta 1140

acctgtgtcc tttcggtgaa gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga 1200

acaggaagag gatctctaac tgcgtcgccg actattccgt gctgtataac agcgcctcct 1260

tctccacatt caaatgctat ggagtgagcc cgacaaaact gaacgatctc tgctttacaa 1320

atgtctacgc cgactctttt gtgatcagag gggacgaggt ccggcagatc gcaccaggac 1380

agacaggcaa gattgctgac tacaactata agctgcctga cgacttcaca ggatgtgtga 1440

tcgcatggaa ctcaaacaat ctggactcca aagtcggggg caactataat tacctgtatc 1500

gcctgttccg gaagtccaac ctgaagccct tcgagaggga catcagtaca gagatctatc 1560

aggctggctc cacccccttgc aatggcgtcg aaggctttaa ttgttatttt cccctgcagt 1620

cttacgggtt tcagcctact aatggagttg ggtaccagcc atacagagtg gtcgtgctca 1680

gcttcgagct cctgcatgct ccagctacag tttgcgggcc aaagaagtcc actaacctgg 1740

tgaagaataa gtgcgtcaac ttcaacttta acgggctcac cggcaccggc gtgctgactg 1800

agagcaacaa gaagtttctg ccatttcaac agtttggacg ggacattgcc gacaccaccg 1860

atgccgttcg ggatccacag accctggaaa ttctggacat tacaccgtgc agcttcgggg 1920

gcgtgagcgt gatcacaccc ggaaccaata caagcaacca ggttgccgtc ctgtatcagg 1980
atgtcaattg cacagaagtg ccagttgcta tccacgcaga ccagctgact cccacatggc 2040
gggtgtatag caccggatcc aacgtgtttc agacccgcgc cggatgtctc attggggccg 2100
agcacgtgaa taacagctac gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt 2160
accagactca gaccaactct cctggctccg cctcttccgt tgctagtcag tctattattg 2220
cctataccat gagcctcgga gctgagaata gcgtggccta ctccaataat tccatcgcaa 2280
tccctactaa cttcactatt tctgtgacca ccgagatcct gcctgtgtct atgactaaga 2340

ctagcgttga ttgtaccatg tatatttgtg gcgactctac cgaatgttct aacctgctgc 2400
ttcagtacgg ctcattttgc acacagctga acagagccct gactgggatc gctgtggagc 2460
aggacaagaa cacacaggag gtgtttgcac aggtgaagca gatctataag acccctccta 2520
ttaaggattt cggcggattc aatttctcac agattctgcc agaccccagt aagccttcca 2580
agaggagctt catcgaggat ctcctgttta acaaggtgac cctggcagac gccggcttta 2640
ttaagcaata tggggattgc ctgggcgaca ttgctgccag agacctgatt tgcgcccaga 2700
aattcaatgg cctcacagtg ctgccacctc tgctgaccga cgagatgatc gctcaataca 2760

ctagcgcact gctggccgga accatcacat caggctggac cttcggggcc ggagcagcac 2820
tgcagattcc attcgccatg cagatggcct atagattcaa cggcattggc gtcacacaga 2880
acgtgctgta cgaaaaccag aagctcatcg ctaaccagtt taattccgca attggaaaga 2940
tccaagattc actcagctca accgcctctg cactcggaaa gctgcaggac gtggtcaacc 3000
agaatgctca ggccctgaac acactcgtca agcagctgtc ctctaacttt ggcgctatca 3060
gctccgttct gaacgacatt ctgagccgcc tggatccccc agaggctgaa gtccagattg 3120
accgcctgat taccggccgg ctgcagtctc tgcaaacata cgtgacccag cagctgatca 3180

gagcagccga gatccgggca tccgcaaatc tggcagcaac taagatgagc gaatgcgtgc 3240
tgggccagtc caagcgggtg gacttttgtg gcaagggcta ccacctgatg agcttccccc 3300
agagcgcccc acatggcgtt gtttttctgc acgtgaccta tgtccctgct caggaaaaga 3360
actttacaac tgctcctgct atctgccatg acggcaaggc ccacttccca cgggagggag 3420
tgtttgtgtc caatggcaca cactggttcg tgacccagag gaacttctat gaaccccaga 3480
tcatcaccac tgacaatacc ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta 3540
acaacaccgt gtacgaccct ctccagccag agctggactc ctttaaggag gaactggata 3600

agtattttaa gaaccacaca agcccagatg tggatctcgg ggacatctcc ggaattaacg 3660

cctccgtggt gaatatccag aaggagattg accgcctaaa tgaagttgcc aagaacctca 3720

atgagtctct gattgatctg caggaactgg gcaagtatga gcagtatatc aaatggccct 3780

ggtacatttg gctggggttt atcgccggac tgattgccat cgtcatggtg accatcatgc 3840

tgtgttgcat gacctcctgt tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct 3900

gtaaattcga cgaagatgat agcgagcccg tgctgaaggg cgtgaagctg cattatacct 3960

gacgggtggc atccctgtga cccctcccca gtgcctctcc tggccctgga agttgccact 4020

ccagtgccca ccagccttgt cctaataaaa ttaagttgca tcaagct 4067

<210> 150

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 150

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aatttccggc actaatggca caaagcggtt cgacaatcca 240

gtcctgcctt tcaacgatgg cgtctacttt gcatctactg agaaatccaa tatcattagg 300

ggatggatct tcggcacaac cctggattct aagacccaga gcctgctgat cgtcaacaac 360

gccacaaacg tggtcattaa ggtttgcgag tttcagttct gtaacgatcc ttttctgggc 420

gtttaccata agaacaataa gagctggatg gagtccgagt ttagagtgta tagctctgca 480

aataattgta cctttgagta cgtgagccag ccctttctga tggacctgga gggaaaacaa 540

ggaaacttca aaaacctgcg ggaattcgtt ttcaaaaaca tcgacggcta tttcaagatc 600

tatagcaagc ataccccaat caacctcgtg agggacctcc cccagggctt tagcgcactg 660

gagccactgg ttgacctgcc tatcggcatt aatatcacaa gatttcagac cctgctggca 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca agattgctga ctacaactat 1260
aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320
aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380
ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440
gaaggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttatggagtt 1500
gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560
gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680
cagtttggac gggacattga tgacaccacc gatgccgttc gggatccaca gaccctggaa 1740
attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800
acaagcaacc aggttgccgt cctgtatcag ggagtcaatt gcacagaagt gccagttgct 1860
atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920
cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980
atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcatagaagg 2040

gccaggtccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agctgagaat 2100
agcgtggcct actccaataa ttccatcgca atccctataa acttcactat ttctgtgacc 2160
accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220
ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280
aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340
caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400
cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520
attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580
ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640
tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700
tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760
gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820
gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctggcacgc 2940

ctggataagg tggaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctcacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 151

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 151

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp Asn Pro
65 70 75 80

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

85 90 95

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr

100 105 110

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

115 120 125

Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr His Lys

130 135 140

Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala

145 150 155 160

Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu

165 170 175

Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys

180 185 190

Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn

195 200 205

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

210 215 220

Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg
340 345 350
Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

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

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

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

545 550 555 560
Gln Phe Gly Arg Asp Ile Asp Asp Thr Thr Asp Ala Val Arg Asp Pro
565 570 575
Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly Val

580 585 590

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

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp

610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe

625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser

645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln

660 665 670

Thr Gln Thr Asn Ser His Arg Arg Ala Arg Ser Val Ala Ser Gln Ser

675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr

690 695 700

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

705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln

740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln

770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser

785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu

805 810 815

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

820 825 830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys
835 840 845
Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
850 855 860
Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr

865 870 875 880
Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
885 890 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
900 905 910
Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
915 920 925
Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys

930 935 940
Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945 950 955 960
Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
965 970 975
Ile Leu Ala Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp Arg
980 985 990
Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005
Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr
1010 1015 1020
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys
1025 1030 1035 1040
Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly
1045 1050 1055
Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070
Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr His Asn Thr Phe Val Ser

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 152

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 152

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aatttccggc actaatggca caaagcggtt cgacaatcca 240

gtcctgcctt tcaacgatgg cgtctacttt gcatctactg agaaatccaa tatcattagg 300

ggatggatct tcggcacaac cctggattct aagacccaga gcctgctgat cgtcaacaac 360

gccacaaacg tggtcattaa ggtttgcgag tttcagttct gtaacgatcc ttttctgggc 420

gtttaccata agaacaataa gagctggatg gagtccgagt ttagagtgta tagctctgca 480

aataattgta cctttgagta cgtgagccag ccctttctga tggacctgga gggaaaacaa 540

ggaaacttca aaaacctgcg ggaattcgtt ttcaaaaaca tcgacggcta tttcaagatc 600

tatagcaagc ataccccaat caacctcgtg agggacctcc cccagggctt tagcgcactg 660

gagccactgg ttgacctgcc tatcggcatt aatatcacaa gatttcagac cctgctggca 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca agattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

gaaggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttatggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattga tgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggagtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcatagaagg 2040

gccaggtccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agctgagaat 2100
agcgtggcct actccaataa ttccatcgca atccctataa acttcactat ttctgtgacc 2160
accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220
ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280
aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340
caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400
cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520
attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580
ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640
tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700
tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760
gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820
gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctggcacgc 2940
ctggatcccc cagaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000
ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060
ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120
ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180
cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240
gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctcacaatac cttcgtgtct 3360
ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420
gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480
gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540
gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600
ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660
ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 153

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 153

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp Asn Pro

65 70 75 80

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

85 90 95

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr

100 105 110

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

115 120 125

Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr His Lys

130 135 140

Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala

145 150 155 160

Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu

165 170 175

Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys

180 185 190

Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn

195 200 205

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

210 215 220

Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

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

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

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

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

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

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

675 680 685
Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr

690 695 700

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

705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln

740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln

770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser

785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu

805 810 815

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

820 825 830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys

835 840 845

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

850 855 860

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

865 870 875 880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala

885 890 895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val

900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

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

930 935 940

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

995 1000 1005
Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr
1010 1015 1020
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys
1025 1030 1035 1040
Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly
1045 1050 1055
Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070
Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg
1075 1080 1085
Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg
1090 1095 1100
Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr His Asn Thr Phe Val Ser
1105 1110 1115 1120
Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135
Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr
1140 1145 1150
Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly
1155 1160 1165
Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn
1170 1175 1180
Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 154

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 154

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aatttccggc actaatggca caaagcggtt cgacaatcca 240

gtcctgcctt tcaacgatgg cgtctacttt gcatctactg agaaatccaa tatcattagg 300

ggatggatct tcggcacaac cctggattct aagacccaga gcctgctgat cgtcaacaac 360

gccacaaacg tggtcattaa ggtttgcgag tttcagttct gtaacgatcc ttttctgggc 420

gtttaccata agaacaataa gagctggatg gagtccgagt ttagagtgta tagctctgca 480

aataattgta cctttgagta cgtgagccag ccctttctga tggacctgga gggaaaacaa 540

ggaaacttca aaaacctgcg ggaattcgtt ttcaaaaaca tcgacggcta tttcaagatc 600

tatagcaagc ataccccaat caacctcgtg agggacctcc cccagggctt tagcgcactg 660

gagccactgg ttgacctgcc tatcggcatt aatatcacaa gatttcagac cctgctggca 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca agattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

gaaggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttatggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattga tgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggagtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcatggctcc 2040

gcctcttccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agctgagaat 2100

agcgtggcct actccaataa ttccatcgca atccctataa acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctggcacgc 2940
ctggataagg tggaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000
ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060
ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120
ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180
cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240
gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctcacaatac cttcgtgtct 3360
ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420
gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480
gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540
gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600
ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660
ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780
gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 155
<211> 1270
<212> PRT
<213> Artificial Sequence
<400> 155

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

65 70 75 80

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

85 90 95

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr

100 105 110

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

115 120 125

Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr His Lys

130 135 140

Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala

145 150 155 160

Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu

165 170 175

Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys

180 185 190

Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn

195 200 205

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

210 215 220

Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe Gly
325 330 335
Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg
340 345 350
Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

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

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

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

545 550 555 560
Gln Phe Gly Arg Asp Ile Asp Asp Thr Thr Asp Ala Val Arg Asp Pro

565 570 575

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

580 585 590

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

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp

610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe

625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser

645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln

660 665 670

Thr Gln Thr Asn Ser His Gly Ser Ala Ser Ser Val Ala Ser Gln Ser

675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr

690 695 700

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

705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln

740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln

770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser

785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu

805 810 815

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

865 870 875 880
Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
885 890 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
900 905 910
Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
915 920 925
Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys

930 935 940
Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945 950 955 960
Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
965 970 975
Ile Leu Ala Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp Arg
980 985 990
Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005
Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr
1010 1015 1020
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys
1025 1030 1035 1040
Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly
1045 1050 1055
Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070

Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr His Asn Thr Phe Val Ser

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 156

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 156

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aatttccggc actaatggca caaagcggtt cgacaatcca 240

gtcctgcctt tcaacgatgg cgtctacttt gcatctactg agaaatccaa tatcattagg 300

ggatggatct tcggcacaac cctggattct aagacccaga gcctgctgat cgtcaacaac 360

gccacaaacg tggtcattaa ggtttgcgag tttcagttct gtaacgatcc ttttctgggc 420

gtttaccata agaacaataa gagctggatg gagtccgagt ttagagtgta tagctctgca 480

aataattgta cctttgagta cgtgagccag ccctttctga tggacctgga gggaaaacaa 540

ggaaacttca aaaacctgcg ggaattcgtt ttcaaaaaca tcgacggcta tttcaagatc 600

tatagcaagc ataccccaat caacctcgtg agggacctcc cccagggctt tagcgcactg 660

gagccactgg ttgacctgcc tatcggcatt aatatcacaa gatttcagac cctgctggca 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca agattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

gaaggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttatggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattga tgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggagtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcatggctcc 2040

gcctcttccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agctgagaat 2100

agcgtggcct actccaataa ttccatcgca atccctataa acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctggcacgc 2940

ctggatcccc cagaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctcacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780
gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 157

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 157

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp Asn Pro
65 70 75 80

Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser
85 90 95

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr
100 105 110

Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile Lys Val
115 120 125

Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr His Lys
130 135 140

Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala
145 150 155 160

Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu
165 170 175

Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys

180 185 190

Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn

195 200 205

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

210 215 220

Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

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

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr
435 440 445
Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp
450 455 460
Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val
465 470 475 480
Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro

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

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

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

675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr

690 695 700

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

705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln

740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln

770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser

785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu

805 810 815

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

820 825 830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys

835 840 845

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

850 855 860

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

865 870 875 880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala

885 890 895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val

900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

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

930 935 940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val

945 950 955 960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp

965 970 975

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

980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005

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

1010 1015 1020

Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys

1025 1030 1035 1040

Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly

1045 1050 1055

Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070

Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr His Asn Thr Phe Val Ser

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 158

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 158

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aatttccggc actaatggca caaagcggtt cgacaatcca 240

gtcctgcctt tcaacgatgg cgtctacttt gcatctactg agaaatccaa tatcattagg 300

ggatggatct tcggcacaac cctggattct aagacccaga gcctgctgat cgtcaacaac 360

gccacaaacg tggtcattaa ggtttgcgag tttcagttct gtaacgatcc ttttctgggc 420

gtttaccata agaacaataa gagctggatg gagtccgagt ttagagtgta tagctctgca 480

aataattgta cctttgagta cgtgagccag ccctttctga tggacctgga gggaaaacaa 540

ggaaacttca aaaacctgcg ggaattcgtt ttcaaaaaca tcgacggcta tttcaagatc 600

tatagcaagc ataccccaat caacctcgtg agggacctcc cccagggctt tagcgcactg 660

gagccactgg ttgacctgcc tatcggcatt aatatcacaa gatttcagac cctgctggca 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca agattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

gaaggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttatggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattga tgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggcgtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcatggctcc 2040

gcctcttccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agctgagaat 2100

agcgtggcct actccaataa ttccatcgca atccctataa acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagcc ctatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggaccagca ctgcagattc cattccctat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aacccccctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctggcacgc 2940

ctggatcccc cagaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctcacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 159

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 159

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

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

20 25 30

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

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60
Phe His Ala Ile Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp Asn Pro
65 70 75 80
Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu Lys Ser
85 90 95
Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr

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

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

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

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

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp

450 455 460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val

465 470 475 480

Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro

485 490 495

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

500 505 510

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

515 520 525

Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr

530 535 540

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

545 550 555 560

Gln Phe Gly Arg Asp Ile Asp Asp Thr Thr Asp Ala Val Arg Asp Pro

565 570 575

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

580 585 590

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

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp

610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe

625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser

645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln

660 665 670

Thr Gln Thr Asn Ser His Gly Ser Ala Ser Ser Val Ala Ser Gln Ser

675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr

690 695 700

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

705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln

740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln

770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser

785 790 795 800

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

805 810 815

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

820 825 830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys

835 840 845

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

850 855 860

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

865 870 875 880

Ser Gly Trp Thr Phe Gly Ala Gly Pro Ala Leu Gln Ile Pro Phe Pro

885 890 895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val

900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

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

930 935 940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val

945 950 955 960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp

965 970 975

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

980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005

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

1010 1015 1020

Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys

1025 1030 1035 1040

Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly

1045 1050 1055

Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070

Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr His Asn Thr Phe Val Ser

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 160

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 160

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgcc 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga ggggcctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca acattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

aagggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttacggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattgc cgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggcgtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcctagaagg 2040

gccaggtccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agtggagaat 2100

agcgtggcct actccaataa ttccatcgca atccctacta acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctgagccgc 2940

ctggataagg tggaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctgacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 161

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 161

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Ala
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

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

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

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

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp

450 455 460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val

465 470 475 480

Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro

485 490 495

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

500 505 510

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

515 520 525

Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr

530 535 540

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

545 550 555 560

Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro

565 570 575

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

580 585 590

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

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp

610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe

625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser

645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln
660 665 670
Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala Ser Gln Ser

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

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

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

865 870 875 880
Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
885 890 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val

900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

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

930 935 940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val

945 950 955 960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp

965 970 975

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

980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005

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

1010 1015 1020

Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys

1025 1030 1035 1040

Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly

1045 1050 1055

Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070

Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

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

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 162

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 162

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cctgacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgcc 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga ggggcctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca acattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

aagggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttacggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattgc cgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggcgtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcctggctcc 2040

gcctcttccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agtggagaat 2100

agcgtggcct actccaataa ttccatcgca atccctacta acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctgagccgc 2940

ctggatcccc cagaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctgacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 163

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 163

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu

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

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

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

225 230 235 240
Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr
245 250 255
Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro Arg Thr Phe

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp

450 455 460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val

465 470 475 480

Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro

485 490 495

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

500 505 510

Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr
515 520 525
Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr
530 535 540
Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln

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

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

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

740 745 750
Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala

```
        755                 760                 765
Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln
    770                 775                 780
Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser
785                 790                 795                 800
Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu

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

865                 870                 875                 880
Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
                885                 890                 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
            900                 905                 910
Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
        915                 920                 925
Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys

    930                 935                 940
Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945                 950                 955                 960
Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
                965                 970                 975
Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp Arg
            980                 985                 990
Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

        995                 1000                1005
```

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

1060 1065 1070
Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg
1075 1080 1085
Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg
1090 1095 1100
Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser
1105 1110 1115 1120
Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135
Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr
1140 1145 1150
Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly
1155 1160 1165
Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn
1170 1175 1180
Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu

1185 1190 1195 1200
Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly
1205 1210 1215
Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys
1220 1225 1230
Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly
1235 1240 1245
Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys Gly

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 164

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 164

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cttcacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgcc 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga ggggcctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca acattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccccttg caatggcgtc 1440

aagggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttacggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattgc cgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggcgtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcctagaagg 2040

gccaggtccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agtggagaat 2100

agcgtggcct actccaataa ttccatcgca atccctacta acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctgagccgc 2940

ctggataagg tggaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctgacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 165

<211> 1270

<212> PRT

<213> Artificial Sequence

<400> 165

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Phe Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Ala
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

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

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

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

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

355 360 365
Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp

450 455 460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val

465 470 475 480

Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro

485 490 495

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

500 505 510

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

515 520 525

Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr

530 535 540

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

545 550 555 560

Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro

565 570 575

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

580 585 590

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

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp

610 615 620

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

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

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

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

865 870 875 880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala

885 890 895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val

900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

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

930 935 940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val

945 950 955 960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp

965 970 975

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

980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

995 1000 1005

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

1010 1015 1020

Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys

1025 1030 1035 1040

Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly

1045 1050 1055

Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe

1060 1065 1070

Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg

1075 1080 1085

Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg

1090 1095 1100

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

1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

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

1170 1175 1180

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

1185 1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly

1205 1210 1215

Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

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

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 166

<211> 3813

<212> DNA

<213> Artificial Sequence

<400> 166

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa cttcacaact 60

aggactcagc tgccaccagc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgcc 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa cgatcctttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgcgg gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga ggggcctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgcatagaa gctatctgac ccctggagac tcctctagtg ggtggactgc cggcgccgct 780

gcctactatg tgggctatct gcagccacgg acattcctgc tgaaatacaa tgagaacggg 840

acaatcacag atgctgttga ttgcgcactc gacccctgt ccgagacaaa gtgcactctc 900

aagagcttta ccgtcgagaa gggcatctat cagacctcaa acttcagggt gcagcccaca 960

gaatctatcg tgcgcttccc taatatcact aacctgtgtc ctttcggtga agtgttcaac 1020

gccaccaggt ttgctagcgt gtatgcctgg aacaggaaga ggatctctaa ctgcgtcgcc 1080

gactattccg tgctgtataa cagcgcctcc ttctccacat tcaaatgcta tggagtgagc 1140

ccgacaaaac tgaacgatct ctgctttaca aatgtctacg ccgactcttt tgtgatcaga 1200

ggggacgagg tccggcagat cgcaccagga cagacaggca acattgctga ctacaactat 1260

aagctgcctg acgacttcac aggatgtgtg atcgcatgga actcaaacaa tctggactcc 1320

aaagtcgggg gcaactataa ttacctgtat cgcctgttcc ggaagtccaa cctgaagccc 1380

ttcgagaggg acatcagtac agagatctat caggctggct ccacccctg caatggcgtc 1440

aagggcttta attgttattt tcccctgcag tcttacgggt ttcagcctac ttacggagtt 1500

gggtaccagc catacagagt ggtcgtgctc agcttcgagc tcctgcatgc tccagctaca 1560

gtttgcgggc caaagaagtc cactaacctg gtgaagaata agtgcgtcaa cttcaacttt 1620

aacgggctca ccggcaccgg cgtgctgact gagagcaaca agaagtttct gccatttcaa 1680

cagtttggac gggacattgc cgacaccacc gatgccgttc gggatccaca gaccctggaa 1740

attctggaca ttacaccgtg cagcttcggg ggcgtgagcg tgatcacacc cggaaccaat 1800

acaagcaacc aggttgccgt cctgtatcag ggcgtcaatt gcacagaagt gccagttgct 1860

atccacgcag accagctgac tcccacatgg cgggtgtata gcaccggatc caacgtgttt 1920

cagacccgcg ccggatgtct cattggggcc gagcacgtga ataacagcta cgagtgcgac 1980

atccccattg gcgccggcat ttgtgcgtct taccagactc agaccaactc tcctggctcc 2040

gcctcttccg ttgctagtca gtctattatt gcctatacca tgagcctcgg agtggagaat 2100

agcgtggcct actccaataa ttccatcgca atccctacta acttcactat ttctgtgacc 2160

accgagatcc tgcctgtgtc tatgactaag actagcgttg attgtaccat gtatatttgt 2220

ggcgactcta ccgaatgttc taacctgctg cttcagtacg gctcattttg cacacagctg 2280

aacagagccc tgactgggat cgctgtggag caggacaaga acacacagga ggtgtttgca 2340

caggtgaagc agatctataa gacccctcct attaaggatt tcggcggatt caatttctca 2400

cagattctgc cagaccccag taagccttcc aagaggagct tcatcgagga tctcctgttt 2460

aacaaggtga ccctggcaga cgccggcttt attaagcaat atggggattg cctgggcgac 2520

attgctgcca gagacctgat ttgcgcccag aaattcaatg gcctcacagt gctgccacct 2580

ctgctgaccg acgagatgat cgctcaatac actagcgcac tgctggccgg aaccatcaca 2640

tcaggctgga ccttcggggc cggagcagca ctgcagattc cattcgccat gcagatggcc 2700

tatagattca acggcattgg cgtcacacag aacgtgctgt acgaaaacca gaagctcatc 2760

gctaaccagt ttaattccgc aattggaaag atccaagatt cactcagctc aaccgcctct 2820

gcactcggaa agctgcagga cgtggtcaac cagaatgctc aggccctgaa cacactcgtc 2880

aagcagctgt cctctaactt tggcgctatc agctccgttc tgaacgacat tctgagccgc 2940

ctggatcccc cagaggctga agtccagatt gaccgcctga ttaccggccg gctgcagtct 3000

ctgcaaacat acgtgaccca gcagctgatc agagcagccg agatccgggc atccgcaaat 3060

ctggcagcaa ctaagatgag cgaatgcgtg ctgggccagt ccaagcgggt ggacttttgt 3120

ggcaagggct accacctgat gagcttcccc cagagcgccc cacatggcgt tgtttttctg 3180

cacgtgacct atgtccctgc tcaggaaaag aactttacaa ctgctcctgc tatctgccat 3240

gacggcaagg cccacttccc acgggaggga gtgtttgtgt ccaatggcac acactggttc 3300

gtgacccaga ggaacttcta tgaaccccag atcatcacca ctgacaatac cttcgtgtct 3360

ggaaattgcg acgtcgtgat cggcatcgtt aacaacaccg tgtacgaccc tctccagcca 3420

gagctggact cctttaagga ggaactggat aagtatttta agaaccacac aagcccagat 3480

gtggatctcg gggacatctc cggaattaac gcctccgtgg tgaatatcca gaaggagatt 3540

gaccgcctaa atgaagttgc caagaacctc aatgagtctc tgattgatct gcaggaactg 3600

ggcaagtatg agcagtatat caaatggccc tggtacattt ggctggggtt tatcgccgga 3660

ctgattgcca tcgtcatggt gaccatcatg ctgtgttgca tgacctcctg ttgttcctgt 3720

ctgaagggct gctgtagttg cggctcttgc tgtaaattcg acgaagatga tagcgagccc 3780

gtgctgaagg gcgtgaagct gcattatacc tga 3813

<210> 167

<211> 1270

<212> PRT

<213> Artificial Sequence

<220><223> Artificial Sequence

<400> 167

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Phe Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Ala
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Gly Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr

305 310 315 320

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

325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg

340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser

355 360 365

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

370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg

385 390 395 400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala

405 410 415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala

420 425 430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr

435 440 445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp

450 455 460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val

465 470 475 480

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

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

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

Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala Ser Gln Ser
675 680 685
Ile Ile Ala Tyr Thr Met Ser Leu Gly Val Glu Asn Ser Val Ala Tyr
690 695 700
Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile Ser Val Thr
705 710 715 720
Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr

725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln
740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala
755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln
770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser
785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu
805 810 815

Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys
820 825 830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys
835 840 845

Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
850 855 860

Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr
865 870 875 880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
885 890 895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
900 905 910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
915 920 925

Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys
930 935 940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945 950 955 960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
965 970 975

Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp Arg
980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln
995 1000 1005
Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala Thr
1010 1015 1020
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys
1025 1030 1035 1040
Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala Pro His Gly
1045 1050 1055

Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu Lys Asn Phe
1060 1065 1070
Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His Phe Pro Arg
1075 1080 1085
Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val Thr Gln Arg
1090 1095 1100
Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe Val Ser
1105 1110 1115 1120

Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp
1125 1130 1135
Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr
1140 1145 1150
Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly
1155 1160 1165
Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn
1170 1175 1180

Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1185 1190 1195 1200
Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly
1205 1210 1215
Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys

1220 1225 1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly

1235 1240 1245

Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys Gly

1250 1255 1260

Val Lys Leu His Tyr Thr

1265 1270

<210> 168

<211> 3822

<212> DNA

<213> Artificial Sequence

<220><223> Artificial Sequence

<400> 168

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa ctttacaaac 60

aggactcagc tgccatccgc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa ctacccttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgagc gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcac cattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtca agggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

tacggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920

aacgtgtttc agacccgcgc cggatgtctc attggggccg agtacgtgaa taacagctac 1980

gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040

cctagaaggg ccaggtccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100

gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160

tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220

tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340

gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400

aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460

ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520

ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580

ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640

accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760

aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820

accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880

acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940

ctgagccgcc tggataaggt ggaggctgaa gtccagattg accgcctgat taccggccgg 3000

ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060

tccgcaaatc tggcagcaat caagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180

gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240

atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300

cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360

ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420

ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480

agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccttcgt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 169

<211> 1273

<212> PRT

<213> Artificial Sequence

<220><223> Artificial Sequence

<400> 169

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Phe Thr Asn Arg Thr Gln Leu Pro Ser Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95
Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110
Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125
Lys Val Cys Glu Phe Gln Phe Cys Asn Tyr Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160
Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175
Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Ser Glu Phe
180 185 190
Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255
Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285
Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300
Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
370 375 380
Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Thr Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605

Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu Tyr Val
645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn

915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala

930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn

945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln

980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val

995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu

1010 1015 1020

Ala Ala Ile Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala

1045 1050 1055

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu

1060 1065 1070

Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val
1090 1095 1100

Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp
1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Phe Val Asn Ile Gln Lys Glu Ile Asp
1170 1175 1180
Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1185 1190 1195 1200
Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile
1205 1210 1215
Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile
1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val
1250 1255 1260
Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 170
<211> 3822
<212> DNA
<213> Artificial Sequence
<220><223> Artificial Sequence
<400> 170

atgttcgtct tcctcgtgct gctcccactc gtttcttccc agtgtgtcaa ctttacaaac 60

aggactcagc tgccatccgc ctacaccaac tccttcacca gaggcgtgta ttacccagac 120

aaggtgttta gaagcagcgt gctgcactct acccaggacc tctttctgcc ctttttcagc 180

aacgtgacat ggtttcacgc aattcacgtg tccggcacta atggcacaaa gcggttcgac 240

aatccagtcc tgcctttcaa cgatggcgtc tactttgcat ctactgagaa atccaatatc 300

attaggggat ggatcttcgg cacaaccctg gattctaaga cccagagcct gctgatcgtc 360

aacaacgcca caaacgtggt cattaaggtt tgcgagtttc agttctgtaa ctacccttt 420

ctgggcgtgt attatcataa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

agctctgcaa ataattgtac ctttgagtac gtgagccagc cctttctgat ggacctggag 540

ggaaaacaag gaaacttcaa aaacctgagc gaattcgttt tcaaaaacat cgacggctat 600

ttcaagatct atagcaagca taccccaatc aacctcgtga gggacctccc ccagggcttt 660

agcgcactgg agccactggt tgacctgcct atcggcatta atatcacaag atttcagacc 720

ctgctggcac tgcatagaag ctatctgacc cctggagact cctctagtgg gtggactgcc 780

ggcgccgctg cctactatgt gggctatctg cagccacgga cattcctgct gaaatacaat 840

gagaacggga caatcacaga tgctgttgat tgcgcactcg acccccctgtc cgagacaaag 900

tgcactctca agagctttac cgtcgagaag ggcatctatc agacctcaaa cttcagggtg 960

cagcccacag aatctatcgt gcgcttccct aatatcacta acctgtgtcc tttcggtgaa 1020

gtgttcaacg ccaccaggtt tgctagcgtg tatgcctgga acaggaagag gatctctaac 1080

tgcgtcgccg actattccgt gctgtataac agcgcctcct tctccacatt caaatgctat 1140

ggagtgagcc cgacaaaact gaacgatctc tgctttacaa atgtctacgc cgactctttt 1200

gtgatcagag gggacgaggt ccggcagatc gcaccaggac agacaggcac cattgctgac 1260

tacaactata agctgcctga cgacttcaca ggatgtgtga tcgcatggaa ctcaaacaat 1320

ctggactcca aagtcggggg caactataat tacctgtatc gcctgttccg gaagtccaac 1380

ctgaagccct tcgagaggga catcagtaca gagatctatc aggctggctc cacccttgc 1440

aatggcgtca agggctttaa ttgttatttt cccctgcagt cttacgggtt tcagcctact 1500

tacggagttg ggtaccagcc atacagagtg gtcgtgctca gcttcgagct cctgcatgct 1560

ccagctacag tttgcgggcc aaagaagtcc actaacctgg tgaagaataa gtgcgtcaac 1620

ttcaacttta acgggctcac cggcaccggc gtgctgactg agagcaacaa gaagtttctg 1680

ccatttcaac agtttggacg ggacattgcc gacaccaccg atgccgttcg ggatccacag 1740

accctggaaa ttctggacat tacaccgtgc agcttcgggg gcgtgagcgt gatcacaccc 1800

ggaaccaata caagcaacca ggttgccgtc ctgtatcagg gcgtcaattg cacagaagtg 1860

ccagttgcta tccacgcaga ccagctgact cccacatggc gggtgtatag caccggatcc 1920
aacgtgtttc agacccgcgc cggatgtctc attggggccg agtacgtgaa taacagctac 1980
gagtgcgaca tccccattgg cgccggcatt tgtgcgtctt accagactca gaccaactct 2040
cctggctccg cctcttccgt tgctagtcag tctattattg cctataccat gagcctcgga 2100
gctgagaata gcgtggccta ctccaataat tccatcgcaa tccctactaa cttcactatt 2160
tctgtgacca ccgagatcct gcctgtgtct atgactaaga ctagcgttga ttgtaccatg 2220
tatatttgtg gcgactctac cgaatgttct aacctgctgc ttcagtacgg ctcattttgc 2280

acacagctga acagagccct gactgggatc gctgtggagc aggacaagaa cacacaggag 2340
gtgtttgcac aggtgaagca gatctataag acccctccta ttaaggattt cggcggattc 2400
aatttctcac agattctgcc agaccccagt aagccttcca agaggagctt catcgaggat 2460
ctcctgttta acaaggtgac cctggcagac gccggcttta ttaagcaata tggggattgc 2520
ctgggcgaca ttgctgccag agacctgatt tgcgcccaga aattcaatgg cctcacagtg 2580
ctgccacctc tgctgaccga cgagatgatc gctcaataca ctagcgcact gctggccgga 2640
accatcacat caggctggac cttcggggcc ggagcagcac tgcagattcc attcgccatg 2700

cagatggcct atagattcaa cggcattggc gtcacacaga acgtgctgta cgaaaaccag 2760
aagctcatcg ctaaccagtt taattccgca attggaaaga tccaagattc actcagctca 2820
accgcctctg cactcggaaa gctgcaggac gtggtcaacc agaatgctca ggccctgaac 2880
acactcgtca agcagctgtc ctctaacttt ggcgctatca gctccgttct gaacgacatt 2940
ctgagccgcc tggatccccc agaggctgaa gtccagattg accgcctgat taccggccgg 3000
ctgcagtctc tgcaaacata cgtgacccag cagctgatca gagcagccga gatccgggca 3060
tccgcaaatc tggcagcaat caagatgagc gaatgcgtgc tgggccagtc caagcgggtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccccc agagcgcccc acatggcgtt 3180
gtttttctgc acgtgaccta tgtccctgct caggaaaaga actttacaac tgctcctgct 3240
atctgccatg acggcaaggc ccacttccca cgggagggag tgtttgtgtc caatggcaca 3300
cactggttcg tgacccagag gaacttctat gaaccccaga tcatcaccac tgacaatacc 3360
ttcgtgtctg gaaattgcga cgtcgtgatc ggcatcgtta acaacaccgt gtacgaccct 3420
ctccagccag agctggactc ctttaaggag gaactggata agtattttaa gaaccacaca 3480
agcccagatg tggatctcgg ggacatctcc ggaattaacg cctccttcgt gaatatccag 3540

aaggagattg accgcctaaa tgaagttgcc aagaacctca atgagtctct gattgatctg 3600

caggaactgg gcaagtatga gcagtatatc aaatggccct ggtacatttg gctggggttt 3660

atcgccggac tgattgccat cgtcatggtg accatcatgc tgtgttgcat gacctcctgt 3720

tgttcctgtc tgaagggctg ctgtagttgc ggctcttgct gtaaattcga cgaagatgat 3780

agcgagcccg tgctgaaggg cgtgaagctg cattatacct ga 3822

<210> 171

<211> 1273

<212> PRT

<213> Artificial Sequence

<220><223> Artificial Sequence

<400> 171

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

Asn Phe Thr Asn Arg Thr Gln Leu Pro Ser Ala Tyr Thr Asn Ser Phe

20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile

115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Tyr Pro Phe Leu Gly Val Tyr

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Ser Glu Phe

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser

245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro

260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala

275 280 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys

290 295 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val

305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala

340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu

355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro

370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly

405 410 415

Thr Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe

450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys

515 520 525
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605
Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620
His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625 630 635 640
Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu Tyr Val

645 650 655
Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Gly Ser Ala Ser Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr

900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val

965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu
1010 1015 1020
Ala Ala Ile Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val

1025 1030 1035 1040
Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1045 1050 1055
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln Glu
1060 1065 1070
Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys Ala His
1075 1080 1085
Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His Trp Phe Val

1090 1095 1100
Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr Thr Asp Asn Thr
1105 1110 1115 1120
Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Val Asn Asn Thr
1125 1130 1135
Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu
1140 1145 1150
Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp

1155 1160 1165

Ile Ser Gly Ile Asn Ala Ser Phe Val Asn Ile Gln Lys Glu Ile Asp

1170 1175 1180

Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu

1185 1190 1195 1200

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile

1205 1210 1215

Trp Leu Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile

1220 1225 1230

Met Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val

1250 1255 1260

Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 172

<211> 4058

<212> RNA

<213> Artificial Sequence

<220><223> Artificial Sequence

<400> 172

ggacagaucg ccuggagacg ccauccacgc uguuuugacc uccauagaag acaccgggac 60

cgauccagcc uccgcggccg ggaacggugc auuggaacgc ggauuccccg ugccaagagu 120

gacucaccgu ccuugacacg auguucgucu uccucgugcu gcucccacuc guuucuuccc 180

agugugucaa cuucacaacu aggacucagc ugccaccagc cuacaccaac uccuucacca 240

gaggcgugua uuacccagac aagguguuua gaagcagcgu gcugcacucu acccaggacc 300

ucuuucugcc cuuuuucagc aacgugacau gguuucacgc aauucacgug uccggcacua 360

auggcacaaa gcgguucgcc aauccagucc ugccuuucaa cgauggcguc uacuuugcau 420

cuacugagaa auccaauauc auuaggggau ggaucuucgg cacaacccug gauucuaaga 480

cccagagccu gcugaucguc aacaacgcca caaacguggu cauuaagguu ugcgaguuuc 540

aguucuguaa cgauccuuuu cugggcgugu auuaucauaa gaacaauaag agcuggaugg 600

aguccgaguu uagaguguau agcucugcaa auaauuguac cuuugaguac gugagccagc 660

ccuuucugau ggaccuggag ggaaaacaag gaaacuucaa aaaccugcgg gaauucguuu 720

ucaaaaacau cgacggcuau uucaagaucu auagcaagca uaccccaauc aaccucguga 780

ggggccuccc ccagggcuuu agcgcacugg agccacuggu ugaccugccu aucggcauua 840

auaucacaag auuucagacc cugcauagaa gcuaucugac cccuggagac uccucuagug 900

gguggacugc cggcgccgcu gccuacuaug ugggcuaucu gcagccacgg acauuccugc 960

ugaaauacaa ugagaacggg acaaucacag augcuguuga uugcgcacuc gacccccugu 1020

ccgagacaaa gugcacucuc aagagcuuua ccgucgagaa gggcaucuau cagaccucaa 1080

acuucagggu gcagcccaca gaaucuaucg ugcgcuuccc uaauaucacu aaccuguguc 1140

cuuucgguga aguguucaac gccaccaggu uugcuagcgu guaugccugg aacaggaaga 1200

ggaucucuaa cugcgucgcc gacuauuccg ugcuguauaa cagcgccucc uucuccacau 1260

ucaaaugcua uggagugagc ccgacaaaac ugaacgaucu cugcuuuaca aaugucuacg 1320

ccgacucuuu ugugaucaga ggggacgagg uccggcagau cgcaccagga cagacaggca 1380

acauugcuga cuacaacuau aagcugccug acgacuucac aggaugugug aucgcaugga 1440

acucaaacaa ucuggacucc aaagucgggg gcaacuauaa uuaccuguau cgccuguucc 1500

ggaaguccaa ccugaagccc uucgagaggg acaucaguac agagaucuau caggcuggcu 1560

ccaccccuug caauggcguc aagggcuuua auuguuauuu uccccugcag ucuuacgggu 1620

uucagccuac uuacggaguu ggguaccagc cauacagagu ggucgugcuc agcuucgagc 1680

uccugcaugc uccagcuaca guuugcgggc caaagaaguc cacuaaccug gugaagaaua 1740

agugcgucaa cuucaacuuu aacgggcuca ccggcaccgg cgugcugacu gagagcaaca 1800

agaaguuucu gccauuucaa caguuuggac gggacauugc cgacaccacc gaugccguuc 1860

gggauccaca gacccuggaa auucuggaca uuacaccgug cagcuucggg ggcgugagcg 1920

ugaucacacc cggaaccaau acaagcaacc agguugccgu ccuguaucag ggcgucaauu 1980

gcacagaagu gccaguugcu auccacgcag accagcugac ucccacaugg cggguguaua 2040

gcaccggauc caacguguuu cagacccgcg ccggaugucu cauuggggcc gagcacguga 2100

auaacagcua cgagugcgac aucccauug gcgccggcau uugugcgucu uaccagacuc 2160

agaccaacuc uccuggcucc gccucuuccg uugcuaguca gucuauuauu gccuauacca 2220

ugagccucgg aguggagaau agcguggccu acuccaauaa uuccaucgca aucccuacua 2280

acuucacuau uucugugacc accgagaucc ugccuguguc uaugacuaag acuagcguug 2340

auuguaccau guauauuugu ggcgacucua ccgaauguuc uaaccugcug cuucaguacg 2400

gcucauuuug cacacagcug aacagagccc ugacugggau cgcuguggag caggacaaga 2460

acacacagga gguguuugca caggugaagc agaucuauaa gaccccuccu auuaaggauu 2520

ucggcggauu caauuucuca cagauucugc cagaccccag uaagccuucc aagaggagcu 2580

ucaucgagga ucuccuguuu aacaagguga cccuggcaga cgccggcuuu auuaagcaau 2640
auggggauug ccuggggcgac auugcugcca gagaccugau uugcgcccag aaauucaaug 2700
gccucacagu gcugccaccu cugcugaccg acgagaugau cgcucaauac acuagcgcac 2760
ugcuggccgg aaccaucaca ucaggcugga ccuucgggggc cggagcagca cugcagauuc 2820
cauucgccau gcagauggcc uauagauuca acggcauugg cgucacacag aacgugcugu 2880
acgaaaacca gaagcucauc gcuaaccagu uuaauuccgc aauuggaaag auccaagauu 2940
cacucagcuc aaccgccucu gcacucggaa agcugcagga cguggucaac cagaaugcuc 3000

aggcccugaa cacacucguc aagcagcugu ccucuaacuu uggcgcuauc agcuccguuc 3060
ugaacgacau ucugagccgc cuggaucccc cagaggcuga aguccagauu gaccgccuga 3120
uuaccggccg gcugcagucu cugcaaacau acgugaccca gcagcugauc agagcagccg 3180
agauccgggc auccgcaaau cuggcagcaa cuaagaugag cgaaugcgug cugggccagu 3240
ccaagcgggu ggacuuuugu ggcaagggcu accaccugau gagcuucccc cagagcgccc 3300
cacauggcgu uguuuuucug cacgugaccu augucccugc ucaggaaaag aacuuuacaa 3360
cugcuccugc uaucugccau gacggcaagg cccacuuccc acgggaggga guguuugugu 3420

ccaauggcac acacugguuc gugacccaga ggaacuucua ugaaccccag aucaucacca 3480
cugacaauac cuucgugucu ggaaauugcg acgucgugau cggcaucguu aacaacaccg 3540
uguacgaccc ucuccagcca gagcuggacu ccuuuaagga ggaacuggau aaguauuuua 3600
agaaccacac aagcccagau guggaucucg gggacaucuc cggaauuaac gccuccgugg 3660
ugaauaucca gaaggagauu gaccgccuaa augaaguugc caagaaccuc aaugagucuc 3720
ugauugaucu gcaggaacug ggcaaguaug agcaguauau caaauggccc ugguacauuu 3780
ggcugggguu uaucgccgga cugauugcca ucgucauggu gaccaucaug cuguguugca 3840

ugaccuccug uuguuccugu cugaagggcu gcuguaguug cggcucuugc uguaaauucg 3900
acgaagauga uagcgagccc gugcugaagg gcgugaagcu gcauuauacc ugacgggugg 3960
caucccugug accccucccc agugccucuc cuggcccugg aaguugccac uccagugccc 4020
accagccuug uccuaauaaa auuaaguugc aucaagcu 4058

<210> 173
<211> 3813
<212> RNA
<213> Artificial Sequence
<220><223> Artificial Sequence

<400> 173

auguucgucu uccucgugcu gcucccacuc guuucuuccc agugugucaa cuucacaacu 60

aggacucagc ugccaccagc cuacaccaac uccuucacca gaggcgugua uuacccagac 120

aagguguuua gaagcagcgu gcugcacucu acccaggacc ucuuucugcc cuuuuucagc 180

aacgugacau gguuucacgc aauucacgug uccggcacua auggcacaaa gcgguucgcc 240

aauccagucc ugccuuucaa cgauggcguc uacuuugcau cuacugagaa auccaauauc 300

auuaggggau ggaucuucgg cacaacccug gauucuaaga cccagagccu gcugaucguc 360

aacaacgcca caaacguggu cauuaagguu ugcgaguuuc aguucuguaa cgauccuuuu 420

cugggcgugu auuaucauaa gaacaauaag agcuggaugg aguccgaguu uagaguguau 480

agcucugcaa auaauuguac cuuugaguac gugagccagc ccuuucugau ggaccuggag 540

ggaaaacaag gaaacuucaa aaaccugcgg gaauucguuu ucaaaaacau cgacggcuau 600

uucaagaucu auagcaagca uaccccaauc aaccucguga ggggccuccc ccagggcuuu 660

agcgcacugg agccacuggu ugaccugccu aucggcauua auaucacaag auuucagacc 720

cugcauagaa gcuaucugac cccuggagac uccucuagug gguggacugc cggcgccgcu 780

gccuacuaug ugggcuaucu gcagccacgg acauuccugc ugaaauacaa ugagaacggg 840

acaaucacag augcuguuga uugcgcacuc gacccccugu ccgagacaaa gugcacucuc 900

aagagcuuua ccgucgagaa gggcaucuau cagaccucaa acuucagggu gcagcccaca 960

gaaucuaucg ugcgcuuccc uaauaucacu aaccuguguc cuuucgguga aguguucaac 1020

gccaccaggu uugcuagcgu guaugccugg aacaggaaga ggaucucuaa cugcgucgcc 1080

gacuauuccg ugcuguauaa cagcgccucc uucuccacau ucaaaugcua uggagugagc 1140

ccgacaaaac ugaacgaucu cugcuuuaca aaugucuacg ccgacucuuu ugugaucaga 1200

ggggacgagg uccggcagau cgcaccagga cagacaggca acauugcuga cuacaacuau 1260

aagcugccug acgacuucac aggaugugug aucgcaugga acucaaacaa ucuggacucc 1320

aaagucgggg gcaacuauaa uuaccuguau cgccuguucc ggaaguccaa ccugaagccc 1380

uucgagaggg acaucaguac agagaucuau caggcuggcu ccaccccuug caauggcguc 1440

aagggcuuua auuguuauuu uccccugcag ucuuacgggu uucagccuac uuacggaguu 1500

ggguaccagc cauacagagu ggucgugcuc agcuucgagc uccugcaugc uccagcuaca 1560

guuugcgggc caaagaaguc cacuaaccug gugaagaaua agugcgucaa cuucaacuuu 1620

aacgggcuca ccggcaccgg cgugcugacu gagagcaaca agaaguuucu gccauuucaa 1680

caguuuggac gggacauugc cgacaccacc gaugccguuc gggauccaca gacccuggaa 1740

auucuggaca uuacaccgug cagcuucggg ggcgugagcg ugaucacacc cggaaccaau 1800

acaagcaacc agguugccgu ccuguaucag ggcgucaauu gcacagaagu gccaguugcu 1860

auccacgcag accagcugac ucccacaugg cgggguguaua gcaccggauc caacguguuu 1920

cagacccgcg ccggaugucu cauuggggcc gagcacguga auaacagcua cgagugcgac 1980

aucccccauug gcgccggcau uugugcgucu uaccagacuc agaccaacuc uccuggcucc 2040

gccucuuccg uugcuaguca gucuauuauu gccuauacca ugagccucgg aguggagaau 2100

agcguggccu acuccaauaa uuccaucgca aucccuacua acuucacuau uucugugacc 2160

accgagaucc ugccuguguc uaugacuaag acuagcguug auuguaccau guauauuugu 2220

ggcgacucua ccgaauguuc uaaccugcug cuucaguacg gcucauuuug cacacagcug 2280

aacagagccc ugacugggau cgcuguggag caggacaaga acacacagga gguguuugca 2340

caggugaagc agaucuauaa gaccccuccu auuaaggauu ucggcggauu caauuucuca 2400

cagauucugc cagaccccag uaagccuucc aagaggagcu ucaucgagga ucuccuguuu 2460

aacaagguga cccuggcaga cgccggcuuu auuaagcaau auggggauug ccugggcgac 2520

auugcugcca gagaccugau uugcgcccag aaauucaaug gccucacagu gcugccaccu 2580

cugcugaccg acgagaugau cgcucaauac acuagcgcac ugcuggccgg aaccaucaca 2640

ucaggcugga ccuucggggc cggagcagca cugcagauuc cauucgccau gcagauggcc 2700

uauagauuca acggcauugg cgucacacag aacgugcugu acgaaaacca gaagcucauc 2760

gcuaaccagu uuaauuccgc aauuggaaag auccaagauu cacucagcuc aaccgccucu 2820

gcacucggaa agcugcagga cguggucaac cagaaugcuc aggcccugaa cacacucguc 2880

aagcagcugu ccucuaacuu uggcgcuauc agcuccguuc ugaacgacau ucugagccgc 2940

cuggaucccc cagaggcuga aguccagauu gaccgccuga uuaccggccg gcugcagucu 3000

cugcaaacau acgugaccca gcagcugauc agagcagccg agauccgggc auccgcaaau 3060

cuggcagcaa cuaagaugag cgaaugcgug cugggccagu ccaagcgggu ggacuuuugu 3120

ggcaagggcu accaccugau gagcuucccc cagagcgccc cacauggcgu uguuuuucug 3180

cacgugaccu augucccugc ucaggaaaag aacuuuacaa cugcuccugc uaucugccau 3240

gacggcaagg cccacuuccc acgggaggga guguuugugu ccaauggcac acacugguuc 3300

gugacccaga ggaacuucua ugaaccccag aucaucacca cugacaauac cuucgugucu 3360

ggaaauugcg acgucgugau cggcaucguu aacaacaccg uguacgaccc ucuccagcca 3420

gagcuggacu ccuuuaagga ggaacuggau aaguauuuua agaaccacac aagcccagau 3480

guggaucucg gggacaucuc cggaauuaac gccuccgugg ugaauaucca gaaggagauu 3540

gaccgccuaa augaaguugc caagaaccuc aaugagucuc ugauugaucu gcaggaacug 3600

ggcaaguaug agcaguauau caaauggccc ugguacauuu ggcugggguu uaucgccgga 3660

cugauugcca ucgucauggu gaccaucaug cuguguugca ugaccuccug uuguuccugu 3720

cugaagggcu gcuguaguug cggcucuugc uguaaauucg acgaagauga uagcgagccc 3780

gugcugaagg gcgugaagcu gcauuauacc uga 3813