(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0006825
(43) 공개일자 2023년01월11일

(51) 국제특허분류(Int. Cl.)
A61K 39/12 (2006.01) *A61K 39/00* (2006.01)
A61P 31/12 (2006.01) *A61P 31/16* (2006.01)
A61P 31/18 (2006.01) *C12N 15/86* (2006.01)
(52) CPC특허분류
A61K 39/12 (2013.01)
A61P 31/12 (2018.01)
(21) 출원번호 10-2022-7038618
(22) 출원일자(국제) 2021년04월05일
심사청구일자 없음
(85) 번역문제출일자 2022년11월03일
(86) 국제출원번호 PCT/US2021/025828
(87) 국제공개번호 WO 2021/203104
국제공개일자 2021년10월07일
(30) 우선권주장
63/005,160 2020년04월03일 미국(US)

(71) 출원인
그릿스톤 바이오, 인코포레이티드
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959
(72) 발명자
퍼거슨, 앤드류
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
루쏘, 라파엘
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
(뒷면에 계속)
(74) 대리인
김영철, 김 순 영

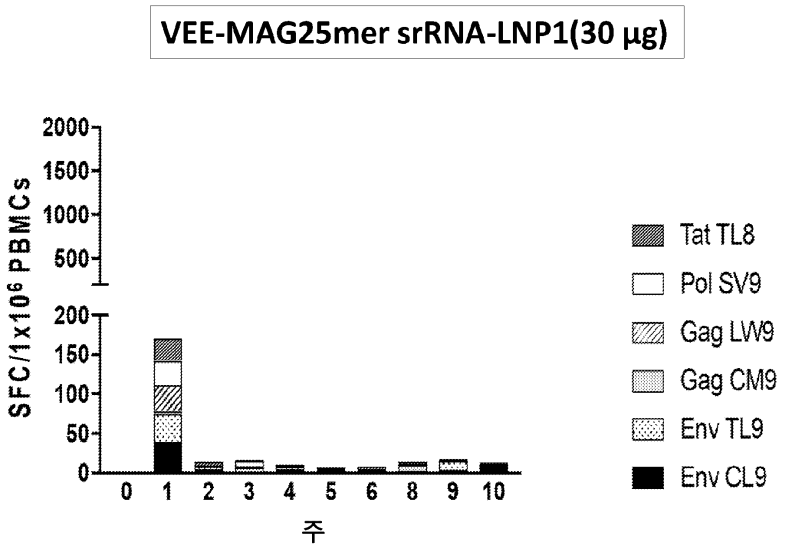
전체 청구항 수 : 총 387 항

(54) 발명의 명칭 **전염성 질병 항원 및 백신**

(57) 요 약

항원-코딩 핵산 서열 및/또는 항원 펩티드를 포함하는 조성물이 본원에 개시된다. 또한, 이종 프라임/부스트 백신 접종 전략을 위한 벡터 및 방법을 포함하는 백신으로서의 용도를 포함하는 조성물과 관련된 뉴클레오티드, 세포 및 방법이 개시된다.

대 표 도 - 도1a



(52) CPC특허분류
A61P 31/16 (2018.01)
A61P 31/18 (2018.01)
C12N 15/86 (2013.01)
A61K 2039/53 (2013.01)
A61K 2039/545 (2013.01)
A61K 2039/572 (2013.01)
C12N 2710/10343 (2013.01)
C12N 2740/16034 (2013.01)
C12N 2740/16134 (2013.01)

(72) 발명자
엘렌스키, 로만
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
선, 제임스 신
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
데이비스, 매튜 조셉
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
주스, 카린
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
라파포트, 에이미 레이첼
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
스캘란, 시아란 다니엘
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
기틀린, 레오니드
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드
팔머, 크리스틴 데니스
미국, 캘리포니아 94608, 에머리빌, 스위트 300, 호튼 스트리트 5959, 그릿스톤 바이오, 인코포레이티드

명 세 서

청구범위

청구항 1

자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물로서,

여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은:

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, RNA 알파바이러스 백본, 및

(b) 카세트로서, 여기서 카세트는

(i) 다음을 포함하는 적어도 하나의 항원-코딩 핵산 서열:

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열;

(ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제2 프로모터 뉴클레오티드 서열; 및

(iii) 알파바이러스에 대해 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열인, 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하는, 카세트를 포함하는, 자가-복제 알파바이러스-기반 발현 시스템; 및

(B) 상기 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 조성물.

청구항 2

자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물로서,

여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은 서열번호: 6에 제시된 핵산 서열을 포함하고, 여기서 RNA 알파바이러스 백본 서열은 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열을 포함하고, 여기서 26S 프로모터 서열은 RNA 알파바이러스 백본에 내인성이며, 여기서 폴리(A) 서열은 RNA 알파바이러스 백본에 내인성인, RNA 알파바이러스 백본; 및

(b) 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합된 카세트로서, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열에 작동가능하게 연결되고, 여기서 카세트는

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열;을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하는, 자가-복제 알파바이러스-기반 발현 시스템; 및

(B) 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 조성물.

청구항 3

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 4

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 5

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 6

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 1-10 사이, 1-20 사이, 1-30 사이, 1-40 사이, 1-50 사이, 1-100 사이, 1-200 사이, 1-300 사이, 1-400 사이, 또는 1-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 7

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 2-10 사이, 2-20 사이, 2-30 사이, 2-40 사이, 2-50 사이, 2-100 사이, 2-200 사이, 2-300 사이, 2-400 사이, 또는 2-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 8

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 전장 단백질을 코딩하는 핵산 서열을 포함하는, 조성물.

청구항 9

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 단백질 도메인을 코딩하는 핵산 서열을 포함하는, 조성물.

청구항 10

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 단백질 서브유닛을 코딩하는 핵산 서열을 포함하는, 조성물.

청구항 11

제 1항 또는 제 2항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인 및 이의 조합으로부터 선택된 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함하는, 조성물.

청구항 12

제 1항 내지 제 11항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 면역 반응을 자극할 수 있는, 조성물.

청구항 13

제 1항 내지 제 11항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 T 세포 반응을 자극할 수 있는, 조성물.

청구항 14

제 1항 내지 제 11항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 B 세포 반응을 자극할 수 있는, 조성물.

청구항 15

제 1항 내지 제 11항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 T 세포 반응 및 B 세포 반응을 자극할 수 있는, 조성물.

청구항 16

제 1항 내지 제 15항 중 어느 한 항에 있어서, 상기 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 조성물.

청구항 17

제 1항 또는 제 3항 내지 제 16항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'부터 3'로 하기를 포함하는 식으로 설명되며,

P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

여기서, P는 제 2 프로모터 뉴클레오티드 서열을 포함하고, 여기서 a = 0 또는 1이고,

N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고,

L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고,

L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고,

G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고,

G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고,

U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고,

X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고, 그리고

Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열인, 조성물.

청구항 18

제 17항에 있어서, 각각의 X에 대해 상응하는 N_c가 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열인, 조성물.

청구항 19

제 17항 또는 제 18항에 있어서, 각각의 Y에 대해 상응하는 U_f가 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열인, 조성물.

청구항 20

제 17항 내지 제 19항 중 어느 한 항에 있어서,

a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2이고,

적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 제공되는 단일 26S 프로모터 뉴클

레오티드 서열이고,

적어도 하나의 폴리아데닐화 폴리(A) 서열은 RNA 알파바이러스 백본에 의해 제공되는 적어도 100개의 연속적인 A 뉴클레오티드의 폴리(A) 서열이고,

카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합되고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열에 작동가능하게 연결되고,

각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고,

L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

U는 각각 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고,

RNA 알파바이러스 백본은 서열번호: 6에 제시된 서열이고, 그리고

각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 길이가 13개 내지 25개 아미노산인 폴리펩티드를 코딩하는, 조성물.

청구항 21

제 1항 내지 제 20항 중 어느 한 항에 있어서, 상기 LNP가 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, PEG-기반 코트 지질, 또는 이들의 조합으로 이루어진 군으로부터 선택된 지질을 포함하는, 조성물.

청구항 22

제 1항 내지 제 21항 중 어느 한 항에 있어서, 상기 LNP가 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, 및 PEG-기반 코트 지질을 포함하는, 조성물.

청구항 23

제 21항 또는 제 22항에 있어서, 상기 이온화가능한 아미노 지질이 MC3-유사(디리놀레일메틸-4-디메틸아미노부티레이트) 분자를 포함하는, 조성물.

청구항 24

제 1항 내지 제 23항 중 어느 한 항에 있어서, LNP-캡슐화 발현 시스템이 약 100 nm의 직경을 갖는, 조성물.

청구항 25

제 1항 내지 제 24항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여용으로 제형화되는, 조성물.

청구항 26

제 1항 내지 제 25항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM) 투여용으로 제형화되는, 조성물.

청구항 27

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 26항 중 어느 한 항에 있어서, 상기 카세트가 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합된, 조성물.

청구항 28

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 27항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 상기 카세트에 작동가능하게 연결된, 조성물.

청구항 29

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 28항 중 어느 한 항에 있어서, 상기 적어도 하나의 벡터가 하나 이상의 +-가닥 RNA 벡터를 포함하는, 조성물.

청구항 30

제 29항에 있어서, 상기 하나 이상의 +-가닥 RNA 벡터가 5'7-메틸구아노신(m7g) 캡을 포함하는, 조성물.

청구항 31

제 29항 또는 제 30항에 있어서, 상기 하나 이상의 +-가닥 RNA 벡터가 시험관내 전사에 의해 생성되는, 조성물.

청구항 32

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 31항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 포유동물 세포 내에서 자가-복제되는, 조성물.

청구항 33

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 32항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함하는, 조성물.

청구항 34

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 32항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 베네수엘라 말 뇌염 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함하는, 조성물.

청구항 35

제 33항 또는 제 34항에 있어서, 상기 RNA 알파바이러스 백본이 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩되는 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열, 비구조적 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자 및 nsP4 유전자를 적어도 포함하는, 조성물.

청구항 36

제 33항 또는 제 34항에 있어서, 상기 RNA 알파바이러스 백본이 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩되는 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 및 폴리(A) 서열을 적어도 포함하는, 조성물.

청구항 37

제 35항 또는 제 36항에 있어서, 비구조적 단백질-매개 증폭을 위한 서열이 알파바이러스 5' UTR, 51-nt CSE, 24-nt CSE, 26S 서브게놈 프로모터 서열, 19-nt CSE, 알파바이러스 3' UTR, 또는 이들의 조합으로 이루어진 군으로부터 선택되는, 조성물.

청구항 38

제 35항 내지 제 37항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 구조적 비리온 단백질 캡시드, E2 및 E1을 코딩하지 않는, 조성물.

청구항 39

제 38항에 있어서, 상기 카세트가 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열 내

구조적 비리온 단백질 대신 삽입되는, 조성물.

청구항 40

제 33항 또는 제 34항에 있어서, 상기 베네수엘라 말 뇌염 바이러스가 서열번호: 3 또는 서열번호: 5의 서열을 포함하는, 조성물.

청구항 41

제 33항 또는 제 34항에 있어서, 상기 베네수엘라 말 뇌염 바이러스가 염기쌍 7544와 11175 사이의 결실을 추가로 포함하는 서열번호: 3 또는 서열번호: 5의 서열을 포함하는, 조성물.

청구항 42

제 41항에 있어서, 상기 RNA 알파바이러스 백본이 서열번호: 6 또는 서열번호: 7에 제시된 서열을 포함하는, 조성물.

청구항 43

제 41항 또는 제 42항에 있어서, 상기 카세트가 서열번호: 3 또는 서열번호: 5의 서열에서 제시된 바와 같이 염기쌍 7544 및 11175 사이의 결실을 대체하기 위해 위치 7544에 삽입되는, 조성물.

청구항 44

제 39항 내지 제 43항 중 어느 한 항에 있어서, 상기 카세트의 삽입이 nsP1-4 유전자 및 적어도 하나의 핵산 서열을 포함하는 폴리시스트론 RNA의 전사를 제공하고, 여기서 nsP1-4 유전자 및 적어도 하나의 핵산 서열은 분리된 열린 해독틀에 있는, 조성물.

청구항 45

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 44항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 RNA 알파바이러스 백본에 의해 코딩되는 천연 26S 프로모터 뉴클레오티드 서열인, 조성물.

청구항 46

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 44항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 외인성 RNA 프로모터인, 조성물.

청구항 47

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 46항 중 어느 한 항에 있어서, 상기 제 2 프로모터 뉴클레오티드 서열이 26S 프로모터 뉴클레오티드 서열인, 조성물.

청구항 48

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 46항 중 어느 한 항에 있어서, 상기 제 2 프로모터 뉴클레오티드 서열이 다수의 26S 프로모터 뉴클레오티드 서열을 포함하고, 여기서 각각의 26S 프로모터 뉴클레오티드 서열이 하나 이상의 분리된 열린 해독틀의 전사를 제공하는, 조성물.

청구항 49

제 1항 내지 제 48항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 적어도 300 nt인, 조성물.

청구항 50

제 1항 내지 제 49항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 적어도 1 kb인, 조성물.

청구항 51

제 1항 내지 제 50항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터의 크기가 각각 2 kb인, 조성물.

청구항 52

제 1항 내지 제 51항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 5 kb 미만인, 조성물.

청구항 53

제 1항 내지 제 19항 또는 제 21항 내지 제 52항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함하는, 조성물.

청구항 54

제 53항에 있어서, 각각의 항원-코딩 핵산 서열이 서로 직접 연결된, 조성물.

청구항 55

제 1항 내지 제 19항 또는 제 21항 내지 제 54항 중 어느 한 항에 있어서, 각각의 항원-코딩 핵산 서열이 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된, 조성물.

청구항 56

제 55항에 있어서, 상기 링커가 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결하는, 조성물.

청구항 57

제 56항에 있어서, 상기 링커가 (1) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 글리신 잔기; (2) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원을 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택되는, 조성물.

청구항 58

제 55항에 있어서, 상기 링커가 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결하는, 조성물.

청구항 59

제 58항에 있어서, 상기 링커가 서열 GPGPG를 포함하는, 조성물.

청구항 60

제 1항 내지 제 19항 또는 제 21항 내지 제 59항 중 어느 한 항에 있어서, 상기 항원-코딩 핵산 서열이 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결된, 조성물.

청구항 61

제 60항에 있어서, 상기 분리 또는 인접 서열이 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열이 위치 76에서 Gly에서 Ala로의 치환을 함유), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열 중 적어도 하나를 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76인, 조성물.

청구항 62

제 1항 내지 제 19항 또는 제 21항 내지 제 61항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각

각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 조성물.

청구항 63

제 1항 내지 제 19항 또는 제 21항 내지 제 61항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 조성물.

청구항 64

제 1항 내지 제 19항 또는 제 21항 내지 제 61항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하는, 조성물.

청구항 65

제 1항 내지 제 19항 또는 제 21항 내지 제 61항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩하는, 조성물.

청구항 66

제 20항에 있어서, 적어도 2개의 MHC 클래스 I 에피토프가 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시되는, 조성물.

청구항 67

제 1항 내지 제 19항 또는 제 21항 내지 제 66항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열이 8 내지 35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개 아미노산 길이의 폴리펩티드 서열을 코딩하는, 조성물

청구항 68

제 17항 내지 제 19항 또는 제 21항 내지 제 67항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하는, 조성물.

청구항 69

제 17항 내지 제 19항 또는 제 21항 내지 제 67항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드 서열과 구별되게 하는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 70

제 1항 내지 제 19항 또는 제 21항 내지 제 69항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열이 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 20-40개 아미노산 길이의 폴리펩티드 서열을 코딩하는, 조성물.

청구항 71

제 1항 내지 제 19항, 또는 제 21항 내지 제 70항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함하는, 조성물.

청구항 72

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 71항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열이 유도성인, 조성물.

청구항 73

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 71항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열이 비유도성인, 조성물.

청구항 74

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 73항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 알파바이러스에 천연인 폴리(A) 서열을 포함하는, 조성물.

청구항 75

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 73항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 알파바이러스에 외인성인 폴리(A) 서열을 포함하는, 조성물.

청구항 76

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 75항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 하나의 핵산 서열 중 적어도 하나에 작동가능하게 연결되는, 조성물.

청구항 77

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 76항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드인, 조성물.

청구항 78

제 1항, 제 3항 내지 제 19항 또는 제 21항 내지 제 76항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 100개의 연속적인 A 뉴클레오티드인, 조성물.

청구항 79

제 1항 내지 제 19항 또는 제 21항 내지 제 78항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 조성물:

(a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 80

제 20항에 있어서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 조성물:

(a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트

각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 81

제 79항에 있어서, 선택된 에피토프 세트의 수가 2-20인, 조성물.

청구항 82

제 79항 내지 제 81항 중 어느 한 항에 있어서, 상기 제시 모델은

(a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및

(b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열의 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타내는, 조성물.

청구항 83

제 79항 내지 제 82항 중 어느 한 항에 있어서, 선택된 에피토프 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 84

제 79항 내지 제 83항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 85

제 79항 내지 제 84항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브(naïve) T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하고, 임의로 여기서 APC는 수지상 세포(DC)인, 조성물.

청구항 86

제 79항 내지 제 85항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 87

제 79항 내지 제 86항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 88

제 79항 내지 제 87항 중 어느 한 항에 있어서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터가 감염된 조직에 대해 시퀀싱을 수행함으로써 수득되는, 조성물.

청구항 89

제 88항에 있어서, 상기 시퀀싱이 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법인, 조성물.

청구항 90

침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물로서,여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은,

ChAdV-기반 발현 시스템을 포함하며, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는,

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은:

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하며,

여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결되는, 조성물.

청구항 91

ChAdV-기반 발현 시스템의 전달을 위한 조성물로서, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은,

ChAdV-기반 발현 시스템을 포함하며, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는,

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은,

(i) 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열로서, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있는, 변형된 ChAdV68 서열;

(ii) CMV 프로모터 뉴클레오티드 서열; 및

(iii) SV40 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아 유래-펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는, 카세트를 포함하며, 여기서 카세트는 E1 결실 내에 삽입되고 카세트는 CMV 프로모터 뉴클레오티드 서열 및 SV40 폴리(A) 서열에 작동가능하게 연결되는, 조성물.

청구항 92

제 90항 또는 제 91항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 93

제 90항 또는 제 91항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 1-10 사이, 1-20 사이, 1-30 사이, 1-40

사이, 1-50 사이, 1-100 사이, 1-200 사이, 1-300 사이, 1-400개 사이, 또는 1-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 펩티드를 코딩하는 핵산 서열은 2-10 사이, 2-20 사이, 2-30 사이, 2-40 사이, 2-50 사이, 2-100 사이, 2-200 사이, 2-300 사이, 2-400 사이, 또는 2-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 94

제 90항 또는 제 91항에 있어서, 상기 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인으로부터 선택된 펩티드를 코딩하는 핵산 서열을 포함하고, 임의로 여기서 감염성 질병 유기체는 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인 및 이들의 조합으로부터 선택된 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함하는, 조성물.

청구항 95

제 90항 내지 제 94항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 면역 반응을 자극할 수 있고, 임의로 여기서 면역 반응이 T 세포 반응 및/또는 B 세포 반응인, 조성물 .

청구항 96

제 90항 내지 제 95항 중 어느 한 항에 있어서, 상기 감염성 질병 유기체가 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 조성물.

청구항 97

제 90항 내지 제 96항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 세포의 표면 상에서 MHC 클래스 I에 의해 제시되는 것으로 알려지거나 의심되는 에피토프를 코딩하고, 임의로 여기서 세포의 표면이 감염된 세포 표면이고, 임의로 여기서 세포는 대상체의 세포인, 조성물.

청구항 98

제 97항에 있어서, 상기 세포가 병원체 감염 세포, 바이러스 감염 세포, 박테리아 감염 세포, 진균 감염 세포, 및 기생충 감염 세포로 이루어진 군으로부터 선택된 감염된 세포인, 조성물.

청구항 99

제 98항에 있어서, 상기 바이러스로 감염된 세포가 HIV 감염 세포, HPV 감염 세포, SARS 감염 세포, SARS-CoV-2 감염 세포, 에볼라 감염 세포, HBV 감염 세포, 인플루엔자 감염 세포, HCV 감염 세포, CMV 감염 세포, 치쿤구니야 바이러스 감염 세포, RSV 감염 세포, 뎅기열 바이러스 감염 세포, 오르티믹소바이러스과 바이러스 감염 세포 및 결핵 감염 세포로 이루어진 군으로부터 선택되는, 조성물.

청구항 100

제 90항 또는 제 92항 내지 제 99항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'부터 3'로 하기를 포함하는 식으로 설명되며,

P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

여기서, P는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 적어도 하나의 프로모터 서열을 포함하고, 여기서 a = 1이고,

N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고,

L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고,

L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고,

G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고,

G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고,

U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고,

X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고, 그리고

Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열인, 조성물.

청구항 101

제 100항에 있어서, 각각의 X에 대해 상응하는 N_c가 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열인, 조성물.

청구항 102

제 100항 또는 제 101항에 있어서, 각각의 Y에 대해 상응하는 U_f가 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열인, 조성물.

청구항 103

제 100항 내지 제 102항 중 어느 한 항에 있어서,

b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2이고,

P는 CMV 프로모터 서열이고,

각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고,

L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

U는 각각 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고,

ChAdV 백터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고, 그리고 항원 카세트는 E1 결실 내에 삽입되고, 그리고

각각의 MHC 클래스 I 항원-코딩 핵산 서열은 길이가 25개 아미노산인 폴리펩티드를 코딩하는, 조성물.

청구항 104

제 90항 내지 제 103항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여용으로 제형화되는, 조성물.

청구항 105

제 90항 내지 제 104항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물은 근육내(IM) 투여용으로 제형화되는, 조성물.

청구항 106

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 105항 중 어느 한 항에 있어서, 상기 카세트가 적어도

하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합된, 조성물.

청구항 107

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 106항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 상기 카세트에 작동가능하게 연결되는, 조성물.

청구항 108

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 107항 중 어느 한 항에 있어서, 상기 ChAdV 백본이 ChAdV68 백터 백본을 포함하는, 조성물.

청구항 109

제 108항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호:1에 제시된 서열을 포함하는, 조성물.

청구항 110

제 108항에 있어서, 상기 ChAdV68 벡터 백본은 ChAdV68 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 아데노바이러스 E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택되는 적어도 하나의 유전자에 기능적 결실을 포함하고, 임의로 여기서 아데노바이러스 백본 또는 변형된 ChAdV68 서열은 아데노바이러스 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 (1) E1A 및 E1B; 또는 (2) E1A, E1B 및 E3에서 완전히 결실되거나 기능적으로 결실되고, 임의로 여기서 E1 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 577 내지 3403의 E1 결실을 통해 기능적으로 결실되고 및 임의로 여기서 E3 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 27,125 내지 31,825의 E3 결실을 통해 기능적으로 결실되는, 조성물.

청구항 111

제 108항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호: 1에 제시된 서열의 염기쌍 번호 577과 3403 사이 또는 염기쌍 456과 3014 사이에 하나 이상의 결실을 포함하고, 임의로 여기서 벡터가 염기쌍 27,125와 31,825 사이 또는 염기쌍 27,816과 31,333 사이에 하나 이상의 결실을 추가로 포함하는, 조성물.

청구항 112

제 108항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518은:

A. E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403;

B. E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825;

C. 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642;

D. 서열번호: 1에 제시된 서열의 뉴클레오티드 456 내지 3014;

E. 서열번호: 1에 제시된 서열의 뉴클레오티드 27,816 내지 31,333;

F. 서열번호: 1에 제시된 서열의 뉴클레오티드 3957 내지 10346;

G. 서열번호: 1에 제시된 서열의 뉴클레오티드 21787 내지 23370;

H. 서열번호: 1에 제시된 서열의 뉴클레오티드 33486 내지 36193; 또는

이들의 조합이 결여된, 조성물.

청구항 113

제 108항 내지 제 112항 중 어느 한 항에 있어서, 상기 ChAdV68 벡터 백본이 부분적으로 결실된 E4 유전자를 포함하고, 임의로 여기서 부분적으로 결실된 E4 유전자는:

A. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642가 결여된 E4

유전자 서열,

B. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 34,942, 뉴클레오티드 34,952 내지 35,305, 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642가 결여된 E4 유전자 서열,

C. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516이 결여된 E4 유전자 서열,

D. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,979 내지 35,642가 결여된 E4 유전자 서열,

E. E4Orf2의 적어도 부분 결실, 완전히 결실된 E4Orf3 및 E4Orf4의 적어도 부분 결실의 E4 결실,

F. E4Orf2의 적어도 부분 결실, E4Orf3의 적어도 부분 결실, 및 E4Orf4의 적어도 부분 결실의 E4 결실,

G. E4Orf1의 적어도 부분 결실, 완전히 결실된 E4Orf2, 및 E4Orf3의 적어도 부분 결실의 E4 결실, 또는

H. E4Orf2의 적어도 부분 결실 및 E4Orf3의 적어도 부분 결실의 E4 결실을 포함하는, 조성물.

청구항 114

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 113항 중 어느 한 항에 있어서, 상기 카세트가 E1 영역, E3 영역 및/또는 카세트의 혼입을 허용하는 임의의 결실된 AdV 영역에서 상기 ChAdV 백본에 삽입되는, 조성물.

청구항 115

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 114항 중 어느 한 항에 있어서, 상기 ChAdV 백본이 1세대, 2세대, 또는 헬퍼-의존성 아데노바이러스 벡터 중 하나로부터 생성되는, 조성물.

청구항 116

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 115항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 CMV, SV40, EF-1, RSV, PGK, HSA, MCK 및 EBV 프로모터 서열로 이루어진 군으로부터 선택되는, 조성물.

청구항 117

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 115항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 CMV 프로모터 서열인, 조성물.

청구항 118

제 90항 내지 제 117항 중 어느 한 항에 있어서, 적어도 하나의 에피토프-코딩 핵산 서열이 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 I에 의해 제시될 수 있는 에피토프를 코딩하는, 조성물.

청구항 119

제 90항 내지 제 118항 중 어느 한 항에 있어서, 적어도 하나의 에피토프-코딩 핵산 서열이 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 II에 의해 제시될 수 있는 에피토프를 코딩하는, 조성물.

청구항 120

제 90항 내지 제 102항 또는 제 104항 내지 제 119항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 2개 이상의 항원-코딩 핵산 서열을 포함하는, 조성물.

청구항 121

제 120항에 있어서, 각각의 항원-코딩 핵산 서열이 서로 직접 연결된, 조성물.

청구항 122

제 90항 내지 제 102항 또는 제 104항 내지 제 121항 중 어느 한 항에 있어서, 각각의 항원-코딩 핵산 서열이 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결되는, 조성물.

청구항 123

제 122항에 있어서, 상기 링커가 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결하는, 조성물.

청구항 124

제 123항에 있어서, 상기 링커가 (1) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 글리신 잔기; (2) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원을 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택되는, 조성물.

청구항 125

제 122항에 있어서, 상기 링커가 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결하는, 조성물.

청구항 126

제 125항에 있어서, 상기 링커가 서열 GPGPG를 포함하는, 조성물.

청구항 127

제 90항 내지 제 102항 또는 제 104항 내지 제 126항 중 어느 한 항에 있어서, 상기 항원-코딩 핵산 서열이 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결되는, 조성물.

청구항 128

제 127항에 있어서, 상기 분리 또는 인접 서열이 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열이 위치 76에서 Gly에서 Ala로의 치환을 함유), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열 중 적어도 하나를 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76인, 조성물.

청구항 129

제 90항 내지 제 128항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 친화도를 갖도록 하는 적어도 하나의 변경을 포함하는, 조성물.

청구항 130

제 90항 내지 제 129항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 안정성을 갖도록 하는 적어도 하나의 변경을 포함하는, 조성물.

청구항 131

제 90항 내지 제 130항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자 상에서 증가된 제시 가능성을 갖도록 하는 적어도 하나의 변경을 포함하는, 조성물.

청구항 132

제 90항 내지 제 131항 중 어느 한 항에 있어서, 상기 적어도 하나의 변경이 점 돌연변이, 프레임시프트 돌연변이, 비-프레임시프트 돌연변이, 결실 돌연변이, 삽입 돌연변이, 스플라이스 변이체, 게놈 재배열 또는 프로테아좀-생성 스플라이싱된 항원을 포함하는, 조성물.

청구항 133

제 90항 내지 제 132항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 감염을 갖는 것으로 알려지거나 의심되는 대상체에서 발현되는 것으로 알려지거나 의심되는 에피토프를 코딩하는, 조성물.

청구항 134

제 133항에 있어서, 상기 감염이 병원체 감염, 바이러스 감염, 박테리아 감염, 진균 감염 및 기생충 감염으로 이루어진 군으로부터 선택되는, 조성물.

청구항 135

제 134항에 있어서, 상기 바이러스 감염이 HIV 감염, HPV 감염, SARS 감염, SARS-CoV-2 감염, 에볼라 감염, HBV 감염, 인플루엔자 감염, HCV 감염, CMV 감염, 치쿤구니야 바이러스 감염, RSV 감염, 뎅기열 바이러스 감염, 오르티믹소바이러스과 바이러스 감염, 및 결핵 감염으로 이루어진 군으로부터 선택되는, 조성물.

청구항 136

제 134항에 있어서, 상기 박테리아 감염이 결핵 감염인, 조성물.

청구항 137

제 90항 내지 제 102항 또는 제 104항 내지 제 136항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 조성물.

청구항 138

제 90항 내지 제 102항 또는 제 104항 내지 제 136항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 조성물.

청구항 139

제90항 내지 제 102항 또는 제 104항 내지 제 136항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하는, 조성물.

청구항 140

제 90항 내지 제 102항 또는 제 104항 내지 제 136항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩하는, 조성물.

청구항 141

제 103에 있어서, 적어도 2개의 MHC 클래스 I 에피토프가 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시되는, 조성물.

청구항 142

제 90항 내지 제 102항 또는 제 104항 내지 제 141항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8 내지 35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,

23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 조성물.

청구항 143

제 100항 내지 제 102항 또는 제 104항 내지 제 142항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하는, 조성물.

청구항 144

제 100항 내지 제 102항 또는 제 104항 내지 제 142항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드 서열과 구별되게 하는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하는, 조성물.

청구항 145

제 90항 내지 제 102항 또는 제 104항 내지 제 144항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 조성물.

청구항 146

제 90항 내지 제 102항 또는 제 104항 내지 제 145항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함하는, 조성물.

청구항 147

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 146항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 유도성인, 조성물.

청구항 148

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 146항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 비-유도성인, 조성물.

청구항 149

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 148항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 소 성장 호르몬(BGH) SV40 폴리A 서열을 포함하는, 조성물.

청구항 150

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 149항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드인, 조성물.

청구항 151

제 90항, 제 92항 내지 제 102항 또는 제 104항 내지 제 149항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 100개의 연속적인 A 뉴클레오티드인, 조성물.

청구항 152

제 90항 내지 제 151항 중 어느 한 항에 있어서, 상기 카세트가 인트론 서열, 우드척 간염 바이러스 전사후 조절 요소(WPRE) 서열, 내부 리보솜 진입 서열(IRES) 서열, 2A 자가 절단 펩티드 서열을 코딩하는 뉴클레오티드 서열, 푸린 절단 부위를 코딩하는 뉴클레오티드 서열, 또는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나

에 작동가능하게 연결된 mRNA의 핵 엑스포트, 안정성 또는 번역 효율을 향상시키는 것으로 알려진 5' 또는 3' 비-코딩 영역의 서열 중 적어도 하나를 추가로 포함하는, 조성물.

청구항 153

제 90항 내지 제 152항 중 어느 한 항에 있어서, 상기 카세트가 녹색 형광 단백질(GFP), GFP 변이체, 분비된 알칼리성 포스파타제, 루시퍼라제, 루시페라제 변이체, 또는 검출가능한 펩티드 또는 에피토프를 포함하나 이에 제한되지 않는 리포터 유전자를 추가로 포함하는, 조성물.

청구항 154

제 153항에 있어서, 상기 검출가능한 펩티드 또는 에피토프가 HA 태그, 플래그(Flag) 태그, His-태그, 또는 V5 태그로 이루어진 군으로부터 선택되는, 조성물.

청구항 155

제 90항 내지 제 154항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 적어도 하나의 면역 조절자를 코딩하는 하나 이상의 핵산 서열을 추가로 포함하는, 조성물.

청구항 156

제 155항에 있어서, 상기 면역 조절자가 항-CTLA4 항체 또는 이의 항원-결합 단편, 항-PD-1 항체 또는 이의 항원-결합 단편, 항-PD-L1 항체 또는 이의 항원-결합 단편, 항-4-1BB 항체 또는 이의 항원-결합 단편, 또는 항-OX-40 항체 또는 이의 항원-결합 단편인, 조성물.

청구항 157

제 156항에 있어서, 상기 항체 또는 이의 항원-결합 단편이 단일 특이성 또는 다중 특이성으로 함께 연결된 Fab 단편, Fab' 단편, 단일 사슬 Fv(scFv), 단일 도메인 항체(sdAb)(예를 들어, 낙타류 항체 도메인), 또는 전장 단일-사슬 항체(예를 들어, 가요성 링커에 의해 연결된 중쇄 및 경쇄가 있는 전장 IgG)인, 조성물.

청구항 158

제 156항 또는 제 157항에 있어서, 상기 항체의 중쇄 및 경쇄 서열이 2A 또는 IRES와 같은 자가-절단 서열에 의해 분리된 연속 서열이거나; 또는 상기 항체의 중쇄 및 경쇄 서열은 연속적인 글리신 잔기와 같은 가요성 링커에 의해 연결되는, 조성물.

청구항 159

제 155항에 있어서, 상기 면역 조절자가 사이토카인인, 조성물.

청구항 160

제 159항에 있어서, 상기 사이토카인이 IL-2, IL-7, IL-12, IL-15, 또는 IL-21 또는 이들 각각의 변이체 중 적어도 하나인, 조성물.

청구항 161

제 90항 내지 제 102항 또는 제 104항 내지 제 160항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 조성물:

(a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 162

제 103항에 있어서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 조성물:

(a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 163

제 161항에 있어서, 선택된 에피토프 세트의 수가 2-20인, 조성물.

청구항 164

제 161항 내지 제 163항 중 어느 한 항에 있어서, 상기 제시 모델은

(a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및

(b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열의 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면에 제시될 가능성 사이의 의존성을 나타내는, 조성물.

청구항 165

제 161항 내지 제 164항 중 어느 한 항에 있어서, 선택된 에피토프 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 166

제 161항 내지 제 165항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 167

제 161항 내지 제 166항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하고, 임의로 여기서 APC는 수지상 세포(DC)인, 조성물.

청구항 168

제 161항 내지 제 167항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 대상체의 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 169

제 161항 내지 제 168항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소

된 에피토프를 선택하는 것을 포함하는, 조성물.

청구항 170

제 161항 내지 제 169항 중 어느 한 항에 있어서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터가 감염된 조직에 대해 시퀀싱을 수행함으로써 수득되는, 조성물.

청구항 171

제 170항에 있어서, 상기 시퀀싱이 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법인, 조성물.

청구항 172

제 90항 내지 제 171항 중 어느 한 항에 있어서, 상기 카세트가 카세트 내의 인접 서열에 의해 형성된 접합 에피토프 서열을 포함하는, 조성물.

청구항 173

제 172항에 있어서, 적어도 하나 또는 각각의 접합 에피토프 서열이 MHC에 대해 500 nM 초과의 친화도를 갖는, 조성물.

청구항 174

제 172항 또는 제 173항에 있어서, 각각의 접합 에피토프 서열이 비-자기인, 조성물.

청구항 175

제 90항 내지 제 174항 중 어느 한 항에 있어서, 상기 카세트가 번역된 야생형 핵산 서열을 포함하는 비-치료적 MHC 클래스 I 또는 클래스 II 에피토프 핵산 서열을 코딩하지 않고, 여기서 비-치료적 에피토프가 대상체의 MHC 대립유전자 상에 표시될 것으로 예측되는, 조성물.

청구항 176

제 175항에 있어서, 상기 비-치료적 예측된 MHC 클래스 I 또는 클래스 II 에피토프 서열이 카세트에서 인접 서열에 의해 형성된 접합 에피토프 서열인, 조성물.

청구항 177

제 172항 내지 제 176항 중 어느 한 항에 있어서, 상기 예측이 비-치료적 에피토프의 서열을 제시 모델에 입력함으로써 생성된 제시 가능성에 기초하는, 조성물.

청구항 178

제172항 내지 제177항 중 어느 한 항에 있어서, 카세트 내의 항원-코딩 핵산 서열의 순서가 다음을 포함하는 일련의 단계에 의해 결정되는, 조성물:

(a) 항원-코딩 핵산 서열의 상이한 순서에 상응하는 후보 카세트 서열 세트를 생성하는 단계;

(b) 각각의 후보 카세트 서열에 대해 후보 카세트 서열에서 비-치료적 에피토프의 제시에 기초하여 제시 점수를 결정하는 단계; 및

(c) 백신에 대한 카세트 서열로서 미리 결정된 임계값 미만의 제시 점수와 관련된 후보 카세트 서열을 선택하는 단계.

청구항 179

제 90항 내지 제 178항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물이 약제학적으로 허용되는 담체를 포함하는 약제학적 조성물로 제형화되는, 조성물.

청구항 180

상기 조성물 청구항 중 어느 한 항의 ChAdV-기반 발현 시스템의 전달을 위한 조성물, 및 사용 설명서를 포함하

는 키트.

청구항 181

제 90항 내지 제 180항 중 어느 한 항에 있어서, 하나 이상의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래된, 조성물.

청구항 182

제 90항 내지 제 181항 중 어느 한 항에 있어서, 각각의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래된, 조성물.

청구항 183

제 90항 내지 제 182항 중 어느 한 항에 있어서, 하나 이상의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되지 않는, 조성물.

청구항 184

제 90항 내지 제 183항 중 어느 한 항에 있어서, 각각의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되지 않는, 조성물.

청구항 185

자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물로서,

여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은:

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, RNA 알파바이러스 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) 다음을 포함하는 적어도 하나의 항원-코딩 핵산 서열:

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열로서, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 핵산 서열;

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열;

(ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제2 프로모터 뉴클레오티드 서열; 및

(iii) 알파바이러스에 대해 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열인, 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하는, 카세트를 포함하는, 자가-복제 알파바이러스-기반 발현 시스템; 및

(B) 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 조성물.

청구항 186

침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물로서,여기서 ChAdV-기반 발현 시스템의

전달을 위한 조성물은,

ChAdV-기반 발현 시스템을 포함하며, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는,

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은:

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열로서, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 핵산 서열;

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하며, 여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결되는, 조성물.

청구항 187

제 185항 또는 제 186항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템 또는 ChAdV-기반 발현 시스템이 제 1항 내지 제 186항 중 어느 한 항의 특징 중 임의의 것을 포함하는, 조성물.

청구항 188

대상체에서 면역 반응을 자극하는 방법으로서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하고/하거나 침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 단계를 포함하고,

a. ChAdV-기반 발현 시스템을 포함하는 ChAdV-기반 발현 시스템의 전달을 위한 조성물,

b. 자가-복제 알파바이러스-기반 발현 시스템을 포함하는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물, 또는

c. ChAdV-기반 발현 시스템을 포함하는 ChAdV-기반 발현 시스템의 전달을 위한 조성물 및 자가-복제 알파바이러스-기반 발현 시스템을 포함하는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물 중 어느 하나인, 방법.

청구항 189

제 188항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물이 프라이밍 용량으로 투여되고, ChAdV-기반 발현 시스템의 전달을 위한 조성물 또는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물 중 어느 하나는 하나 이상의 부스팅 용량으로 투여되는, 방법.

청구항 190

제 188항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 상기 조성물이 프라이밍 용량으로 투여되고, ChAdV-기반 발현 시스템의 전달을 위한 조성물 또는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물중 어느 하나는 하나 이상의 부스팅 용량으로 투여되는, 방법.

청구항 191

제 189항 내지 제 190항 중 어느 한 항에 있어서, 2회 이상의 부스팅 용량이 투여되는, 방법.

청구항 192

제 189항 내지 제 190항 중 어느 한 항에 있어서, 1, 2, 3, 4, 5, 6, 7, 또는 8회의 부스팅 용량이 투여되는, 방법.

청구항 193

제 188항 내지 제 192항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물이 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여되는, 방법.

청구항 194

제 188항 내지 제 193항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물이 투여(IM)되는, 방법.

청구항 195

제 194항에 있어서, 상기 IM 투여가 분리된 주사 부위에서 투여되는, 방법.

청구항 196

제 195항에 있어서, 상기 분리된 주사 부위가 대향 삼각근에 있는, 방법.

청구항 197

제 195항에 있어서, 상기 분리된 주사 부위가 각 측면의 둔근 또는 대퇴직근 부위에 있는, 방법.

청구항 198

제 188항 내지 제 197항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 상기 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여되는, 방법.

청구항 199

제 188항 내지 제 198항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 상기 조성물이 투여(IM)되는, 방법.

청구항 200

제 199항에 있어서, 상기 IM 투여가 분리된 주사 부위에서 투여되는, 방법.

청구항 201

제 200항에 있어서, 상기 분리된 주사 부위가 대향 삼각근에 있는, 방법.

청구항 202

제 200항에 있어서, 상기 분리된 주사 부위가 각 측면의 둔근 또는 대퇴직근 부위에 있는, 방법.

청구항 203

제 199항 내지 제 202항 중 어느 한 항에 있어서, 하나 이상의 부스팅 용량의 상기 주사 부위가 프라이밍 용량의 주사 부위에 가능한 한 가까운, 방법.

청구항 204

제 188항 내지 제 203항 중 어느 한 항에 있어서, 대상체의 HLA-일배체형을 결정하거나 결정한 것을 추가로 포함하는, 방법.

청구항 205

제 188항 내지 제 204항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조

성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은,

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, RNA 알파바이러스 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) 다음을 포함하는 적어도 하나의 항원-코딩 핵산 서열:

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열;

(ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제2 프로모터 뉴클레오티드 서열; 및

(iii) 알파바이러스에 대해 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열인, 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하는 카세트를 포함하는 자가-복제 알파바이러스-기반 발현 시스템, 및

(B) 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 방법.

청구항 206

제 188항 내지 제 205항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은 서열번호: 6에 제시된 핵산 서열을 포함하고, 여기서 RNA 알파바이러스 백본 서열은 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열을 포함하고, 여기서 26S 프로모터 서열은 RNA 알파바이러스 백본에 내인성이고, 여기서 폴리(A) 서열은 RNA 알파바이러스 백본에 내인성인 RNA 알파바이러스 백본, 및

(b) 26S 프로모터 뉴클레오티드 서열에 작동가능하게 연결되는, 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합된 카세트로서,

여기서 카세트는,

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하는 자가-복제 알파바이러스-기반 발현 시스템; 및

(B) 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 방법.

청구항 207

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 208

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 209

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 210

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 1-10 사이, 1-20 사이, 1-30 사이, 1-40 사이, 1-50 사이, 1-100 사이, 1-200 사이, 1-300 사이, 1-400 사이, 또는 1-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 211

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 2-10 사이, 2-20 사이, 2-30 사이, 2-40 사이, 2-50 사이, 2-100 사이, 2-200 사이, 2-300 사이, 2-400 사이, 또는 2-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 212

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 전장 단백질을 코딩하는 핵산 서열을 포함하는, 방법.

청구항 213

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 단백질 도메인을 코딩하는 핵산 서열을 포함하는, 방법.

청구항 214

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 단백질 서브유닛을 코딩하는 핵산 서열을 포함하는 방법.

청구항 215

제 188항 내지 제 206항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인, 및 이들의 조합으로부터 선택된 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함하는, 방법.

청구항 216

제 188항 내지 제 215항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 면역 반응을 자극할 수 있는, 방법.

청구항 217

제 188항 내지 제 215항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 T 세포 반응을 자극할 수 있는, 방법.

청구항 218

제 188항 내지 제 215항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 B 세포 반응을 자극할 수 있는, 방법.

청구항 219

제 188항 내지 제 215항 중 어느 한 항에 있어서, 상기 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 T 세포 반응 및 B 세포 반응을 자극할 수 있는, 방법.

청구항 220

제 188항 내지 제 219항 중 어느 한 항에 있어서, 상기 감염성 질병 유기체가 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 방법.

청구항 221

상기 방법 청구항 또는 제 207항 내지 제 220항 중 어느 한 항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'부터 3'로 하기를 포함하는 식으로 설명되며,

P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

여기서, P는 제 2 프로모터 뉴클레오티드 서열을 포함하고, 여기서 a = 0 또는 1이고,

N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고,

L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고,

L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고,

G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고,

G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고,

U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고,

X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고, 및

Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열인, 방법.

청구항 222

제 221항에 있어서, 각각의 X에 대해 상응하는 N_c가 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열인, 방법.

청구항 223

제 221항 또는 제 222항에 있어서, 각각의 Y에 대해 상응하는 U_f가 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열인, 방법.

청구항 224

제 221항 내지 제 223항 중 어느 한 항에 있어서,

a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2이고,

적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 제공되는 단일 26S 프로모터 뉴클레오티드 서열이고,

적어도 하나의 폴리아데닐화 폴리(A) 서열은 RNA 알파바이러스 백본에 의해 제공되는 적어도 100개의 연속적인 A 뉴클레오티드의 폴리(A) 서열이고,

카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합되고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열에 작동가능하게 연결되고,

각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고,

L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

U는 각각 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고,

RNA 알파바이러스 백본은 서열번호: 6에 제시된 서열이고, 그리고

각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 길이가 13개 내지 25개 아미노산인 폴리펩티드를 코딩하는, 방법.

청구항 225

제 188항 내지 제 224항 중 어느 한 항에 있어서, 상기 LNP가 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, PEG-기반 코트 지질, 또는 이들의 조합으로 이루어진 군으로부터 선택된 지질을 포함하는, 방법.

청구항 226

제 188항 내지 제 225항 중 어느 한 항에 있어서, 상기 LNP가 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, 및 PEG-기반 코트 지질을 포함하는, 방법.

청구항 227

제 225항 또는 제 226항에 있어서, 상기 이온화가능한 아미노 지질이 MC3-유사(디리놀레일메틸-4-디메틸아미노부티레이트) 분자를 포함하는, 방법.

청구항 228

제 188항 내지 제 227항 중 어느 한 항에 있어서, 상기 LNP-캡슐화 발현 시스템이 약 100 nm의 직경을 갖는, 방법.

청구항 229

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 228항 중 어느 한 항에 있어서, 상기 카세트가 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합되는, 방법.

청구항 230

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 229항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 상기 카세트에 작동가능하게 연결되는, 방법.

청구항 231

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 230항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 하나 이상의 +-가닥 RNA 벡터를 포함하는, 방법.

청구항 232

제 231항에 있어서, 상기 하나 이상의 +-가닥 RNA 벡터가 5'7-메틸구아노신(m7g) 캡을 포함하는, 방법.

청구항 233

제 231항 또는 제 232항에 있어서, 상기 하나 이상의 +-가닥 RNA 벡터가 시험관내 전사에 의해 생성되는, 방법.

청구항 234

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 233항 중 어느 한 항에 있어서, 상기 하나 이상의 벡

터가 포유동물 세포 내에서 자가-복제되는, 방법.

청구항 235

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 234항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함하는, 방법.

청구항 236

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 234항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 베네수엘라 말 뇌염 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함하는, 방법.

청구항 237

제 235항 또는 제 236항에 있어서, 상기 RNA 알파바이러스 백본이 적어도 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩되는 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열, 비구조적 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자 및 nsP4 유전자를 적어도 포함하는, 방법.

청구항 238

제 235항 또는 제 236항에 있어서, 상기 RNA 알파바이러스 백본이 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩되는 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 및 폴리(A) 서열을 적어도 포함하는, 방법.

청구항 239

제 237항 또는 제 238항에 있어서, 비구조적 단백질-매개 증폭을 위한 서열이 알파바이러스 5' UTR, 51-nt CSE, 24-nt CSE, 26S 서브게놈 프로모터 서열, 19-nt CSE, 알파바이러스 3' UTR, 또는 이들의 조합으로 이루어진 군으로부터 선택되는, 방법.

청구항 240

제 237항 내지 제 239항 중 어느 한 항에 있어서, 상기 RNA 알파바이러스 백본이 구조적 비리온 단백질 캡시드, E2 및 E1을 코딩하지 않는, 방법.

청구항 241

제 240항에 있어서, 상기 카세트가 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스 또는 마야로 바이러스의 뉴클레오티드 서열 내 구조적 비리온 단백질 대신 삽입되는, 방법.

청구항 242

제 235항 또는 제 236항에 있어서, 상기 베네수엘라 말 뇌염 바이러스가 서열번호: 3 또는 서열번호: 5의 서열을 포함하는, 방법.

청구항 243

제 235항 또는 제 236항에 있어서, 상기 베네수엘라 말 뇌염 바이러스가 염기쌍 7544와 11175 사이의 결실을 추가로 포함하는 서열번호: 3 또는 서열번호: 5의 서열을 포함하는, 방법.

청구항 244

제 243항에 있어서, 상기 RNA 알파바이러스 백본이 서열번호: 6 또는 서열번호: 7에 제시된 서열을 포함하는,

방법.

청구항 245

제 243항 또는 제 244항에 있어서, 상기 카세트가 서열번호: 3 또는 서열번호: 5의 서열에 제시된 바와 같이 염기쌍 7544 및 11175 사이의 결실을 대체하기 위해 위치 7544에 삽입되는, 방법.

청구항 246

제 241항 내지 제 245항 중 어느 한 항에 있어서, 상기 카세트의 삽입이 nsP1-4 유전자 및 적어도 하나의 핵산 서열을 포함하는 폴리시스트론 RNA의 전사를 제공하고, 여기서 상기 nsP1-4 유전자 및 상기 적어도 하나의 핵산 서열은 분리된 열린 해독틀에 있는, 방법.

청구항 247

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 246항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 RNA 알파바이러스 백본에 의해 코딩되는 천연 26S 프로모터 뉴클레오티드 서열인, 방법.

청구항 248

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 246항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 외인성 RNA 프로모터인, 방법.

청구항 249

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 248항 중 어느 한 항에 있어서, 상기 제 2 프로모터 뉴클레오티드 서열이 26S 프로모터 뉴클레오티드 서열인, 방법.

청구항 250

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 248항 중 어느 한 항에 있어서, 상기 제 2 프로모터 뉴클레오티드 서열이 다수의 26S 프로모터 뉴클레오티드 서열을 포함하고, 여기서 각각의 26S 프로모터 뉴클레오티드 서열이 하나 이상의 분리된 열린 해독틀의 전사를 제공하는, 방법.

청구항 251

제 188항 내지 제 250항 중 어느 하나 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 적어도 300 nt인, 방법.

청구항 252

제 188항 내지 제 251항 중 어느 하나 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 적어도 1 kb인, 방법.

청구항 253

제 188항 내지 제 252항 중 어느 하나 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 2 kb인, 방법.

청구항 254

제 188항 내지 제 253항 중 어느 하나 한 항에 있어서, 상기 하나 이상의 벡터가 각각 크기가 5 kb 미만인, 방법.

청구항 255

제 205항 내지 제 223항 또는 제 225항 내지 제 254항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함하는, 방법.

청구항 256

제 255항에 있어서, 각각의 항원-코딩 핵산 서열이 서로 직접 연결된, 방법.

청구항 257

제 205항 내지 제 223항 또는 제 225항 내지 제 256항 중 어느 한 항에 있어서, 각각의 항원-코딩 핵산 서열이 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된, 방법.

청구항 258

제 257항에 있어서, 상기 링커가 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결하는, 방법.

청구항 259

제 258항에 있어서, 상기 링커가 (1) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 글리신 잔기; (2) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원을 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택되는, 방법.

청구항 260

제 257항에 있어서, 상기 링커가 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결하는, 방법.

청구항 261

제 260항에 있어서, 상기 링커가 서열 GPGPG를 포함하는, 방법.

청구항 262

제 205항 내지 제 223항 또는 제 225항 내지 제 261항 중 어느 한 항에 있어서, 상기 항원-코딩 핵산 서열이 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결되는, 방법.

청구항 263

제 262항에 있어서, 상기 분리 또는 인접 서열이 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열이 위치 76에서 Gly에서 Ala로의 치환을 함유), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열 중 적어도 하나를 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76인, 방법.

청구항 264

제 205항 내지 제 223항 또는 제 225항 내지 제 263항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 방법.

청구항 265

제 205항 내지 제 223항 또는 제 225항 내지 제 263항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩핵산 서열을 코딩하는, 방법.

청구항 266

제 205항 내지 제 223항 또는 제 225항 내지 제 263항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하는, 방법.

청구항 267

제 205항 내지 제 223항 또는 제 225항 내지 제 263항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩하는, 방법.

청구항 268

제 224항에 있어서, 적어도 2개의 MHC 클래스 I 에피토프가 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시되는, 방법.

청구항 269

제 205항 내지 제 223항 또는 제 225항 내지 제 268항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열이 8 내지 35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 방법.

청구항 270

제 205항 내지 제 223항 또는 제 225항 내지 제 269항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하는, 방법.

청구항 271

제 221항 내지 제 223항 또는 제 225항 내지 제 269항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드 서열과 구별되게 하는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 272

제 205항 내지 제 223항 또는 제 225항 내지 제 271항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열이 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 20-40개의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 방법.

청구항 273

제 205항 내지 제 223항 또는 제 225항 내지 제 272항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함하는, 방법.

청구항 274

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 273항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열이 유도성인, 방법.

청구항 275

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 273항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열이 비-유도성인, 방법.

청구항 276

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 275항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 알파바이러스에 대해 천연인 폴리(A) 서열을 포함하는, 방법.

청구항 277

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 275항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 알파바이러스에 대해 외인성 폴리(A) 서열을 포함하는, 방법.

청구항 278

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 277항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 하나의 핵산 서열 중 적어도 하나에 작동가능하게 연결되는, 방법.

청구항 279

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 278항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드인, 방법.

청구항 280

제 205항, 제 207항 내지 제 223항 또는 제 225항 내지 제 278항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 100개의 연속적인 A 뉴클레오티드인, 방법.

청구항 281

제 205항 내지 제 223항 또는 제 225항 내지 제 280항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 방법:

(a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상에서 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 282

제 224항에 있어서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 방법:

(a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상에서 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 283

제 281항에 있어서, 선택된 에피토프 세트의 수가 2-20인, 방법.

청구항 284

제 281항 내지 제 283항 중 어느 한 항에 있어서, 상기 제시 모델은

(a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및

(b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열의 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타내는, 방법.

청구항 285

제 281항 내지 제 284항 중 어느 한 항에 있어서, 선택된 에피토프 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 286

제 281항 내지 제 285항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 287

제 281항 내지 제 286항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하고, 임의로 여기서 APC는 수지상 세포(DC)인, 방법.

청구항 288

제 281항 내지 제 287항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 대상체의 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 289

제 281항 내지 제 288항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 290

제 281항 내지 제 289항 중 어느 한 항에 있어서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터가 감염된 조직에 대해 시퀀싱을 수행함으로써 수득되는, 방법.

청구항 291

제 290항에 있어서, 상기 시퀀싱이 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근인, 방법.

청구항 292

제 188항 내지 제 291항 중 어느 한 항에 있어서, 상기 ChAdV 벡터는:

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래

펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하며,

여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결되는, 방법.

청구항 293

제 188항 내지 제 291항 중 어느 한 항에 있어서, 상기 ChAdV 벡터는,

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은

(i) 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열로서, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있는, 변형된 ChAdV68 서열;

(ii) CMV 프로모터 뉴클레오티드 서열; 및

(iii) SV40 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는,

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아 유래-펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열,

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하는 카세트를 포함하며, 여기서 카세트는 E1 결실 내에 삽입되고 카세트는 CMV 프로모터 뉴클레오티드 서열 및 SV40 폴리(A) 서열에 작동가능하게 연결되는, 방법.

청구항 294

제 188항 내지 제 293항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 2개 이상의 별개의 에피토프를 코딩하는 핵산 서열을 포함하는, 방법.

청구항 295

제 188항 내지 제 293항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 1-10 사이, 1-20 사이, 1-30 사이, 1-40 사이, 1-50 사이, 1-100 사이, 1-200 사이, 1-300 사이, 1-400 사이, 또는 1-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열은 2-10 사이, 2-20 사이, 2-30 사이, 2-40 사이, 2-50 사이, 2-100 사이, 2-200 사이, 2-300 사이, 2-400 사이, 또는 2-500개 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 296

제 188항 내지 제 293항 중 어느 한 항에 있어서, 감염성 질병 유기체에서 발현되는 펩티드를 코딩하는 핵산 서열이 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인으로부터 선택된 펩티드를 코딩하는 핵산 서열을 포함하고, 임의로 여기서 감염성 질병 유기체는 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인 및 이들의 조합으로부터 선택된 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함하는, 방법.

청구항 297

제 188항 내지 제 293항 중 어느 한 항에 있어서, 코딩된 펩티드 또는 펩티드들이 대상체에서 발현될 때 면역

반응을 자극할 수 있고, 임의로 여기서 면역 반응이 T 세포 반응 및/또는 B 세포 반응인, 방법.

청구항 298

제 188항 내지 제 297항 중 어느 하나 한 항에 있어서, 상기 감염성 질병 유기체가 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 방법.

청구항 299

제 188항 내지 제 298항 중 어느 하나 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 세포의 표면 상에서 MHC 클래스 I에 의해 제시되는 것으로 알려지거나 의심되는 에피토프를 코딩하고, 임의로 여기서 세포의 표면이 감염된 세포 표면 또는 감염된 세포 표면이고, 임의로 여기서 세포는 대상체의 세포인, 방법.

청구항 300

제 299항에 있어서, 상기 세포가 병원체 감염 세포, 바이러스 감염 세포, 박테리아 감염 세포, 진균 감염 세포, 및 기생충 감염 세포로 이루어진 군으로부터 선택된 감염된 세포인, 방법.

청구항 301

제 300항에 있어서, 상기 바이러스로 감염된 세포가 HIV 감염 세포, HPV 감염 세포, SARS 감염 세포, SARS-CoV-2 감염 세포, 에볼라 감염 세포, HBV 감염 세포, 인플루엔자 감염 세포, HCV 감염 세포, CMV 감염 세포, 치쿤구니야 바이러스 감염 세포, RSV 감염 세포, 뎅기열 바이러스 감염 세포, 오르티믹소바이러스과 바이러스 감염 세포 및 결핵 감염 세포로 이루어진 군으로부터 선택되는, 방법.

청구항 302

제 292항 또는 제 294항 내지 제 301항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'부터 3'로 하기를 포함하는 식으로 설명되며,

P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

여기서, P는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 적어도 하나의 프로모터 서열을 포함하고, 여기서 a = 1이고,

N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고,

L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고,

L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고,

G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고,

G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고,

U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고,

X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고, 및

Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열인, 방법.

청구항 303

제 302항에 있어서, 각각의 X에 대해 상응하는 N_c가 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열인, 방법.

청구항 304

제 302항 또는 제 303항에 있어서, 각각의 Y에 대해 상응하는 U_f가 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열인, 방법.

청구항 305

제 302항 내지 제 304항 중 어느 한 항에 있어서,

b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2이고,

P는 CMV 프로모터 서열이고,

각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고,

L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고,

U는 각각 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고,

ChAdV 백터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고, 및 항원 카세트는 E1 결실 내에 삽입되고, 그리고

각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 길이가 25개 아미노산인 폴리펩티드를 코딩하는, 방법.

청구항 306

제 292항 또는 제 294항 내지 제 304항 중 어느 한 항에 있어서, 상기 카세트가 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합되는, 방법.

청구항 307

제 292항, 제 294항 내지 제 304항 또는 제 306항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 상기 카세트에 작동가능하게 연결된, 방법.

청구항 308

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 307항 중 어느 한 항에 있어서, 상기 ChAdV 백본이 ChAdV68 벡터 백본을 포함하는, 방법

청구항 309

제 308항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호: 1에 제시된 서열을 포함하는, 방법.

청구항 310

제 308항에 있어서, 상기 ChAdV68 벡터 백본은 ChAdV68 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 아데노바이러스 E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택되는 적어도 하나의 유전자에서 기능적 결실을 포함하고, 임의로 여기서 상기 아데노바이러스 백본 또는 변형된 ChAdV68 서열은 아데노바이러스 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 (1) E1A 및 E1B; 또는 (2) E1A, E1B 및 E3에서 완전히 결실되거나 기능적으로 결실되고, 임의로 여기서 E1 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 577 내지 3403의 E1 결실을 통해 기능적으로 결실되고 및 임의로 여기서 E3 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 27,125 내지 31,825의 E3 결실을 통해 기능적으로 결실되는, 방법.

청구항 311

제 308항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호: 1에 제시된 서열의 염기쌍 번호 577과 3403 사이 또는 염기쌍 456과 3014 사이에 하나 이상의 결실을 포함하고, 임의로 여기서 벡터가 염기쌍 27,125와 31,825 사이 또는 염기쌍 27,816과 31,333 사이에 하나 이상의 결실을 추가로 포함하는, 방법.

청구항 312

제 308항에 있어서, 상기 ChAdV68 벡터 백본이 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518은:

A. E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403;

B. E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825;

C. 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642;

D. 서열번호: 1에 제시된 서열의 뉴클레오티드 456 내지 3014;

E. 서열번호: 1에 제시된 서열의 뉴클레오티드 27,816 내지 31,333;

F. 서열번호: 1에 제시된 서열의 뉴클레오티드 3957 내지 10346;

G. 서열번호: 1에 제시된 서열의 뉴클레오티드 21787 내지 23370;

H. 서열번호: 1에 제시된 서열의 뉴클레오티드 33486 내지 36193; 또는

이들의 조합이 결여된, 방법.

청구항 313

제 308항 내지 제 312항 중 어느 한 항에 있어서, 상기 ChAdV68 벡터 백본이 부분적으로 결실된 E4 유전자를 포함하고, 임의로 여기서 부분적으로 결실된 E4 유전자는:

A. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642가 결여된 E4 유전자 서열,

B. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 34,942, 뉴클레오티드 34,952 내지 35,305, 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642가 결여된 E4 유전자 서열,

C. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516이 결여된 E4 유전자 서열,

D. 벡터가 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는, 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,979 내지 35,642가 결여된 E4 유전자 서열,

E. E4Orf2의 적어도 부분 결실, 완전히 결실된 E4Orf3 및 E4Orf4의 적어도 부분 결실의 E4 결실,

F. E4Orf2의 적어도 부분 결실, E4Orf3의 적어도 부분 결실, 및 E4Orf4의 적어도 부분 결실의 E4 결실,

G. E4Orf1의 적어도 부분 결실, 완전히 결실된 E4Orf2, 및 E4Orf3의 적어도 부분 결실의 E4 결실, 또는

H. E4Orf2의 적어도 부분 결실 및 E4Orf3의 적어도 부분 결실의 E4 결실을 포함하는, 방법.

청구항 314

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 313항 중 어느 한 항에 있어서, 상기 카세트가 E1 영역, E3 영역 및/또는 카세트의 혼입을 허용하는 임의의 결실된 AdV 영역에서 상기 ChAdV 백본에 삽입되는, 방법.

청구항 315

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 314항 중 어느 한 항에 있어서, 상기 ChAdV 백본이 1세대, 2세대, 또는 헬퍼-의존성 아데노바이러스 벡터 중 하나로부터 생성되는, 방법.

청구항 316

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 315항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 CMV, SV40, EF-1, RSV, PGK, HSA, MCK 및 EBV 프로모터 서열로 이루어진 군으로부터 선택되는, 방법.

청구항 317

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 315항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 CMV 프로모터 서열인, 방법.

청구항 318

제 188항 내지 제 317항 중 어느 한 항에 있어서, 적어도 하나의 에피토프-코딩 핵산 서열이 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 I에 의해 제시될 수 있는 에피토프를 코딩하는, 방법.

청구항 319

제 188항 내지 제 318항 중 어느 한 항에 있어서, 적어도 하나의 에피토프-코딩 핵산 서열이 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 II에 의해 제시될 수 있는 에피토프를 코딩하는, 방법.

청구항 320

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 319항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 2개 이상의 항원-코딩 핵산 서열을 포함하는, 방법.

청구항 321

제 320항에 있어서, 각각의 항원-코딩 핵산 서열이 서로 직접 연결된, 방법.

청구항 322

제 292항 내지 제 304항 또는 제 306항 내지 제 321항 중 어느 한 항에 있어서, 각각의 항원-코딩 핵산 서열이 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결되는, 방법.

청구항 323

제 322항에 있어서, 상기 링커가 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결하는, 방법.

청구항 324

제 323항에 있어서, 상기 링커가 (1) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 글리신 잔기; (2) 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이의 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원을 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택되는, 방법.

청구항 325

제 322항에 있어서, 상기 링커가 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결하는, 방법.

청구항 326

제 325항에 있어서, 상기 링커가 서열 GPGPG를 포함하는, 방법.

청구항 327

제 292항 내지 제 304항 또는 제 306항 내지 제 326항 중 어느 한 항에 있어서, 상기 항원-코딩 핵산 서열이 항

원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결되는, 방법.

청구항 328

제 327항에 있어서, 상기 분리 또는 인접 서열이 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열이 위치 76에서 Gly에서 Ala로의 치환을 함유), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열 중 적어도 하나를 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76인, 방법.

청구항 329

제 188항 내지 제 328항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 친화도를 갖도록 하는 적어도 하나의 변경을 포함하는, 방법.

청구항 330

제 188항 내지 제 329항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 안정성을 갖도록 하는 적어도 하나의 변경을 포함하는, 방법.

청구항 331

제 188항 내지 제 330항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자 상에서 증가된 제시 가능성을 갖도록 하는 적어도 하나의 변경을 포함하는, 방법.

청구항 332

제 188항 내지 제 331항 중 어느 한 항에 있어서, 상기 적어도 하나의 변경이 점 돌연변이, 프레임시프트 돌연변이, 비-프레임시프트 돌연변이, 결실 돌연변이, 삽입 돌연변이, 스플라이스 변이체, 게놈 재배열 또는 프로테아좀-생성 스플라이싱된 항원을 포함하는, 방법.

청구항 333

제 188항 내지 제 332항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 감염을 갖는 것으로 알려지거나 의심되는 대상체에서 발현되는 것으로 알려지거나 의심되는 에피토프를 코딩하는, 방법.

청구항 334

제 333항에 있어서, 상기 감염이 병원체 감염, 바이러스 감염, 박테리아 감염, 진균 감염 및 기생충 감염으로 이루어진 군으로부터 선택되는, 방법.

청구항 335

제 334항에 있어서, 상기 바이러스 감염이 HIV 감염, HPV 감염, SARS 감염, SARS-CoV-2 감염, 에볼라 감염, HBV 감염, 인플루엔자 감염, HCV 감염, CMV 감염, 치쿤구니야 바이러스 감염, RSV 감염, 뎅기열 바이러스 감염, 오르티믹소바이러스과 바이러스 감염, 및 결핵 감염로 이루어진 군으로부터 선택되는, 방법.

청구항 336

제 334항에 있어서, 상기 박테리아 감염이 결핵 감염인, 방법.

청구항 337

제 292항 내지 제 304항 또는 제 306항 내지 제 336항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기

서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 방법.

청구항 338

제 292항 내지 제 304항 또는 제 306항 내지 제 336항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩하는, 방법.

청구항 339

제 292항 내지 제 304항 또는 제 306항 내지 제 336항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열이 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하는, 방법.

청구항 340

제 292항 내지 제 304항 또는 제 306항 내지 제 336항 중 어느 한 항에 있어서, 상기 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩하는, 방법.

청구항 341

제 340에 있어서, 적어도 2개의 MHC 클래스 I 에피토프가 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시되는, 조성물.

청구항 342

제 292항 내지 제 304항 또는 제 306항 내지 제 341항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8 내지 35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 방법.

청구항 343

제 292항 내지 제 304항 또는 제 306항 내지 제 342항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하는, 방법.

청구항 344

제 292항 내지 제 304항 또는 제 306항 내지 제 342항 중 어느 한 항에 있어서, 상기 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드 서열과 구별되게 하는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하는, 방법.

청구항 345

제 292항 내지 제 304항 또는 제 306항 내지 제 344항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40의 아미노산 길이의 폴리펩티드 서열을 코딩하는, 방법.

청구항 346

제 292항 내지 제 304항 또는 제 306항 내지 제 345항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 적어도 하나의 보편적인 서열은 파상풍

톡소이드 및 PADRE 중 적어도 하나를 포함하는, 방법.

청구항 347

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 346항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 유도성인, 방법.

청구항 348

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 346항 중 어느 한 항에 있어서, 상기 적어도 하나의 프로모터 뉴클레오티드 서열이 비-유도성인, 방법.

청구항 349

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 348항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 소 성장 호르몬(BGH) SV40 폴리A 서열을 포함하는, 방법.

청구항 350

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 349항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드인, 방법.

청구항 351

제 292항, 제 294항 내지 제 304항 또는 제 306항 내지 제 349항 중 어느 한 항에 있어서, 상기 적어도 하나의 폴리(A) 서열이 적어도 100개의 연속적인 A 뉴클레오티드인, 방법.

청구항 352

제 188항 내지 제 351항 중 어느 한 항에 있어서, 상기 카세트가 인트론 서열, 우드척 간염 바이러스 전사후 조절 요소(WPRE) 서열, 내부 리보솜 진입 서열(IRES) 서열, 2A 자가-절단 펩티드 서열을 코딩하는 뉴클레오티드 서열, 푸린 절단 부위를 코딩하는 뉴클레오티드 서열, 또는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 mRNA의 핵 엑스포트, 안정성 또는 번역 효율을 향상시키는 것으로 알려진 5' 또는 3' 비-코딩 영역의 서열 중 적어도 하나를 추가로 포함하는, 방법.

청구항 353

제 188항 내지 제 352항 중 어느 한 항에 있어서, 상기 카세트가 녹색 형광 단백질(GFP), GFP 변이체, 분비된 알칼리성 포스파타제, 루시퍼라제, 루시페라제 변이체, 또는 검출가능한 펩티드 또는 에피토프를 포함하나 이에 제한되지 않는 리포터 유전자를 추가로 포함하는, 방법.

청구항 354

제 353항에 있어서, 상기 검출가능한 펩티드 또는 에피토프가 HA 태그, 플래그(Flag) 태그, His-태그, 또는 V5 태그로 이루어진 군으로부터 선택되는, 방법.

청구항 355

제 188항 내지 제 354항 중 어느 한 항에 있어서, 상기 하나 이상의 벡터가 적어도 하나의 면역 조절자를 코딩하는 하나 이상의 핵산 서열을 추가로 포함하는, 방법.

청구항 356

제 355항에 있어서, 상기 면역 조절자가 항-CTLA4 항체 또는 이의 항원-결합 단편, 항-PD-1 항체 또는 이의 항원-결합 단편, 항-PD-L1 항체 또는 이의 항원-결합 단편, 항-4-1BB 항체 또는 이의 항원-결합 단편, 또는 항-OX-40 항체 또는 이의 항원-결합 단편인, 방법.

청구항 357

제 356항에 있어서, 상기 항체 또는 이의 항원-결합 단편이 단일 특이성 또는 다중 특이성으로 함께 연결된 Fab 단편, Fab' 단편, 단일 사슬 Fv(scFv), 단일 도메인 항체(sdAb)(예를 들어, 낙타류 항체 도메인), 또는 전장 단일-사슬 항체(예를 들어, 가요성 링커에 의해 연결된 중쇄 및 경쇄가 있는 전장 IgG)인, 방법.

청구항 358

제 356항 또는 제 357항에 있어서, 상기 항체의 중쇄 및 경쇄 서열이 2A 또는 IRES와 같은 자가-절단 서열에 의해 분리된 연속 서열이거나; 또는 상기 항체의 중쇄 및 경쇄 서열은 연속적인 글리신 잔기와 같은 가요성 링커에 의해 연결되는, 방법.

청구항 359

제 355항에 있어서, 상기 면역 조절자가 사이토카인인, 방법.

청구항 360

제 359항에 있어서, 상기 사이토카인이 IL-2, IL-7, IL-12, IL-15, 또는 IL-21 또는 이들 각각의 변이체 중 적어도 하나인, 방법.

청구항 361

제 292항 내지 제 304항 또는 제 306항 내지 제 360항 중 어느 한 항에 있어서, 상기 에피토프-코딩 핵산 서열이 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열 하기 단계를 수행함으로써 선택되는, 방법:

(a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 362

제 305항에 있어서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택되는, 방법:

(a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계로서, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되는, 단계;

(b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계로서, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되는, 단계; 및

(c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계.

청구항 363

제 361항에 있어서, 선택된 에피토프 세트의 수가 2-20인, 방법.

청구항 364

제 361항 내지 제 363항 중 어느 한 항에 있어서, 상기 제시 모델은

(a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및

(b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열의 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타내는, 방법.

청구항 365

제 361항 내지 제 364항 중 어느 한 항에 있어서, 선택된 에피토프 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 366

제 361항 내지 제 365항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 367

제 361항 내지 제 366항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하고, 임의로 여기서 APC는 수지상 세포(DC)인, 방법.

청구항 368

제 361항 내지 제 367항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 369

제 361항 내지 제 368항 중 어느 한 항에 있어서, 선택된 에피토프의 세트를 선택하는 것이 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함하는, 방법.

청구항 370

제 361항 내지 제 369항 중 어느 한 항에 있어서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터가 감염된 조직에 대해 시퀀싱을 수행함으로써 수득되는, 방법.

청구항 371

제 370항에 있어서, 상기 시퀀싱이 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법인, 방법.

청구항 372

제 188항 내지 제 371항 중 어느 한 항에 있어서, 상기 카세트가 카세트 내의 인접 서열에 의해 형성된 접합 에피토프 서열을 포함하는, 방법.

청구항 373

제 372항에 있어서, 적어도 하나 또는 각각의 접합 에피토프 서열이 MHC에 대해 500 nM 초과의 친화도를 갖는, 방법.

청구항 374

제 372항 또는 제 373항에 있어서, 각각의 접합 에피토프 서열이 비-자기인, 방법.

청구항 375

제 188항 내지 제 374항 중 어느 한 항에 있어서, 상기 카세트가 번역된 야생형 핵산 서열을 포함하는 비-치료적 MHC 클래스 I 또는 클래스 II 에피토프 핵산 서열을 코딩하지 않고, 여기서 비-치료적 에피토프는 대상체의

MHC 대립유전자 상에 표시될 것으로 예측되는, 방법.

청구항 376

제 375항에 있어서, 상기 비-치료적 예측된 MHC 클래스 I 또는 클래스 II 에피토프 서열이 카세트에서 인접 서열에 의해 형성된 접합 에피토프 서열인, 방법.

청구항 377

제 372항 내지 제 376항 중 어느 한 항에 있어서, 상기 예측이 비-치료적 에피토프의 서열을 제시 모델에 입력함으로써 생성된 제시 가능성에 기초하는, 방법.

청구항 378

제 372항 내지 제 377항 중 어느 한 항에 있어서, 카세트 내의 항원-코딩 핵산 서열의 순서가 다음을 포함하는 일련의 단계에 의해 결정되는, 방법:

(a) 항원-코딩 핵산 서열의 상이한 순서에 상응하는 후보 카세트 서열 세트를 생성하는 단계;

(b) 각각의 후보 카세트 서열에 대해 후보 카세트 서열에서 비-치료적 에피토프의 제시에 기초하여 제시 점수를 결정하는 단계; 및

(c) 백신에 대한 카세트 서열로서 미리 결정된 임계값 미만의 제시 점수와 연관된 후보 카세트 서열을 선택하는 단계.

청구항 379

제 188항 내지 제 378항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 상기 조성물이 약제학적으로 허용되는 담체를 포함하는 약제학적 조성물로 제형화되는, 방법.

청구항 380

제 188항 내지 제 379항 중 어느 한 항에 있어서, 하나 이상의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되는, 방법.

청구항 381

제 188항 내지 제 380항 중 어느 한 항에 있어서, 각각의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되는, 방법.

청구항 382

제 188항 내지 제 381항 중 어느 한 항에 있어서, 하나 이상의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되지 않는, 방법.

청구항 383

제 188항 내지 제 382항 중 어느 한 항에 있어서, 각각의 에피토프-코딩 핵산 서열이 대상체의 감염 또는 감염된 세포로부터 유래되지 않는, 방법.

청구항 384

대상체에서 면역 반응을 자극하는 방법으로서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 단계를 포함하며, 여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은,

(A) 하나 이상의 벡터를 포함하는 자가-복제 알파바이러스-기반 발현 시스템으로서, 여기서 하나 이상의 벡터는,

(a) RNA 알파바이러스 백본으로서, 여기서 RNA 알파바이러스 백본은:

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, RNA 알파바이러스 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) 다음을 포함하는 적어도 하나의 항원-코딩 핵산 서열:

a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군에서 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열로서, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 핵산 서열;

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열;

(ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제2 프로모터 뉴클레오티드 서열; 및

(iii) 알파바이러스에 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열인, 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하는, 카세트를 포함하는, 자가-복제 알파바이러스-기반 발현 시스템; 및

(B) 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화하는 지질-나노입자(LNP)를 포함하는, 방법.

청구항 385

대상체에서 면역 반응을 자극하는 방법으로서, 침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 단계를 포함하며, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은,

ChAdV-기반 발현 시스템을 포함하며, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는,

(a) ChAdV 백본으로서, 여기서 ChAdV 백본은

(i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및

(ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하는, ChAdV 백본; 및

(b) 카세트로서, 여기서 카세트는

(i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열로서, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되는, 핵산 서열;

b. 임의로 5' 링커 서열, 및

c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하눈 카세트를 포함하며,

여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결되는, 방법.

청구항 386

제 384항 또는 제 385항에 있어서, 자가-복제 알파바이러스-기반 발현 시스템 또는 ChAdV-기반 발현 시스템이 상기 청구항 중 어느 한 항의 특징 중 임의의 것을 포함하는, 방법.

청구항 387

제 188항 내지 제 386항 중 어느 한 항에 있어서, ChAdV-기반 발현 시스템의 전달을 위한 조성물의 카세트가 자

가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물의 카세트와 동일한, 방법.

발명의 설명

기 술 분 야

배 경 기 술

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2020년 4월 3일에 출원된 미국 가출원 번호 63/005,160의 이익을 주장하며, 이는 모든 목적을 위해 그 전체가 참고로 본원에 포함된다.

[0003] 서열 목록

[0004] 본 출원은 ASCII 형식으로 전자적으로 제출된 서열 목록을 포함하고, 이는 그 전체가 참고로 포함된다. 2021년 4월 5일에 생성된 ASCII 사본의 명칭은 GSO_090WO_sequencelisting.txt이고 크기는 422,179바이트이다.

[0005] 항원 백신 디자인에 대한 한 가지 질문은 존재하는 많은 코딩 돌연변이 중 어느 것이 "최상의" 치료 항원, 예를 들어 면역을 유발할 수 있는 항원을 생성하는지 여부이다.

[0006] 현재의 항원 예측 방법의 문제에 더하여, 인간에서 항원 전달을 위해 사용될 수 있는 이용가능한 벡터 시스템에 대한 특정 도전도 존재하며, 그 중 다수는 인간으로부터 유래된다. 예를 들어, 많은 인간은 이전의 자연 노출의 결과로 인간 바이러스에 대한 기존 면역성을 갖고 있으며, 이 면역성은 감염성 질병 백신과 같은 백신 접종을 위한 항원 전달을 위한 재조합 인간 바이러스의 사용에 주요 장애물이 될 수 있다.

발명의 내용

[0007] 본원에서는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물을 개시되고, 여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) 다음을 포함하는 적어도 하나의 항원-코딩 핵산 서열: a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열; (ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제 2 프로모터 뉴클레오티드 서열; 및 (iii) 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하고, 여기서 제 2 폴리(A) 서열은 알파바이러스에 대한 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열이고, 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0008] 또한, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물이 본원에 개시되어 있으며, 여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 여기서 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 서열번호: 6에 제시된 핵산 서열을 포함하고, 여기서 RNA 알파바이러스 백본 서열은 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열을 포함하고, 여기서 26S 프로모터 서열은 RNA 알파바이러스 백본에 내인성이고, 폴리(A) 서열은 RNA 알파바이러스 백본에 내인성이고; 및 (b) 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합된 카세트를 포함하고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열에 작동가능하게 연결되고, 여기서 카세트는 a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0009] 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를

코딩하는 핵산 서열은 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 1-10, 1-20, 1-30, 1-40, 1-50, 1-100, 1-200, 1-300, 1-400개, 또는 1-500 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 2-10, 2-20, 2-30, 2-40, 2-50, 2-100, 2-200, 2-300, 2-400, 또는 2-500 사이의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 전장 단백질을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 단백질의 단백질 도메인을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 단백질의 단백질 서브유닛을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 하기로부터 선택되는 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함한다: 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인, 및 이의 조합.

[0010] 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 면역 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 T 세포 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 B 세포 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 T 세포 반응 및 B 세포 반응을 자극할 수 있다.

[0011] 일부 측면에서, 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택된다.

[0012] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'에서 3'로 식 P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$를 포함하고, 여기서 P는 제 2 프로모터 뉴클레오티드 서열을 포함하고, 여기서 a = 0 또는 1이고, N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고, L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고, L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고, G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고, G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고, U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고, X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고 Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 X에 대해 상응하는 N_c는 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 Y에 대해 상응하는 U_f는 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열이다. a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2인 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 제공되는 단일 26S 프로모터 뉴클레오티드 서열이고, 적어도 하나의 폴리아데닐화 폴리(A) 서열은 RNA 알파바이러스 백본에 의해 제공되는 적어도 100개의 연속적인 A 뉴클레오티드의 폴리(A) 서열이고, 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합되고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열에 작동가능하게 연결되고, 각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고, L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, U는 각각의 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고, RNA 알파바이러스 백본은 서열번호: 6에 제시된 서열이고, 각각의 MHC는 클래스 I 에피토프-코딩 핵산 서열은 길이가 13-25개 아미노산인 폴리펩티드를 코딩한다.

[0013] 일부 측면에서, LNP는 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, PEG-기반 코트 지질, 또는 이들의 조합으로 이루어진 군으로부터 선택된 지질을 포함한다. 일부 측면에서, LNP는 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, 및 PEG-기반 코트 지질을 포함한다. 일부 측면에서, 이온화가능한 아미노 지질은 MC3-유사(디리놀레일메틸-4-디메틸아미노부티레이트) 분자를 포함한다. 일부 측면에서, LNP-캡슐화된 발현 시스템은 약 100 nm의 직경을 갖는다.

[0014] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여를 위해 제형화된다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM) 투여를 위해 제형화된다.

[0015] 일부 측면에서, 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합된다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 카세트에 작동가능하게 연결된다.

[0016] 일부 측면에서, 하나 이상의 벡터는 하나 이상의 +-가닥 RNA 벡터를 포함한다. 일부 측면에서, 하나 이상의 +-가닥 RNA 벡터는 5'7-메틸구아노신(m7g) cap을 포함한다. 일부 측면에서, 하나 이상의 +-가닥 RNA 벡터는 시험관내 전사에 의해 생성된다. 일부 측면에서, 하나 이상의 벡터는 포유동물 세포 내에서 자가-복제한다. 일부 측면에서, RNA 알파바이러스 백본은 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 베네수엘라 말 뇌염 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩된 적어도 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열, 비구조 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자, 및 nsP4 유전자를 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩된 적어도 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열을 포함한다. 일부 측면에서, 비구조적 단백질-매개 증폭을 위한 서열은 알파바이러스 5' UTR, 51-nt CSE, 24-nt CSE, 26S 서브게놈 프로모터 서열, 19-nt CSE, 알파바이러스 3' UTR, 또는 이들의 조합으로 이루어진 군으로부터 선택된다. 일부 측면에서, RNA 알파바이러스 백본은 구조적 비리온 단백질 캡시드, E2 및 E1을 코딩하지 않는다. 일부 측면에서, 카세트는 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열 내의 구조적 비리온 단백질 대신 삽입된다. 일부 측면에서, 베네수엘라 말 뇌염 바이러스는 서열번호: 3 또는 서열번호: 5의 서열을 포함한다. 일부 측면에서, 베네수엘라 말 뇌염 바이러스는 염기쌍 7544와 11175 사이의 결실을 추가로 포함하는 서열번호: 3 또는 서열번호: 5의 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 서열번호: 6 또는 서열번호: 7의 서열을 포함한다. 일부 측면에서, 카세트는 서열번호: 3 또는 서열번호: 5의 서열에 제시된 염기쌍 7544와 11175 사이의 결실을 대체하기 위해 위치 7544에 삽입된다. 일부 측면에서, 카세트의 삽입은 nsP1-4 유전자 및 적어도 하나의 핵산 서열을 포함하는 폴리시스트론 RNA의 전사를 제공하고, 여기서 nsP1-4 유전자 및 적어도 하나의 핵산 서열은 분리된 열린 해독틀에 있다.

[0017] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 코딩되는 천연 26S 프로모터 뉴클레오티드 서열이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 외인성 RNA 프로모터이다. 일부 측면에서, 제 2 프로모터 뉴클레오티드 서열은 26S 프로모터 뉴클레오티드 서열이다. 일부 측면에서, 제 2 프로모터 뉴클레오티드 서열은 다수의 26S 프로모터 뉴클레오티드 서열을 포함하고, 여기서 각각의 26S 프로모터 뉴클레오티드 서열은 하나 이상의 분리된 열린 해독틀의 전사를 제공한다.

[0018] 일부 측면에서, 하나 이상의 벡터는 각각 크기가 적어도 300 nt이다. 일부 측면에서, 하나 이상의 벡터는 각각 크기가 적어도 1 kb이다. 일부 양상에서, 하나 이상의 벡터는 각각 크기가 2 kb이다. 일부 측면에서, 하나 이상의 벡터는 각각 크기가 5 kb 미만이다.

[0019] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 서로 직접 연결된다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된다. 일부 측면에서, 링커는 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 (1) 연속적인 글리신 잔기, 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 잔기; (2) 연속적인 알라닌 잔기, 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기 길이; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래되고 항원을 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 링커는 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 II 서열을

MHC 클래스 I 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 서열 GPGPG를 포함한다.

[0020] 일부 측면에서, 항원-코딩 핵산 서열은 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결된다. 일부 측면에서, 분리 또는 인접 서열은 적어도 하나의 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열은 위치 76에서 Gly에서 Ala로의 치환을 함유한다), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열을 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76이다.

[0021] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열를 포함하고, 임의로 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 최대 400개의 항원-코딩 핵산 서열를 포함한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩한다. 일부 측면에서, MHC 클래스 I 에피토프는 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시된다.

[0022] 일부 측면에서, 에피토프-코딩 핵산 서열은 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8-35, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩한다.

[0023] 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재한다. 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드와 구별되게 만드는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40개 아미노산 길이인 폴리펩티드 서열을 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함한다.

[0024] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열은 유도성이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열은 비-유도성이다.

[0025] 일부 측면에서, 적어도 하나의 폴리(A) 서열은 알파바이러스에 천연 폴리(A) 서열을 포함한다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 알파바이러스에 외인성인 폴리(A) 서열을 포함한다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 하나의 핵산 서열 중 적어도 하나에 작동가능하게 연결된다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 100개의 연속적인 A 뉴클레오티드이다.

[0026] 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생

성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되며; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및; 및 (c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 선택된 에피토프 세트의 수는 2-20이다. 일부 측면에서, 제시 모델은 (a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및 (b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열이 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타낸다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브(naïve) T 세포에 제시될 수 있는 증가된 가능성을 갖는 에피토프를 선택하는 것을 포함하며, 임의로 여기서 APC는 수지상 세포(DC)이다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터는 감염된 조직에 대해 시퀀싱을 수행함으로써 수득된다. 일부 측면에서, 시퀀싱은 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법이다.

[0027] 침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물이 또한 본원에 개시되며, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은 다음을 포함한다: ChAdV-기반 발현 시스템, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결된다.

[0028] 또한, ChAdV-기반 발현 시스템의 전달을 위한 조성물이 본원에 개시되며, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은 하기를 포함한다: ChAdV-기반 발현 시스템, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자 바이러스를 포함하고, 여기서 ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 서열번호:1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고; (ii) CMV 프로모터 뉴클레오티드 서열; 및 (iii) SV40 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아 유래-펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고, 여기서 카세트는 E1 결실 내에 삽입되고 카세트는 CMV 프로모터 뉴클레오티드 서열 및 SV40 폴리(A) 서열에 작동가능하게 연결된다.

[0029] 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩

하는 핵산 서열은 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 1-10, 1-20, 1-30, 1-40, 1-50, 1-100, 1-200, 1-300, 1-400개, 또는 1-500의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 2-10, 2-20, 2-30, 2-40, 2-50, 2-100, 2-200, 2-300, 2-400, 또는 2-500의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 전장 단백질을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 단백질의 단백질 도메인을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 단백질의 단백질 서브유닛을 코딩하는 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 전염병 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인, 및 이의 조합으로부터 선택되는 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함한다.

[0030] 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 면역 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 T 세포 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 B 세포 반응을 자극할 수 있다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 T 세포 반응 및 B 세포 반응을 자극할 수 있다.

[0031] 일부 측면에서, 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택된다.

[0032] 일부 측면에서, 에피토프-코딩 핵산 서열은 세포의 표면 상에서 MHC 클래스 I에 의해 제시되는 것으로 알려지거나 의심되는 에피토프를 코딩하고, 임의로 여기서 세포의 표면은 감염된 세포 표면이고, 임의로 여기서 세포는 대상체의 세포이다. 일부 측면에서, 세포는 병원체 감염 세포, 바이러스 감염 세포, 박테리아 감염 세포, 진균 감염 세포, 및 기생충 감염 세포로 이루어진 군으로부터 선택된 감염된 세포이다. 일부 측면에서, 바이러스 감염 세포는 HIV 감염 세포, HPV 감염 세포, SARS 감염 세포, SARS-CoV-2 감염 세포, 에볼라 감염 세포, HBV 감염 세포, 인플루엔자 감염 세포, HCV 감염 세포, CMV 감염 세포, 치쿤구니야 바이러스 감염 세포, RSV 감염 세포, 뎅기열 바이러스 감염 세포, 오르티믹소바이러스과 바이러스 감염 세포 및 결핵 감염 세포를 포함한다.

[0033] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'에서 3'로 식 P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$를 포함하고, 여기서 P는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 적어도 하나의 프로모터 서열을 포함하고, 여기서 a = 1이고, N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고, L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고, L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고, G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고, G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고, U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고, X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고 Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 X에 대해 상응하는 N_c는 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 Y에 대해 상응하는 U_f는 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열이다. b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2인 일부 측면에서, P는 CMV 프로모터 서열이고, 각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고, L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, U는 각각의 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고, ChAdV 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고, 항원 카세트는 E1 결실 내에 삽입되고, 각각의 MHC 클래스

I 항원-코딩 핵산 서열은 길이가 25개 아미노산인 폴리펩티드를 코딩한다.

[0034] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여를 위해 제형화된다. 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM) 투여를 위해 제형화된다.

[0035] 일부 측면에서, 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합된다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 카세트에 작동가능하게 연결된다.

[0036] 일부 측면에서, ChAdV 백본은 ChAdV68 벡터 백본을 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열을 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 ChAdV68 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 아데노바이러스 E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택되는 적어도 하나의 유전자에서 기능적 결실을 포함하고, 임의로 아데노바이러스 백본 또는 변형된 ChAdV68 서열은 아데노바이러스 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 (1) E1A 및 E1B; 또는 (2) E1A, E1B 및 E3에서 완전히 결실되거나 기능적으로 결실되고, 임의로 여기서 E1 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 577 내지 3403의 E1 결실을 통해 기능적으로 결실되고 및 임의로 E3 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 27,125 내지 31,825의 E3 결실을 통해 기능적으로 결실된다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 염기쌍 번호 577과 3403 사이 또는 염기쌍 456과 3014 사이에 하나 이상의 결실을 포함하고, 임의로 벡터는 염기쌍 27,125와 31,825 사이 또는 염기쌍 27,816과 31,333 사이에 하나 이상의 결실을 추가로 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 다음이 결여되어 있다: A. E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; B. E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; C. 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642; D. 서열번호:1에 제시된 서열의 뉴클레오티드 456 내지 3014; E. 서열번호:1에 제시된 서열의 뉴클레오티드 27,816 내지 31,333; F. 서열번호: 1에 제시된 서열의 뉴클레오티드 3957 내지 10346; G. 서열번호: 1에 제시된 서열의 뉴클레오티드 21787 내지 23370; H. 서열번호: 1에 제시된 서열의 뉴클레오티드 33486 내지 36193; 또는 이들의 조합.

[0037] 일부 측면에서, ChAdV68 벡터 백본은 부분적으로 결실된 E4 유전자를 포함한다. 일부 측면에서, 부분적으로 결실된 E4 유전자는 다음을 포함한다: A. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642가 결여된 E4 유전자 서열, B. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 34,942, 뉴클레오티드 34,952 내지 35,305, 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642가 결여된E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, C. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516이 결여된 E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, D. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,979 내지 35,642가 결여된 E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, E. E4Orf2의 적어도 부분 결실, E4Orf3의 완전히 결실 및 E4Orf4의 적어도 부분 결실의 E4 결실, F. E4Orf2의 적어도 부분 결실, E4Orf3의 적어도 부분 결실, 및 E4Orf4의 적어도 부분 결실의 E4 결실, G. E4Orf1의 적어도 부분 결실, 완전히 결실된 E4Orf2, 및 E4Orf3의 적어도 부분 결실의 E4 결실, 또는 H. E4Orf2의 적어도 부분 결실 및 E4Orf3의 적어도 부분 결실의 E4 결실.

[0038] 일부 측면에서, 카세트는 E1 영역, E3 영역, 및/또는 카세트의 혼입을 허용하는 임의의 결실된 AdV 영역에서 ChAdV 백본에 삽입된다. 일부 측면에서, ChAdV 백본은 1세대, 2세대, 또는 헬퍼-의존성 아데노바이러스 벡터 중 하나로부터 생성된다.

[0039] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 CMV, SV40, EF-1, RSV, PGK, HSA, MCK, 및 EBV 프로모터 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 CMV 프로모터 서열이다.

[0040] 일부 측면에서, 에피토프-코딩 핵산 서열 중 적어도 하나는 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 I에 의해 제시될 수 있는 에피토프를 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열 중 적어도 하나는 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 II에 의해 제시될 수 있는 에피토프를 코딩한다.

[0041] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 서로 직접 연결된다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된다. 일부 측면에서, 링커는 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 (1) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 잔기의 연속적인 글리신 잔기; (2) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 잔기의 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원에 플랭킹하고 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기인 하나 이상의 천연 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 링커는 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 서열 GPGPG를 포함한다.

[0042] 일부 측면에서, 항원-코딩 핵산 서열은 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결된다. 일부 측면에서, 분리 또는 인접 서열은 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열은 위치 76에서 Gly에서 Ala로의 치환을 함유한다), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열 중 적어도 하나의 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76이다.

[0043] 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 친화도를 갖도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 안정성을 갖도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자 상에서 제시될 가능성이 증가하도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 적어도 하나의 변경은 점 돌연변이, 프레임시프트 돌연변이, 비-프레임시프트 돌연변이, 결실 돌연변이, 삽입 돌연변이, 스플라이스 변이체, 게놈 재배열, 또는 프로테아좀-생성 스플라이싱 항원을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 감염이 있는 것으로 알려지거나 의심되는 대상체에서 발현되는 것으로 알려지거나 의심되는 에피토프를 코딩한다. 일부 측면에서, 감염은 병원체 감염, 바이러스 감염, 박테리아 감염, 진균 감염 및 기생충 감염으로 이루어진 군으로부터 선택된다. 일부 측면에서, 바이러스 감염은 HIV 감염, HPV 감염, SARS 감염, SARS-CoV-2 감염, 에볼라 감염, HBV 감염, 인플루엔자 감염, HCV 감염, CMV 감염, 치쿤구니야 바이러스 감염, RSV 감염, 뎅기열 바이러스 감염, 오르티믹소바이러스과 바이러스 감염, 및 결핵 감염으로 이루어진 군으로부터 선택된다. 일부 측면에서, 박테리아 감염은 결핵 감염이다.

[0044] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열을 포함하고, 임의로 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 최대 400개의 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 항원-코딩 핵산 서열 중 적어도 2개는 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩한다. 일부 측면에서, MHC 클래스 I 에피토프 중 적어도 2개는 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시된다. 일부 측면에서, 에피토프-코딩 핵산 서열은 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8-35개 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩한다.

[0045] 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재한다. 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코

딩된 상응하는 펩티드 서열과 구별되게 만드는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40개 아미노산 길이인 폴리펩티드 서열을 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함한다.

[0046] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 유도성이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 비-유도성이다.

[0047] 일부 측면에서, 적어도 하나의 폴리(A) 서열은 소 성장 호르몬(BGH) SV40 폴리A 서열을 포함한다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 100개의 연속적인 A 뉴클레오티드이다.

[0048] 일부 측면에서, 카세트는 인트론 서열, 우드척 간염 바이러스 전사후 조절 요소(WPRE) 서열, 내부 리보솜 진입 서열(IRES) 서열, 2A 자가 절단 펩티드 서열을 코딩하는 뉴클레오티드 서열, 푸린 절단 부위를 코딩하는 뉴클레오티드 서열, 또는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동 가능하게 연결된 mRNA의 핵 엑스포트(nuclear export), 안정성 또는 번역 효율을 향상시키는 것으로 알려진 5' 또는 3' 비-코딩 영역의 서열을 추가로 포함한다. 일부 측면에서, 카세트는 녹색 형광 단백질(GFP), GFP 변이체, 분비된 알칼리성 포스파타제, 루시퍼라제, 루시퍼라제 변이체, 또는 검출가능한 펩티드 또는 에피토프를 포함하나 이에 제한되지 않는 리포터 유전자를 추가로 포함한다. 일부 측면에서, 검출가능한 펩티드 또는 에피토프는 HA 태그, 플래그(Flag) 태그, His-태그, 또는 V5 태그로 이루어진 군으로부터 선택된다.

[0049] 일부 측면에서, 하나 이상의 벡터는 적어도 하나의 면역 조절자를 코딩하는 하나 이상의 핵산 서열을 추가로 포함한다. 일부 측면에서, 면역 조절자는 항-CTLA4 항체 또는 이의 항원-결합 단편, 항-PD-1 항체 또는 이의 항원-결합 단편, 항-PD-L1 항체 또는 이의 항원-결합 단편, 항-4-1BB 항체 또는 이의 항원 결합 단편, 또는 항-OX-40 항체 또는 이의 항원 결합 단편이다. 일부 측면에서, 항체 또는 이의 항원-결합 단편은 단일 특이적 또는 다중 특이성으로서 함께 연결된 Fab 단편, Fab' 단편, 단일 사슬 Fv(scFv), 단일 도메인 항체(sdAb)(예를 들어, 낙타류 항체 도메인), 또는 전장 단일-사슬 항체(예를 들어, 가요성 링커에 의해 연결된 중쇄 및 경쇄를 갖는 전장 IgG)이다. 일부 측면에서, 항체의 중쇄 및 경쇄 서열은 2A 또는 IRES와 같은 자가-절단 서열에 의해 분리된 연속 서열이거나; 또는 항체의 중쇄 및 경쇄 서열은 연속적인 글리신 잔기와 같은 가요성 링커에 의해 연결된다. 일부 측면에서, 면역 조절자는 사이토카인이다. 일부 측면에서, 사이토카인은 적어도 하나의 IL-2, IL-7, IL-12, IL-15 또는 IL-21 또는 각각의 이의 변이체이다.

[0050] 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되며; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 선택된 에피토프 세트의 수는 2-20이다. 일부 측면에서, 제시 모델은 (a) 펩티드

서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및 (b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열이 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타낸다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하며, 임의로 여기서 APC는 수지상 세포(DC)이다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터는 감염된 조직에 대해 시퀀싱을 수행함으로써 획득된다. 일부 측면에서, 시퀀싱은 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법이다.

[0051] 일부 측면에서, 카세트는 카세트의 인접 서열에 의해 형성된 접합 에피토프 서열을 포함한다. 일부 측면에서, 적어도 하나 또는 각각의 접합 에피토프 서열은 MHC에 대해 500 nM 초과의 친화도를 갖는다. 일부 측면에서, 각각의 접합 에피토프 서열은 비-자기이다. 일부 측면에서, 카세트는 번역된 야생형 핵산 서열을 포함하는 비-치료적 MHC 클래스 I 또는 클래스 II 에피토프 핵산 서열을 코딩하지 않으며, 여기서 비-치료적 에피토프는 대상체의 MHC 대립유전자 상에 표시될 것으로 예측된다. 일부 측면에서, 비-치료적 예측 MHC 클래스 I 또는 클래스 II 에피토프 서열은 카세트에서 인접한 서열에 의해 형성된 접합 에피토프 서열이다. 일부 측면에서, 예측은 비-치료적 에피토프의 서열을 제시 모델에 입력함으로써 생성된 제시 가능성에 기초한다. 일부 측면에서, 카세트 내의 항원-코딩 핵산 서열의 순서는 (a) 항원-코딩 핵산 서열의 상이한 순서에 상응하는 후보 카세트 서열 세트를 생성하는 단계; (b) 각각의 후보 카세트 서열에 대해 후보 카세트 서열에서 비-치료적 에피토프의 제시에 기초하여 제시 점수를 결정하는 단계; 및 (c) 백신에 대한 카세트 서열로서 미리 결정된 임계값 미만의 제시 점수와 관련된 후보 카세트 서열을 선택하는 단계를 포함하는 일련의 단계에 의해 결정된다.

[0052] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 약제학적으로 허용되는 담체를 포함하는 약제학적 조성물로 제형화된다.

[0053] 일부 측면에서, 하나 이상의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래된다. 일부 측면에서, 각각의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래된다. 일부 측면에서, 하나 이상의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래되지 않는다. 일부 측면에서, 각각의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래되지 않는다.

[0054] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 서로 선형으로 연결된 적어도 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개의 감염성 질병 유기체-특이적 MHC 클래스 I 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 적어도 하나의 에피토프-코딩 핵산 서열은 서로 선형으로 연결된 적어도 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개의 감염성 질병 유기체-특이적 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함한다.

[0055] 또한, 본원에 기재된 ChAdV-기반 발현 시스템의 전달을 위한 임의의 조성물, 및 사용 설명서를 포함하는 키트가 본원에 제공된다.

[0056] 또한, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물이 본원에 제공되며, 여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과

바이러스 및 결핵으로 이루어진 군으로부터 선택되고; b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열; (ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제 2 프로모터 뉴클레오티드 서열; 및 (iii) 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하고, 여기서 제 2 폴리(A) 서열은 알파바이러스에 대해 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열이고, 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0057] 침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물이 또한 본원에 개시되며, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은: ChAdV-기반 발현 시스템, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되고; b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 여기서 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결된다.

[0058] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템 또는 ChAdV-기반 발현 시스템은 상기 청구항 중 어느 하나의 특징 중 임의의 것을 포함한다.

[0059] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물의 카세트는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물의 카세트와 동일하다.

[0060] 또한, 본원에는 대상체에서 면역 반응을 자극하는 방법이 본원에 개시되어 있으며, 이 방법은 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 단계 및/또는 침팬지-아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 단계를 포함하여, 여기서: a. ChAdV-기반 발현 시스템의 전달을 위한 조성물은 ChAdV 기반 발현 시스템을 포함하고, b. 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 자가-복제 알파바이러스-기반 발현 시스템을 포함하고, 또는 c. ChAdV-기반 발현 시스템의 전달을 위한 조성물은 ChAdV-기반 발현 시스템을 포함하고, 및 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 자가-복제 알파바이러스-기반 발현 시스템을 포함한다.

[0061] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 프라이밍 용량으로 투여되고, ChAdV-기반 발현 시스템의 전달을 위한 조성물 또는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 부스팅 용량으로 투여된다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 프라이밍 용량으로 투여되고 ChAdV-기반 발현 시스템의 전달을 위한 조성물 또는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 부스팅 용량으로 투여된다. 일부 측면에서, 2회 이상의 부스팅 용량이 투여된다. 일부 측면에서, 1, 2, 3, 4, 5, 6, 7 또는 8회의 부스팅 용량이 투여된다.

[0062] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여된다. 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 투여(IM)된다. 일부 측면에서, IM 투여는 분리된 주사 부위에서 투여된다. 일부 측면에서, 분리된 주사 부위는 대향 삼각근에 있다. 일부 측면에서, 분리된 주사 부위는 각 측면의 둔근 또는 대퇴직근 부위에 있다.

[0063] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 근육내(IM), 피내(ID), 피하(SC), 또는 정맥내(IV) 투여된다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물이 투여된다(IM). 일부 측면에서, IM 투여는 분리된 주사 부위에서 투여된다. 일부 측면에서, 분리된 주사 부위는 대향 삼각근에 있다. 일부 측면에서, 분리된 주사 부위는 각 측면의 둔근 또는 대퇴직근 부위에 있다. 일부 측면에서, 하나 이상의 부스팅 용량의 주사 부위는 프라이밍 용량의 주사 부위에 가능한 한 가깝다.

[0064] 일부 측면에서, 방법은 대상체의 HLA-일배체형을 결정하거나 결정한 것을 추가로 포함한다.

[0065] 일부 측면에서, 방법은 니볼루맙을 투여하는 것을 추가로 포함한다. 일부 측면에서, 니볼루맙은 정맥내(IV) 주입으로 투여된다. 일부 측면에서, 니볼루맙은 480 mg의 용량으로 투여된다. 일부 측면에서, 니볼루맙은 1일에

투여된다. 일부 측면에서, 니볼루맙은 1일에 투여되고 프라이밍 용량 후 4주마다(Q4W) 투여된다. 일부 측면에서, 니볼루맙은 프라이밍 투여와 동일한 날 또는 하나 이상의 부스팅 투여와과 동일한 날에 투여된다. 일부 측면에서, 니볼루맙은 10 mg/mL의 용액으로 제형화된다.

[0066] 일부 측면에서, 방법은 이필리무맙을 투여하는 것을 추가로 포함한다. 일부 측면에서, 이필리무맙은 정맥내(IV) 주입으로 투여된다. 일부 측면에서, 이필리무맙은 피하(SC) 투여된다. 일부 측면에서, SC 투여는 프라이밍 용량 주사 부위 또는 하나 이상의 부스팅 용량 주사 부위에 근위(~2cm 이내)로 주사된다. 일부 측면에서, SC 투여는 4회의 개별 주사로 투여되거나 6회의 개별 주사로 투여된다. 일부 측면에서, 이필리무맙은 30 mg의 용량으로 투여된다. 일부 측면에서, 이필리무맙은 1일에 투여된다. 일부 측면에서, 이필리무맙은 1일에 투여되고 프라이밍 용량 후 4주마다(Q4W) 투여된다. 일부 측면에서, 이필리무맙은 프라이밍 용량과 동일한 날 또는 하나 이상의 부스팅 투여와 동일한 날에 투여된다. 일부 측면에서, 이필리무맙은 5 mg/mL의 용액으로 제형화된다.

[0067] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열; (ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제 2 프로모터 뉴클레오티드 서열; 및 (iii) 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하고, 여기서 제 2 폴리(A) 서열은 알파바이러스에 대해 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열이고, 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0068] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 서열번호: 6에 제시된 핵산 서열을 포함하고, RNA 알파바이러스 백본 서열은 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열을 포함하고, 여기서 26S 프로모터 서열은 RNA 알파바이러스 백본에 내인성이고, 폴리(A) 서열은 RNA 알파바이러스 백본에 내인성이며; 및 (b) 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합된 카세트를 포함하고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열에 작동가능하게 연결되고, 카세트는 a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0069] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 적어도 30 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 적어도 100 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 적어도 300 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 적어도 400 μg, 적어도 500 μg, 적어도 600 μg, 적어도 700 μg, 적어도 800 μg, 적어도 900 μg, 적어도 1000 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 10-30 μg, 10-100 μg, 10-300 μg, 30-100 μg, 30-300 μg, 또는 100-300 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스 기반-발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 10-500 μg, 10-1000 μg, 30-500 μg, 30-1000 μg, 또는 500-1000 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 400 μg, 500 μg, 600 μg, 700 μg, 800 μg, 900 μg, 또는 1000 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 10 μg, 30 μg, 100 μg, 또는 300 μg 포함한다. 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은 하나 이상의 벡터를 각각 300 μg 이하 포함한다.

[0070] 일부 측면에서, LNP의 중량 대 하나 이상의 벡터의 총 중량에 대한 중량 비는 10-40 대 1이다. 일부 측면에서, LNP의 중량 대 하나 이상의 벡터의 총 중량에 대한 중량 비는 16-32 대 1이다. 일부 측면에서, LNP의 중량 대 하나 이상의 벡터의 총 중량에 대한 중량비는 약 24 대 1이다. 일부 측면에서, LNP의 중량 대 하나 이상의 벡터

의 총 중량에 대한 중량 비는 24 대 1이다.

[0071] 일부 측면에서, 하나 이상의 벡터는 1 mg/mL의 농도로 존재한다.

[0072] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'에서 3'로 식 P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$를 포함하고, 여기서 P는 제 2 프로모터 뉴클레오티드 서열을 포함하고, 여기서 a = 0 또는 1이고, N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고, L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고, L3은 3' 링커 서열을 포함하고, 여기서 d = 0 또는 1이고, G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고, G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고, U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고, X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고, Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 X에 대해 상응하는 N_c는 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 Y에 대해 상응하는 U_f는 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열이다. a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2인 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 제공되는 단일 26S 프로모터 뉴클레오티드 서열이고, 적어도 하나의 폴리아데닐화 폴리(A) 서열은 RNA 알파바이러스 백본에 의해 제공되는 적어도 100개의 연속적인 A 뉴클레오티드의 폴리(A) 서열이고, 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열 사이에 통합되고, 여기서 카세트는 26S 프로모터 뉴클레오티드 서열 및 폴리(A) 서열에 작동가능하게 연결되고, 각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고, L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, U는 각각의 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고, RNA 알파바이러스 백본은 서열번호: 6에 제시된 서열이고, 각각의 MHC는 클래스 I 에피토프-코딩 핵산 서열은 길이가 13-25개 아미노산인 폴리펩티드를 코딩한다.

[0073] 일부 측면에서, LNP는 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, PEG-기반 코트 지질, 또는 이들의 조합으로 이루어진 군으로부터 선택된 지질을 포함한다. 일부 측면에서, LNP는 이온화가능한 아미노 지질, 포스파티딜콜린, 콜레스테롤, 및 PEG-기반 코트 지질을 포함한다. 일부 측면에서, 이온화가능한 아미노 지질은 MC3-유사(디리놀레일메틸-4-디메틸아미노부티레이트) 분자를 포함한다. 일부 측면에서, LNP-캡슐화된 발현 시스템은 약 100 nm의 직경을 갖는다.

[0074] 일부 측면에서, 카세트는 하나 이상의 프로모터 뉴클레오티드 서열 및 하나 이상의 폴리(A) 서열 사이에 통합된다.

[0075] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 카세트에 작동가능하게 연결된다.

[0076] 일부 측면에서, 하나 이상의 벡터는 하나 이상의 +-가닥 RNA 벡터를 포함한다. 일부 측면에서, 하나 이상의 +-가닥 RNA 벡터는 5'7-메틸구아노신(m7g) cap을 포함한다. 일부 측면에서, 하나 이상의 +-가닥 RNA 벡터는 시험관내 전사에 의해 생성된다. 일부 측면에서, 하나 이상의 벡터는 포유동물 세포 내에서 자가-복제한다. 일부 측면에서, RNA 알파바이러스 백본은 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 베네수엘라 말 뇌염 바이러스의 적어도 하나의 뉴클레오티드 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 적어도 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩된 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열, 비구조 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자, 및 nsP4 유전자를 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열에 의해 코딩된 적어도 비구조적 단백질-매개 증폭을 위한 서열, 26S 프로모터 서열, 폴리(A) 서열을 포함한다. 일부 측면에서, 비구조적 단백질-매개 증폭을 위한 서열은 알파바이러스 5' UTR, 51-nt CSE, 24-nt

CSE, 26S 서브게놈 프로모터 서열, 19-nt CSE, 알파바이러스 3' UTR, 또는 이들의 조합으로 이루어진 군으로부터 선택된다. 일부 측면에서, RNA 알파바이러스 백본은 구조적 비리온 단백질 캡시드, E2 및 E1을 코딩하지 않는다. 일부 측면에서, 카세트는 아우라 바이러스, 포트 모건 바이러스, 베네수엘라 말 뇌염 바이러스, 로스 리버 바이러스, 셈리키 포레스트 바이러스, 신드비스 바이러스, 또는 마야로 바이러스의 뉴클레오티드 서열 내의 구조적 비리온 단백질 대신 삽입된다. 일부 측면에서, 베네수엘라 말 뇌염 바이러스는 서열번호: 3 또는 서열번호: 5의 서열을 포함한다. 일부 측면에서, 베네수엘라 말 뇌염 바이러스는 염기쌍 7544와 11175 사이의 결실을 추가로 포함하는 서열번호: 3 또는 서열번호: 5의 서열을 포함한다. 일부 측면에서, RNA 알파바이러스 백본은 서열번호: 6 또는 서열번호: 7의 서열을 포함한다. 일부 측면에서, 카세트는 서열번호: 3 또는 서열번호: 5의 서열에 제시된 염기쌍 7544와 11175 사이의 결실을 대체하기 위해 위치 7544에 삽입된다. 일부 측면에서, 카세트의 삽입은 nsP1-4 유전자 및 적어도 하나의 핵산 서열을 포함하는 폴리시스트론 RNA의 전사를 제공하고, 여기서 nsP1-4 유전자 및 적어도 하나의 핵산 서열은 분리된 열린 해독틀에 있다.

[0077] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 RNA 알파바이러스 백본에 의해 코딩되는 천연 26S 프로모터 뉴클레오티드 서열이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 외인성 RNA 프로모터이다. 일부 측면에서, 제 2 프로모터 뉴클레오티드 서열은 26S 프로모터 뉴클레오티드 서열이다. 일부 측면에서, 제 2 프로모터 뉴클레오티드 서열은 다수의 26S 프로모터 뉴클레오티드 서열을 포함하고, 여기서 각각의 26S 프로모터 뉴클레오티드 서열은 하나 이상의 분리된 열린 해독틀의 전사를 제공한다.

[0078] 일부 측면에서, 하나 이상의 벡터는 각각 크기가 적어도 300 nt이다. 일부 측면에서, 하나 이상의 벡터는 각각 크기가 적어도 1 kb이다. 일부 양상에서, 하나 이상의 벡터는 각각 크기가 2 kb이다. 일부 측면에서, 하나 이상의 벡터는 각각 크기가 5 kb 미만이다.

[0079] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 서로 직접 연결된다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된다. 일부 측면에서, 링커는 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 (1) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 잔기인 연속적인 글리신 잔기; (2) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 잔기인 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원에 플랭킹하고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 링커는 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 서열 GPGPG를 포함한다.

[0080] 일부 측면에서, 항원-코딩 핵산 서열은 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결된다. 일부 측면에서, 분리 또는 인접 서열은 적어도 하나의 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열은 위치 76에서 Gly에서 Ala로의 치환을 함유한다), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열을 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76이다.

[0081] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열를 포함하고, 임의로 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 최대 400개의 항원-코딩 핵산 서열를 포함한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩한다. 일부 측면에서, 적어도 2개의 MHC 클래스 I 에피토프는 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시된다.

[0082] 일부 측면에서, 에피토프-코딩 핵산 서열은 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8-35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩한다.

[0083] 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재한다. 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드와 구별되게 만드는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40개 아미노산 길이인 폴리펩티드 서열을 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함한다.

[0084] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열은 유도성이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열 또는 제 2 프로모터 뉴클레오티드 서열은 비-유도성이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 알파바이러스에 천연인 폴리(A) 서열을 포함한다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 하나의 핵산 서열 중 적어도 하나에 작동가능하게 연결된다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 100개의 연속적인 A 뉴클레오티드이다.

[0085] 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 획득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되며; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및; (c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 선택된 에피토프 세트의 수는 2-20이다. 일부 측면에서, 제시 모델은 (a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및 (b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열이 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면 상에 제시될 가능성 사이의 의존성을 나타낸다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함하며, 임의로 여기서 APC는 수지상 세포(DC)이다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은

제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 엑솜 또는 전사체 뉴클레오티드 시퀀싱 데이터는 감염된 조직에 대해 시퀀싱을 수행함으로써 획득된다. 일부 측면에서, 시퀀싱은 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법이다.

[0086] 일부 측면에서, ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균- 유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩핵산 서열을 포함하고; 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결된다.

[0087] 일부 측면에서, ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열, 여기서 뉴클레오티드 2 내지 36,518에는: (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고; (ii) CMV 프로모터 뉴클레오티드 서열; 및 (iii) SV40 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균- 유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 여기서 카세트는 E1 결실 내에 삽입되고 카세트는 CMV 프로모터 뉴클레오티드 서열 및 SV40 폴리(A) 서열에 작동가능하게 연결된다.

[0088] 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서 펩티드를 코딩하는 핵산 서열은 2개 이상의 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 1-10, 1-20, 1-30, 1-40, 1-50, 1-100, 1-200, 1-300, 1-400, 또는 1-500의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함하고, 임의로 여기서, 펩티드를 코딩하는 핵산 서열은 2-10, 2-20, 2-30, 2-40, 2-50, 2-100, 2-200, 2-300, 2-400, 또는 2-500의 별개의 에피토프-코딩 핵산 서열 별개의 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 펩티드를 코딩하는 핵산 서열은 감염성 질병 유기체에서 발현되는 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인으로부터 선택된 펩티드를 코딩하는 핵산 서열을 포함하고, 임의로 여기서 감염성 질병 유기체는 감염성 질병 유기체에서 발현된 단백질의 에피토프, 전장 단백질, 단백질 서브유닛, 단백질 도메인, 및 이들의 조합으로부터 선택된 펩티드를 코딩하는 2개 이상의 별개의 핵산 서열을 포함한다. 일부 측면에서, 코딩된 펩티드 또는 펩티드들은 대상체에서 발현될 때 면역 반응을 자극할 수 있으며, 임의로 면역 반응은 T 세포 반응 및/또는 B 세포 반응이다. 일부 측면에서, 감염성 질병 유기체는 중증 급성 호흡기 증후군-관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택된다.

[0089] 일부 측면에서, 에피토프-코딩 핵산 서열은 세포의 표면 상에서 MHC 클래스 I에 의해 제시되는 것으로 알려지거나 의심되는 에피토프를 코딩하고, 임의로 여기서 세포의 표면은 감염된 세포 표면이고, 임의로 여기서 세포는 대상체의 세포이다. 일부 측면에서, 세포는 병원체 감염 세포, 바이러스 감염 세포, 박테리아 감염 세포, 진균 감염 세포, 및 기생충 감염 세포로 이루어진 군으로부터 선택된 감염된 세포이다. 일부 측면에서, 바이러스 감염 세포는 HIV 감염 세포, HPV 감염 세포, SARS 감염 세포, SARS-CoV-2 감염 세포, 에볼라 감염 세포, HBV 감염 세포, 인플루엔자 감염 세포, HCV 감염 세포, CMV 감염 세포, 치쿤구니야 바이러스 감염 세포, RSV 감염 세포, 뎅기열 바이러스 감염 세포, 오르티믹소바이러스과 바이러스 감염 세포 및 결핵 감염 세포로 이루어진 군으로부터 선택된다.

[0090] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물에서 카세트의 각 요소의 정렬된 서열은 5'에서 3'로 식 P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$를 포함하고, 여기서 P는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 적어도 하나의 프로모터 서열을 포함하고, 여기서 a = 1이고, N은 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1이고, L5는 5' 링커 서열을 포함하고, 여기서 b = 0 또는 1이고, L3은 3' 링커 서열을 포

함하고, 여기서 d = 0 또는 1이고, G5는 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 e = 0 또는 1이고, G3은 GPGPG 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, 여기서 g = 0 또는 1이고, U는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열 중 하나를 포함하고, 여기서 f = 1이고, X = 1 내지 400이고, 여기서 각각의 X에 대해 상응하는 N_c는 MHC 클래스 I 에피토프-코딩 핵산 서열이고 Y = 0, 1 또는 2이고, 여기서 각각의 Y에 대해 상응하는 U_f는 MHC 클래스 II 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 X에 대해 상응하는 N_c는 별개의 MHC 클래스 I 에피토프-코딩 핵산 서열이다. 일부 측면에서, 각각의 Y에 대해 상응하는 U_f는 별개의 MHC 클래스 II 에피토프-코딩 핵산 서열이다. b = 1, d = 1, e = 1, g = 1, h = 1, X = 10, Y = 2인 일부 측면에서, P는 CMV 프로모터 서열이고, 각각의 N은 MHC 클래스 I 에피토프 7-15개 아미노산 길이를 코딩하고, L5는 MHC I 에피토프의 천연 N-말단 아미노산 서열을 코딩하는 천연 5' 링커 서열이고, 여기서 5' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, L3은 MHC I 에피토프의 천연 C-말단 아미노산 서열을 코딩하는 천연 3' 링커 서열이고, 여기서 3' 링커 서열은 길이가 적어도 3개의 아미노산인 펩티드를 코딩하고, U는 각각의 PADRE 클래스 II 서열 및 파상풍 톡소이드 MHC 클래스 II 서열이고, ChAdV 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하는 변형된 ChAdV68 서열을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; 및 임의로 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642가 결여되어 있고, 항원 카세트는 E1 결실 내에 삽입되고, 각각의 MHC 클래스 I 항원-코딩 핵산 서열은 길이가 25개 아미노산인 폴리펩티드를 코딩한다.

[0091] 일부 측면에서, 카세트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열 사이에 통합된다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 카세트에 작동가능하게 연결된다.

[0092] 일부 측면에서, ChAdV 백본은 ChAdV68 벡터 백본을 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열을 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 ChAdV68 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 아데노바이러스 E1A, E1B, E2A, E2B, E3, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택되는 적어도 하나의 유전자에 기능적 결실을 포함하고, 임의로 아데노바이러스 백본 또는 변형된 ChAdV68 서열은 아데노바이러스 게놈에 대한 참조 또는 서열번호: 1에 제시된 서열을 참조하여 (1) E1A 및 E1B; 또는 (2) E1A, E1B 및 E3에서 완전히 결실되거나 기능적으로 결실되고, 임의로 여기서 E1 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 577 내지 3403의 E1 결실을 통해 기능적으로 결실되고 및 임의로 E3 유전자는 서열번호: 1에 제시된 서열을 참조하여 적어도 뉴클레오티드 27,125 내지 31,825의 E3 결실을 통해 기능적으로 결실된다. 일부 측면에서, ChAdV68 벡터 백본은 ChAdV68 게놈에 참조하여 또는 서열번호: 1에 제시된 서열에 참조하여 하나 이상의 유전자 또는 조절 서열을 포함하고, 임의로 하나 이상의 유전자 또는 조절 서열은 침팬지 아데노바이러스 역 말단 반복부(ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5 유전자로 구성된 군으로부터 선택된다.

[0093] 일부 측면에서, ChAdV68 벡터 백본은 부분적으로 결실된 E4 유전자를 포함한다. 일부 측면에서, 부분적으로 결실된 E4 유전자는 다음을 포함한다: A. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642가 결여된 E4 유전자 서열, B. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 34,942, 뉴클레오티드 34,952 내지 35,305, 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642가 결여된E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, C. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516이 결여된 E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 나타낸 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, D. 서열번호: 1에 제시되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,979 내지 35,642가 결여된 E4 유전자 서열, 및 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, E. E4Orf2의 적어도 부분 결실, 완전히 결실된 E4Orf3 및 E4Orf4의 적어도 부분 결실의 E4 결실, F. E4Orf2의 적어도 부분 결실, E4Orf3의 적어도 부분 결실, 및 E4Orf4의 적어도 부분 결실의 E4 결실, G. E4Orf1의 적어도 부분 결실, 완전히 결실된 E4Orf2, 및 E4Orf3의 적어도 부분 결실의 E4 결실, 또는 H. E4Orf2의 적어도 부분 결실 및 E4Orf3의 적어도 부분 결실의 E4 결실. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 다음이 결여되어 있다: (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지

31,825; 및 (3) 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642; 임의로 여기서 항원 카세트는 E1 결실 내에 삽입된다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 29369에 제시된 서열을 포함하고, 임의로 항원 카세트는 E1 결실 내에 삽입된다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 다음이 결여되어 있다: A. E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; B. E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825; C. 부분적 E4 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642; D. 서열번호: 1에 제시된 서열의 뉴클레오티드 456 내지 3014; E. 서열번호: 1에 제시된 서열의 뉴클레오티드 27,816 내지 31,333; F. 서열번호: 1에 제시된 서열의 뉴클레오티드 3957 내지 10346; G. 서열번호: 1에 제시된 서열의 뉴클레오티드 21787 내지 23370; H. 서열번호: 1에 제시된 서열의 뉴클레오티드 33486 내지 36193; 또는 이들의 조합. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함하고, 여기서 뉴클레오티드 2 내지 36,518에는 다음이 결여되어 있다: (1) E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403; (2) E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825.

[0094] 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열을 포함하고, 단 서열이 서열번호: 1에 제시된 서열의 침팬지 아데노바이러스 E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택된 적어도 하나의 유전자가 완전히 결실되거나 기능적으로 결실되고, 임의로 서열은 서열번호: 1에 제시된 서열의 (1) E1A 및 E1B; (2) E1A, E1B 및 E3; 또는 (3) E1A, E1B, E3, 및 E4에서 완전히 결실되거나 기능적으로 결실된다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1의 서열로부터 수득된 유전자 또는 조절 서열을 포함하고, 임의로 여기서 유전자는 서열번호: 1에 제시된 서열의 침팬지 아데노바이러스 역 말단 반복부(ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, 및 L5 유전자로 이루어진 군으로부터 선택된다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 염기쌍 번호 577과 3403 사이 또는 염기쌍 456과 3014 사이에 하나 이상의 결실을 포함하고, 임의로 벡터는 염기쌍 27,125와 31,825 사이 또는 염기쌍 27,816과 31,333 사이에 하나 이상의 결실을 포함한다. 일부 측면에서, ChAdV68 벡터 백본은 서열번호: 1에 제시된 서열의 염기쌍 번호 3957과 10346, 염기쌍 번호 21787과 23370, 및 염기쌍 번호 33486과 36193 사이에 하나 이상의 결실을 포함한다.

[0095] 일부 측면에서, 카세트는 E1 영역, E3 영역, 및/또는 카세트의 통합을 허용하는 임의의 결실된 AdV 영역에서 ChAdV 백본에 삽입된다. 일부 측면에서, ChAdV 백본은 1세대, 2세대, 또는 헬퍼-의존성 아데노바이러스 벡터 중 하나로부터 생성된다.

[0096] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 CMV, SV40, EF-1, RSV, PGK, HSA, MCK, 및 EBV 프로모터 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 CMV 프로모터 서열이다.

[0097] 일부 측면에서, 에피토프-코딩 핵산 서열 중 적어도 하나는 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 I에 의해 제시될 수 있는 에피토프를 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열 중 적어도 하나는 발현되고 번역될 때 대상체의 세포 상에서 MHC 클래스 II에 의해 제시될 수 있는 에피토프를 코딩한다.

[0098] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 2개 이상의 항원-코딩 핵산 서열을 포함한다. 일부 측면에서, 각각의 항원-코딩 핵산 서열은 서로 직접 연결된다.

[0099] 일부 측면에서, 각각의 항원-코딩 핵산 서열은 링커를 코딩하는 핵산 서열을 갖는 별개의 항원-코딩 핵산 서열에 연결된다. 일부 측면에서, 링커는 2개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 I 에피토프-코딩 핵산 서열을 MHC 클래스 II 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 (1) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 잔기인 연속적인 글리신 잔기; (2) 길이가 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 잔기인 연속적인 알라닌 잔기; (3) 2개의 아르기닌 잔기(RR); (4) 알라닌, 알라닌, 티로신(AAY); (5) 포유동물 프로테아좀에 의해 효율적으로 처리되는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산 잔기 길이의 공통 서열; 및 (6) 기원의 동족 단백질로부터 유래된 항원에 플랭킹되고 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 2-20개의 아미노산 잔기 길이인 하나 이상의 천연 서열로 이루어진 군으로부터 선택된다. 일부 측면에서, 링커는 2개의 MHC 클래스 II 에피토프-코딩 핵산 서열을 연결하거나 또는 MHC 클래스 II 서열을 MHC 클래스 I 에피토프-코딩 핵산 서열에 연결한다. 일부 측면에서, 링커는 서열 GPGPG를 포함한다.

[0100] 일부 측면에서, 항원-코딩 핵산 서열은 항원-코딩 핵산 서열의 발현, 안정성, 세포 트래피킹, 처리 및 제시, 및/또는 면역원성을 향상시키는 분리 또는 인접 서열에 작동가능하게 또는 직접적으로 연결된다. 일부 측면에서, 분리 또는 인접 서열은 적어도 하나의 유비퀴틴 서열, 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열(예를 들어, 유비퀴틴 서열은 위치 76에서 Gly에서 Ala로의 치환을 함유한다), 면역글로불린 신호 서열(예를 들어, IgK), 주요 조직적합성 클래스 I 서열, 리소좀-관련 막 단백질(LAMP)-1, 인간 수지상 세포 리소좀-관련 막 단백질, 및 주요 조직적합성 클래스 II 서열을 포함하고; 임의로 여기서 프로테아좀 표적화를 증가시키도록 변형된 유비퀴틴 서열은 A76이다.

[0101] 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 친화도을 갖도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자에 대해 증가된 결합 안정성을 갖도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 코딩된 에피토프가 번역된 상응하는 야생형 핵산 서열에 비해 이의 상응하는 MHC 대립유전자 상에서 제시될 가능성이 증가하도록 하는 적어도 하나의 변경을 포함한다. 일부 측면에서, 적어도 하나의 변경은 점 돌연변이, 프레임시프트 돌연변이, 비-프레임시프트 돌연변이, 결실 돌연변이, 삽입 돌연변이, 스플라이스 변이체, 게놈 재배열, 또는 프로테아좀-생성 스플라이싱 항원을 포함한다.

[0102] 일부 측면에서, 에피토프-코딩 핵산 서열은 감염이 있는 것으로 알려지거나 의심되는 대상체에서 발현되는 것으로 알려지거나 의심되는 에피토프를 코딩한다. 일부 측면에서, 감염은 병원체 감염, 바이러스 감염, 박테리아 감염, 진균 감염 및 기생충 감염으로 이루어진 군으로부터 선택된다. 일부 측면에서, 바이러스 감염은 HIV 감염, HPV 감염, SARS 감염, SARS-CoV-2 감염, 에볼라 감염, HBV 감염, 인플루엔자 감염, HCV 감염, CMV 감염, 치쿤구니야 바이러스 감염, RSV 감염, 뎅기열 바이러스 감염, 오르티믹소바이러스과 바이러스 감염, 및 결핵 감염을 포함한다.

[0103] 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-10, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 항원-코딩 핵산 서열을 포함하고, 임의로 여기서 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 400개 이하의 항원-코딩 핵산 서열를 포함하고, 임의로 각각의 항원-코딩 핵산 서열은 별개의 항원-코딩 핵산 서열을 코딩한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 최대 400개의 항원-코딩 핵산 서열를 포함한다. 일부 측면에서, 적어도 하나의 항원-코딩 핵산 서열은 적어도 2-400개의 항원-코딩 핵산 서열을 포함하고, 여기서 적어도 2개의 항원-코딩 핵산 서열은 세포 표면 상에서 MHC 클래스 I에 의해 제시되는 에피토프 서열 또는 이의 부분을 코딩한다. 일부 측면에서, 적어도 2개의 MHC 클래스 I 에피토프는 감염된 세포 표면 상에서 MHC 클래스 I에 의해 제시된다.

[0104] 일부 측면에서, 에피토프-코딩 핵산 서열은 적어도 하나의 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 8-35개의 아미노산 길이, 임의로 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 또는 35개의 아미노산 길이의 폴리펩티드 서열을 코딩한다.

[0105] 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재한다. 일부 측면에서, 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 코딩된 에피토프 서열을 야생형 핵산 서열에 의해 코딩된 상응하는 펩티드와 구별되게 만드는 적어도 하나의 변경을 포함하는 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 각각의 항원-코딩 핵산 서열은 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 20-40개 아미노산 길이인 폴리펩티드 서열을 코딩한다. 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열이 존재하고, 여기서 적어도 하나의 MHC 클래스 II 에피토프-코딩 핵산 서열은 적어도 하나의 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열을 포함하고, 임의로 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함한다.

[0106] 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 유도성이다. 일부 측면에서, 적어도 하나의 프로모터 뉴클레오티드 서열은 비-유도성이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 소 성장 호르몬(BGH) SV40 폴리A 서열을 포함한다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 20, 적어도 30, 적어도 40,

적어도 50, 적어도 60, 적어도 70, 적어도 80, 또는 적어도 90개의 연속적인 A 뉴클레오티드이다. 일부 측면에서, 적어도 하나의 폴리(A) 서열은 적어도 100개의 연속적인 A 뉴클레오티드이다.

[0107] 일부 측면에서, 카세트는 인트론 서열, 우드척 간염 바이러스 전사후 조절 요소(WPRE) 서열, 내부 리보솜 진입 서열(IRES) 서열, 2A 자가 절단 펩티드 서열을 코딩하는 뉴클레오티드 서열, 푸린 절단 부위를 코딩하는 뉴클레오티드 서열, 또는 적어도 하나의 항원-코딩 핵산 서열 중 적어도 하나에 작동가능하게 연결된 mRNA의 핵 엑스포트, 안정성 또는 번역 효율을 향상시키는 것으로 알려진 5' 또는 3' 비-코딩영역의 서열 중 적어도 하나를 추가로 포함한다. 일부 측면에서, 카세트는 녹색 형광 단백질(GFP), GFP 변이체, 분비된 알칼리성 포스파타제, 루시퍼라제, 루시퍼라제 변이체, 또는 검출가능한 펩티드 또는 에피토프를 포함하나 이에 제한되지 않는 리포터 유전자를 추가로 포함한다. 일부 측면에서, 검출가능한 펩티드 또는 에피토프는 HA 태그, 플래그(Flag) 태그, His-태그, 또는 V5 태그로 이루어진 군으로부터 선택된다.

[0108] 일부 측면에서, 하나 이상의 벡터는 적어도 하나의 면역 조절자를 코딩하는 하나 이상의 핵산 서열을 추가로 포함한다. 일부 측면에서, 면역 조절자는 항-CTLA4 항체 또는 이의 항원-결합 단편, 항-PD-1 항체 또는 이의 항원-결합 단편, 항-PD-L1 항체 또는 이의 항원-결합 단편, 항-4-1BB 항체 또는 이의 항원 결합 단편, 또는 항-OX-40 항체 또는 이의 항원 결합 단편이다. 일부 측면에서, 항체 또는 이의 항원-결합 단편은 단일 특이적 또는 다중 특이성으로서 함께 연결된 Fab 단편, Fab' 단편, 단일 사슬 Fv(scFv), 단일 도메인 항체(sdAb)(예를 들어,낙타류 항체 도메인), 또는 전장 단일-사슬 항체(예를 들어, 가요성 링커에 의해 연결된 중쇄 및 경쇄를 갖는 전장 IgG)이다. 일부 측면에서, 항체의 중쇄 및 경쇄 서열은 2A 또는 IRES와 같은 자가-절단 서열에 의해 분리된 연속 서열이거나; 또는 항체의 중쇄 및 경쇄 서열은 연속적인 글리신 잔기와 같은 가요성 링커에 의해 연결된다. 일부 측면에서, 면역 조절자는 사이토카인이다. 일부 측면에서, 사이토카인은 적어도 하나의 IL-2, IL-7, IL-12, IL-15 또는 IL-21 또는 각각의 이의 변이체이다.

[0109] 일부 측면에서, 에피토프-코딩 핵산 서열은 MHC 클래스 I 에피토프-코딩 핵산 서열을 포함하고, 여기서 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터의 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 획득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 각각의 MHC 클래스 I 에피토프-코딩 핵산 서열은 하기 단계를 수행함으로써 선택된다: (a) 감염성 질병 유기체로부터 엑솜, 전사체, 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 데이터는 에피토프 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되며; (b) 각각의 에피토프가 감염된 세포 표면 상의 MHC 대립유전자 중 하나 이상에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각 에피토프의 펩티드 서열을 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 (c) 적어도 20개의 MHC 클래스 I 에피토프-코딩 핵산 서열을 생성하는 데 사용되는 선택된 에피토프 세트를 생성하기 위해 수치적 가능성의 세트에 기초하여 에피토프 세트의 서브세트를 선택하는 단계. 일부 측면에서, 선택된 에피토프 세트의 수는 2-20이다. 일부 측면에서, 제시 모델은 (a) 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및 (b) 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열이 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면에 제시될 가능성 사이의 의존성을 나타낸다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 증가된 가능성을 갖는 에피토프를 선택하는 것을 포함하며, 임의로 여기서 APC는 수지상 세포(DC)이다. 일부 측면에서, 선택된 에피토프의 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 선택된 에피토프 세트를 선택하는 것은 제시 모델에 기초하여 선택되지 않은 에피토프에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 에피토프를 선택하는 것을 포함한다. 일부 측면에서, 엑솜 또는 전사체 뉴클레

오티드 시퀀싱 데이터는 감염된 조직에 대해 시퀀싱을 수행함으로써 획득된다. 일부 측면에서, 시퀀싱은 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법이다.

[0110] 일부 측면에서, 카세트는 카세트의 인접 서열에 의해 형성된 접합 에피토프 서열을 포함한다. 일부 측면에서, 적어도 하나 또는 각각의 접합 에피토프 서열은 MHC에 대해 500 nM 초과의 친화도를 갖는다. 일부 측면에서, 각각의 접합 에피토프 서열은 비-자기이다.

[0111] 일부 측면에서, 카세트는 번역된 야생형 핵산 서열을 포함하는 비-치료적 MHC 클래스 I 또는 클래스 II 에피토프 핵산 서열을 코딩하지 않으며, 여기서 비-치료적 에피토프는 대상체의 MHC 대립유전자 상에 표시될 것으로 예측된다. 일부 측면에서, 비-치료적 예측 MHC 클래스 I 또는 클래스 II 에피토프 서열은 카세트에서 인접한 서열에 의해 형성된 접합 에피토프 서열이다. 일부 측면에서, 예측은 비-치료적 에피토프의 서열을 제시 모델에 입력함으로써 생성된 제시 가능성에 기초한다. 일부 측면에서, 카세트 내의 항원-코딩 핵산 서열의 순서는 (a) 항원-코딩 핵산 서열의 상이한 순서에 상응하는 후보 카세트 서열 세트를 생성하는 단계; (b) 각각의 후보 카세트 서열에 대해 후보 카세트 서열에서 비-치료적 에피토프의 제시에 기초하여 제시 점수를 결정하는 단계; 및 (c) 백신에 대한 카세트 서열로서 미리 결정된 임계값 미만의 제시 점수와 관련된 후보 카세트 서열을 선택하는 단계를 포함하는 일련의 단계에 의해 결정된다.

[0112] 일부 측면에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물은 약제학적으로 허용되는 담체를 포함하는 약제학적 조성물로 제형화된다.

[0113] 일부 측면에서, 하나 이상의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래된다. 일부 측면에서, 각각의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래된다. 일부 측면에서, 하나 이상의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래되지 않는다. 일부 측면에서, 각각의 에피토프-코딩 핵산 서열은 대상체의 감염 또는 감염된 세포로부터 유래되지 않는다.

[0114] 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 것을 포함하여, 대상체에서 면역 반응을 자극하는 방법이 또한 본원에 제공되며, 여기서 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물은: (A) 자가-복제 알파바이러스-기반 발현 시스템, 여기서 자가-복제 알파바이러스-기반 발현 시스템은 하나 이상의 벡터를 포함하고, 하나 이상의 벡터는 (a) RNA 알파바이러스 백본, 여기서 RNA 알파바이러스 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되고, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열; (ii) 임의로, 적어도 하나의 항원-코딩 핵산 서열에 작동가능하게 연결된 제 2 프로모터 뉴클레오티드 서열; 및 (iii) 임의로, 적어도 하나의 제 2 폴리(A) 서열을 포함하고, 여기서 제 2 폴리(A) 서열은 알파바이러스에 대한 천연 폴리(A) 서열 또는 외인성 폴리(A) 서열이고, 및 (B) 지질-나노입자(LNP)를 포함하고, 여기서 LNP는 자가-복제 알파바이러스-기반 발현 시스템을 캡슐화한다.

[0115] 침팬지 아데노바이러스(ChAdV)-기반 발현 시스템의 전달을 위한 조성물을 대상체에게 투여하는 것을 포함하는, 대상체에서 면역 반응을 자극하는 방법이 또한 본원에 제공되며, 여기서 ChAdV-기반 발현 시스템의 전달을 위한 조성물은 다음을 포함한다: ChAdV-기반 발현 시스템, 여기서 ChAdV-기반 발현 시스템은 ChAdV 벡터를 포함하는 바이러스 입자를 포함하고, 여기서 ChAdV 벡터는 (a) ChAdV 백본, 여기서 ChAdV 백본은 (i) 적어도 하나의 프로모터 뉴클레오티드 서열, 및 (ii) 적어도 하나의 폴리아데닐화(폴리(A)) 서열을 포함하고; 및 (b) 카세트를 포함하고, 여기서 카세트는 (i) a. 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드로 이루어진 군으로부터 선택되는 감염성 질병 유기체 펩티드를 코딩하는 핵산 서열, 여기서 감염성 질병 유기체는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵으로 이루어진 군으로부터 선택되고, b. 임의로 5' 링커 서열, 및 c. 임의로 3' 링커 서열을 포함하는 적어도 하나의 항원-코딩 핵산 서열을 포함하고; 여기서 카세

트는 적어도 하나의 프로모터 뉴클레오티드 서열 및 적어도 하나의 폴리(A) 서열에 작동가능하게 연결된다.

[0116] 일부 측면에서, 자가-복제 알파바이러스-기반 발현 시스템 또는 ChAdV-기반 발현 시스템은 상기 청구항 중 어느 하나의 특징 중 임의의 것을 포함한다.

[0117] 상기 청구항 중 어느 하나의 방법에서, ChAdV-기반 발현 시스템의 전달을 위한 조성물의 카세트는 자가-복제 알파바이러스-기반 발현 시스템의 전달을 위한 조성물의 카세트와 동일하다.

도면의 간단한 설명

[0118] 본 발명의 이들 및 다른 특징, 측면 및 이점은 다음의 설명 및 첨부 도면과 관련하여 더 잘 이해될 것이다.

도 1은 ELISpot을 이용하여 측정한 항원-특이적 세포 면역 반응을 나타낸다. 6개의 다른 mamu A01 제한된 에피토프에 대한 항원-특이적 IFN-감마 생산은 ELISpot를 사용하여 첫 번째 부스트 면역접종 후 1, 2, 3, 4, 5, 6, 8, 9 또는 10주(군 당 6마리의 붉은 털 원숭이)에 VEE-MAG25mer srRNA-LNP1(30 μg)(**도 1a**), VEE-MAG25mer srRNA-LNP1(100 μg)(**도 1b**), 또는 VEE-MAG25mer srRNA-LNP2(100 μg)(**도 1c**) 상동 프라임/부스트 또는 ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA 이종 프라임/부스트 군(**도 1d**)에 대해 PBMC에서 측정된다. 결과는 누적 막대 그래프 형식의 각 에피토프에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)로 표시된다. 각 동물에 대한 값은 채혈 전(0주)의 수준으로 정규화되었다.

도 2는 ELISpot을 이용하여 측정한 항원-특이적 세포 면역 반응을 나타낸다. 6개의 상이한 mamu A01 제한된 에피토프에 대한 항원-특이적 IFN-감마 생산은 ELISpot을 사용하여 면역화 전, ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA 이종 프라임/부스트 요법으로 면역화 이후 및 초기 면역화 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 또는 24주 후 PBMC에서 측정되었다. 결과는 누적 막대 그래프 형식에서 각 에피토프(군 당 6마리의 붉은털 원숭이)에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)로 표시된다.

도 3은 ELISpot을 이용하여 측정한 항원-특이적 세포 면역 반응을 나타낸다. 6개의 다른 mamu A01 제한된 에피토프에 대한 항원-특이적 IFN-감마 생산은 ELISpot을 사용하여 면역화 전에, VEE-MAG25mer srRNA LNP2 상동 프라임/부스트 요법으로 면역화 이후 및 초기 면역화 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 또는 15주 후 PBMC에서 측정되었다. 결과는 누적 막대 그래프 형식에서 각 에피토프(군당 6마리의 붉은털 원숭이)에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)로 표시된다.

도 4는 ELISpot을 이용하여 측정한 항원-특이적 세포 면역 반응을 나타낸다. 6개의 다른 mamu A01 제한된 에피토프에 대한 항원-특이적 IFN-감마 생산은 ELISpot을 사용하여 면역화 전에, VEE-MAG25mer srRNA LNP1 상동 프라임/부스트 요법으로 면역화 이후 및 초기 면역화 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 또는 15주 후 PBMC에서 측정되었다. 결과는 누적 막대 그래프 형식에서 각 에피토프(군당 6마리의 붉은털 원숭이)에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)로 표시된다.

도 5는 붉은털 원숭이에서 항원-카세트 함유 벡터의 면역원성을 평가하기 위해 사용된 백신접종 전략을 예시한다. 삼각형은 0주 및 32주에 chAd68 백신접종(1e12 vp/동물)을 나타낸다. 원은 0, 4, 12, 20, 28 및 32주에 알파바이러스 백신접종을 나타낸다. 사각형은 항-CTLA4 항체의 투여를 나타낸다.

도 6은 chAd-MAG 단독으로 투여된 붉은털 원숭이(군 4)에서 CD8+ 항-에피토프 반응의 시간 경과를 보여준다. 평균 SFC/1e6 비장세포가 표시된다.

도 7은 chAd-MAG + 항-CTLA4 항체(이필리무맙)가 IV로 투여된 붉은털 원숭이에서 CD8+ 항-에피토프 반응의 시간 경과를 보여준다(군 5). 평균 SFC/1e6 비장세포가 표시된다.

도 8은 chAd-MAG + 항-CTLA4 항체(이필리무맙)가 SC로 투여된 붉은털 원숭이에서 CD8+ 항-에피토프 반응의 시간 경과를 보여준다(군 6). 평균 SFC/1e6 비장세포가 표시된다.

도 9는 ELISpot에 의해 측정된 ChAdV68/samRNA 백신 프로토콜에 의해 생성된 항원-특이적 기억 반응을 나타낸다. 결과는 개별 점 플롯으로 표시되며 각 점은 단일 동물을 나타낸다. 면역화 전 기준선(왼쪽 패널) 및 프라임 후 18개월에서의 기억 반응(오른쪽 패널)이 표시된다.

도 10은 조합 사량체 염색 및 CD45RA/CCR7 공동-염색을 사용한 유세포 분석에 의한 항원-특이적 CD8+ T-세포의 기억 세포 표현형을 나타낸다.

도 11은 연구 18개월에 4개의 Mamu-A*01 사량체+ CD8+ T-세포 집단의 합 내에서 기억 세포 유형의 분포를 보여준다. 기억 세포는 다음과 같이 특성화되었다: CD45RA+CCR7+=나이브, CD45RA+CCR7-=효과기 (Teff), CD45RA-CCR7+=중앙 기억(Tcm), CD45RA-CCR7-=효과기 기억(Tem).

도 12a는 5e10 vp/동물로 투여한 후 인플루엔자 A H1N1 균주 CA/7/09에 대한 HAI 반응을 보여준다. HAI 5는 탐지 한계이다. HAI 40은 치료 임계값이다.

도 12b는 (b) 1e9 vp/동물로 투여한 후 인플루엔자 A H1N1 균주 CA/7/09에 대한 HAI 반응을 보여준다. HAI 5는 탐지 한계이다. HAI 40은 치료 임계값이다.

도 13a는 5e10 vp/동물로 투여한 후 인플루엔자 A H7N9 균주 Anhui/1/13에 대한 HAI 반응을 보여준다. HAI 5는 탐지 한계이다. HAI 40은 치료 임계값이다. 백신은 Anhui/1/13에 대한 것이지만 HAI 검정에 사용된 균주는 PR8 인플루엔자 균주에서 밀접하게 관련된 A/Shanghai/2/13 HA이다.

도 13b는 1e9 vp/동물로 투여한 후 인플루엔자 A H7N9 균주 Anhui/1/13에 대한 HAI 반응을 보여준다. HAI 5는 탐지 한계이다. HAI 40은 치료 임계값이다. 백신은 Anhui/1/13에 대한 것이지만 HAI 검정에 사용된 균주는 PR8 인플루엔자 균주에서 밀접하게 관련된 A/Shanghai/2/13 HA이다.

도 14a는 ChAd 프라임 2주 후 각 마우스(n = 6/군)에 대한 항-HA T-세포 반응을 나타낸다. 비장세포는 HA 항원에 걸쳐 있는 6개의 중첩 펩티드 풀을 사용하여 수행된 밤새 IFNγ ELISpot을 생체외에서 분리하였다. SFU/1e6 비장세포, 6개 풀에 대한 응답이 누적된다.

도 14b는 SAM 프라임 2주 후 각 마우스(n = 6/군)에 대한 항-HA T-세포 반응을 나타낸다. 비장세포는 HA 항원에 걸쳐 있는 6개의 중첩 펩티드 풀을 사용하여 수행된 밤새 IFNγ ELISpot을 생체외에서 분리하였다. SFU/1e6 비장세포, 6개 풀에 대한 응답이 누적된다.

도 14c는 SAM-HA(ChAd 프라임 후 10주)(c)로 부스트 후 2주에 각각의 마우스(n = 6/군)에 대한 항-HA T-세포 반응을 보여준다. 비장세포는 HA 항원에 걸쳐 있는 6개의 중첩 펩티드 풀을 사용하여 수행된 밤새 IFNγ ELISpot을 생체외에서 분리하였다. SFU/1e6 비장세포, 6개 풀에 대한 응답이 누적된다.

도 15a는 chAd 프라임/SAM 부스트 후 T-세포 반응을 보여준다(부스트 후 2주, 프라임 후 10주). 생체외 자극 5시간 후 ICS에 의해 측정된 IFNg, TNFa 및 IL2를 발현하는 CD8+의 백분율(2개의 중첩 펩티드 풀의 합). 상자 및 수염 및 중앙값, IQR 및 범위.

도 15b는 chAd 프라임/SAM 부스트 후 T-세포 반응을 보여준다(부스트 후 2주, 프라임 후 10주). 생체 외 자극 5시간 후 ICS에 의해 측정된 IFNg, TNFa 및 IL2를 발현하는 CD4+ 세포의 백분율(2개의 중첩 펩티드 풀의 합). 상자 및 수염 및 중앙값, IQR 및 범위

발명을 실시하기 위한 구체적인 내용

[0119] **I. 정의**

[0120] 일반적으로, 청구범위 및 명세서에서 사용된 용어는 당업자가 이해하는 명백한 의미를 갖는 것으로 해석되도록 의도된다. 추가 명확성을 제공하기 위해 특정 용어가 아래에 정의되어 있다. 일반 의미와 제공된 정의가 충돌하는 경우 제공된 정의를 사용한다.

[0121] 본원에 사용된 용어 "항원"은 면역 반응을 유도하는 물질이다. 항원은 신생항원일 수 있다. 항원은 특정 집단, 예를 들어 감염성 질병에 감염되거나 감염될 위험이 있는 환자의 특정 집단 사이에서 발견되는 항원인 "공유 항원(shared antigen)"일 수 있다.

[0122] 본원에 사용된 항원으로서 용어 "항원-기반 백신"은 하나 이상의 항원, 예를 들어, 다수의 항원을 기반으로 하는 백신 조성물이다. 백신은 뉴클레오티드-기반(예를 들어, 바이러스 기반, RNA 기반 또는 DNA 기반), 단백질-기반(예를 들어, 펩티드 기반), 또는 이들의 조합일 수 있다.

[0123] 본원에 사용된 용어 "후보 항원"은 항원을 나타낼 수 있는 서열을 발생시키는 돌연변이 또는 기타 이상이다.

[0124] 본원에 사용된 용어 "코딩 영역"은 단백질을 코딩하는 유전자 부분(들)입니다.

[0125] 본원에 사용된 용어 "코딩 돌연변이"는 코딩 영역에서 발생하는 돌연변이이다.

[0126] 본원에서 사용된 용어 "ORF"는 열린 해독틀(open reading frame)을 의미한다.

[0127] 본원에 사용된 용어 "미스센스 돌연변이"는 하나의 아미노산에서 다른 아미노산으로의 치환을 야기하는 돌연변이이다.

[0128] 본원에 사용된 용어 "넌센스 돌연변이"는 아미노산에서 종결 코돈으로의 치환을 야기하거나 정규 개시 코돈의 제거를 야기하는 돌연변이이다.

[0129] 본원에 사용된 용어 "프레임시프트 돌연변이"는 단백질의 프레임에 변화를 야기하는 돌연변이이다.

[0130] 본원에 사용된 용어 "인델"은 하나 이상의 핵산의 삽입 또는 결실이다.

[0131] 본원에 사용된 용어 퍼센트 "동일성"은 아래에 설명된 서열 비교 알고리즘(예를 들어, BLASTP 및 BLASTN 또는 당업자가 사용할 수 있는 기타 알고리즘) 중 하나를 사용하여 또는 육안 검사에 의해 측정된 최대 일치성을 위해 비교 및 정렬될 때, 2개 이상의 핵산 또는 폴리펩티드 서열의 맥락에서 동일한 뉴클레오티드 또는 아미노산 잔기의 특정 백분율을 갖는 2개 이상의 서열 또는 하위서열을 지칭한다. 적용에 따라, 퍼센트 "동일성"은 비교되는 서열의 영역, 예를 들어 기능적 도메인에 걸쳐 존재할 수 있거나, 대안적으로 비교될 2개의 서열의 전체 길이에 걸쳐 존재할 수 있다.

[0132] 서열 비교를 위해, 전형적으로 하나의 서열은 테스트 서열이 비교되는 참조 서열로서 작용한다. 서열 비교 알고리즘을 사용하는 경우, 테스트 서열 및 참조 서열을 컴퓨터에 입력하고 필요에 따라 하위서열 좌표를 지정하고 서열 알고리즘 프로그램 매개변수를 지정한다. 그런 다음 서열 비교 알고리즘은 지정된 프로그램 매개변수를 기반으로 참조 서열에 대한 테스트 서열(들)의 서열 동일성 백분율을 계산한다. 대안적으로, 서열 유사성 또는 비유사성은 특정 뉴클레오티드의 조합된 존재 또는 부재, 또는 번역된 서열의 경우 선택된 서열 위치(예를 들어, 서열 모티프)의 아미노산에 의해 확립될 수 있다.

[0133] 비교를 위한 서열의 최적 정렬은, 예를 들어 스미스 및 워터만(Smith & Waterman, Adv. Appl. Math. 2:482 (1981))의 로컬 상동성 알고리즘에 의해, 니들만 및 분쉬(Needleman & Wunsch, J. Mol. Biol. 48:443 (1970))의 상동성 알고리즘에 의해, 피어슨 및 립만(Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988))의 유사성 방법의 연구에 의해, 이러한 알고리즘의 컴퓨터 구현(GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.)에 의해, 또는 육안 검사(Ausubel 등, infra)에 의해 수행될 수 있다.

[0134] 퍼센트 서열 동일성 및 서열 유사성을 결정하기에 적합한 알고리즘의 한 예는 BLAST 알고리즘이며, 이는 문헌(Altschul 등, J. Mol. Biol. 215:403-410 (1990))에 기술되어 있다. BLAST 분석을 수행하기 위한 소프트웨어는 국립 생명 공학 정보 센터를 통해 공개적으로 제공된다.

[0135] 본원에 사용된 용어 "논스톱 또는 리드-스루(read-through)"는 자연 종결 코돈의 제거를 야기하는 돌연변이이다.

[0136] 본원에 사용된 용어 "에피토프"는 항체 또는 T 세포 수용체에 의해 전형적으로 결합된 항원의 특정 부분이다.

[0137] 본원에 사용된 용어 "면역원성"은, 예를 들어 T 세포, B 세포, 또는 둘 다를 통해 면역 반응을 유도하는 능력이다.

[0138] 본원에 사용된 용어 "HLA 결합 친화도" "MHC 결합 친화도"는 특정 항원과 특정 MHC 대립유전자 사이의 결합 친화도를 의미한다.

[0139] 본원에서 사용되는 용어 "베이트(bait)"는 샘플로부터 DNA 또는 RNA의 특정 서열을 풍부하게 하는 데 사용되는 핵산 프로브이다.

[0140] 본원에 사용된 용어 "변이체"는 대상체의 핵산 및 대조군으로 사용된 기준 인간 게놈 간의 차이이다.

[0141] 본원에서 사용된 용어 "변이체 호출"은 일반적으로 시퀀싱으로부터 변이체의 존재에 대한 알고리즘 결정이다.

[0142] 본원에 사용된 용어 "다형성"은 생식계열 변이체, 즉 개체의 모든 DNA-보유 세포에서 발견되는 변이체이다.

[0143] 본원에 사용된 용어 "체세포 변이체"는 개체의 비-생식계열 세포에서 발생하는 변이체이다.

[0144] 본원에 사용된 용어 "대립유전자"는 유전자의 버전 또는 유전자 서열의 버전 또는 단백질의 버전이다.

[0145] 본원에 사용된 용어 "HLA 유형"은 HLA 유전자 대립유전자의 보체이다.

[0146] 본원에 사용된 용어 "넌센스-매개 붕괴" 또는 "NMD"는 조기 종결 코돈으로 인한 세포에 의한 mRNA의 분해이다.

[0147] 본원에 사용된 용어 "엑솜"은 단백질을 코딩하는 게놈의 서브세트다. 엑솜은 게놈의 집합적 엑손일 수 있다.

[0148] 본원에서 사용된 용어 "로지스틱 회귀"는 종속 변수가 1일 확률의 로짓이 종속 변수의 선형 함수로 모델링되는 통계로부터의 이진 데이터에 대한 회귀 모델이다.

[0149] 본원에서 사용된 용어 "신경망"은 확률적 경사 하강법 및 역전파를 통해 일반적으로 훈련된 요소별 비선형성이 뒤따르는 여러 계층의 선형 변환으로 구성된 분류 또는 회귀를 위한 기계 학습 모델이다.

[0150] 본원에 사용된 용어 "프로테옴"은 세포, 세포 그룹 또는 개체에 의해 발현 및/또는 번역된 모든 단백질의 세트이다.

[0151] 본원에 사용된 용어 "펩티돔"는 세포 표면 상에서 MHC-I 또는 MHC-II에 의해 제시되는 모든 펩티드의 세트이다. 펩티돔은 세포 또는 세포의 집합체(예를 들어, 감염성 질병 펩티돔, 감염성 질병에 감염된 모든 세포의 펩티돔의 결합을 의미)의 특성을 나타낼 수 있다.

[0152] 본원에 사용된 용어 "ELISpot"은 효소-결합 면역흡착 스폿 검정을 의미하며, 이는 인간 및 동물의 면역 반응을 모니터링하기 위한 일반적인 방법이다.

[0153] 본원에 사용된 용어 "덱스트라머"는 유세포 분석에서 항원-특이적 T-세포 염색에 사용되는 덱스트란-기반 펩티드-MHC 다량체이다.

[0154] 본원에 사용된 용어 "내성 또는 면역 내성"은 하나 이상의 항원, 예를 들어, 자가-항원에 대한 면역 무반응의 상태이다.

[0155] 본원에 사용된 용어 "중추 내성"은 자가-반응성 T-세포 클론을 결실시키거나 자가-반응성 T-세포 클론을 촉진하여 면역억제성 조절 T-세포(Treg)로 분화되도록 촉진함으로써 흉선에서 영향을 받는 내성이다.

[0156] 본원에 사용된 용어 "말초 내성"은 중추 내성에서 생존하는 자가-반응성 T-세포를 하향조절 또는 무반응화하거나(anergizing) 또는 이들 T 세포가 Treg로 분화하도록 촉진함으로써 말초에서 영향을 받는 내성이다.

[0157] 용어 "샘플"은 정맥천자, 배설, 사정, 마사지, 생검, 바늘 흡인, 세척 샘플, 긁힘, 외과적 절개, 또는 개입 또는 당업계에 공지된 기타 수단을 포함하는 수단을 통해 대상체로부터 채취한 단일 세포 또는 다중 세포 또는 세포의 단편 또는 체액의 분취량을 포함할 수 있다.

[0158] 용어 "대상체"는 생체내, 생체외 또는 시험관내, 남성 또는 여성이든지 간에 인간 또는 비인간의 세포, 조직 또는 유기체를 포함한다. 대상체라는 용어는 인간을 포함한 포유동물을 포함한다.

[0159] 용어 "포유류"는 인간 및 비인간 둘 모두를 포함하며 인간, 비인간 영장류, 개과, 고양이과, 뮤린, 소, 말 및 돼지를 포함하지만 이에 제한되지는 않는다.

[0160] 용어 "임상 인자"는 대상체의 상태, 예를 들어 질병 활성 또는 중증도의 척도를 지칭한다. "임상 요인"은 비-샘플 마커 및/또는 이에 제한되지 않는 연령 및 성별과 같은 대상체의 기타 특성을 포함하여 대상체의 건강 상태에 대한 모든 마커를 포함한다. 임상 요인은 결정된 조건 하에서 대상체 또는 대상체로부터 샘플(또는 샘플 집단)의 평가에서 얻을 수 있는 점수, 값 또는 값 세트일 수 있다. 임상 인자는 또한 마커 및/또는 유전자 발현 대리와 같은 다른 매개변수에 의해 예측될 수 있다. 임상 요인에는 감염 유형, 감염 하위유형 및 흡연 이력이 포함될 수 있다.

[0161] 용어 "감염으로부터 유래된 항원-코딩 핵산 서열"은 감염된 세포 또는 감염성 질병 유기체, 예를 들어, RT-PCR을 통해 감염성 질병 유기체로부터 수득된 핵산 서열; 또는 감염된 세포 또는 감염성 질병 유기체를 시퀀싱한 다음, 예를 들어 당업계에 공지된 다양한 합성 또는 PCR-기반 방법을 통해 시퀀싱 데이터를 사용하여 핵산 서열을 합성함으로써 얻은 서열 데이터을 지칭한다. 유래된 서열은 상응하는 천연 감염성 질병 유기체 핵산 서열과 동일한 폴리펩티드 서열을 코딩하는 서열-최적화된 핵산 서열 변이체(예를 들어, 코돈-최적화 및/또는 발현에 대해 최적화된)와 같은 핵산 서열 변이체를 포함할 수 있다. 유도된 서열은 천연 감염성 질병 유기체 폴리펩티드 서열에 비해 하나 이상의(예를 들어, 1, 2, 3, 4, 또는 5개) 돌연변이를 갖는 변형된 감염성 질병 유기체 폴리펩티드 서열을 코딩하는 핵산 서열 변이체를 포함할 수 있다. 예를 들어, 변형된 폴리펩티드 서열은 감염성 질병 유기체 단백질의 천연 폴리펩티드 서열에 비해 하나 이상의 미스센스 돌연변이를 가질 수 있다.

[0162] 용어 "알파바이러스"는 토가비리과(*Togaviridae*)의 구성원을 말하며, 양성-센스 단일-가닥 RNA 바이러스이다. 알파바이러스는 일반적으로 신드비스(Sindbis), 로스 리버(Ross River), 마야로(Mayaro), 치쿤구니야(Chikungunya) 및 셈리키 포레스트(Semliki Forest) 바이러스와 같은 구세계 또는 동부 말 뇌염, 아우라(Aura), 포트 모건(Fort Morgan) 또는 베네수엘라 말 뇌염 및 그 파생 균주 TC-83와 같은 신세계로 분류된다. 알파바이러스는 일반적으로 자가-복제 RNA 바이러스이다.

[0163] 용어 "알파바이러스 백본"은 바이러스 게놈의 자가-복제를 허용하는 알파바이러스의 최소 서열(들)을 의미한다. 최소 서열은 비구조적 단백질-매개 증폭을 위한 보존된 서열, 비구조적 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자, nsP4 유전자, 및 폴리A 서열 뿐만 아니라 26S 프로모터 요소를 포함하는 서브게놈 바이러스 RNA의 발현을 위한 서열을 포함할 수 있다.

[0164] "비구조적 단백질-매개 증폭을 위한 서열"이라는 용어는 당업자에게 잘 알려진 알파바이러스 보존 서열 요소(CSE)를 포함한다. CSE에는 알파바이러스 5' UTR, 51-nt CSE, 24-nt CSE 또는 기타 26S 서브게놈 프로모터 서열, 19-nt CSE 및 알파바이러스 3' UTR이 포함되지만 이에 제한되지 않는다.

[0165] 용어 "RNA 중합효소"는 DNA 주형으로부터 RNA 폴리뉴클레오티드의 생산을 촉매하는 중합효소를 포함한다. RNA 중합효소는 T3, T7 및 SP6을 포함하는 박테리오파지 유래 중합효소를 포함하지만 이에 제한되지는 않는다.

[0166] 용어 "지질"은 소수성 및/또는 양친매성 분자를 포함한다. 지질은 양이온성, 음이온성 또는 중성일 수 있다. 지질은 합성 또는 천연 유래일 수 있으며 경우에 따라 생분해성이다. 지질은 콜레스테롤, 인지질, 폴리에틸렌글리콜(PEG) 접합체(PEG화 지질), 왁스, 오일, 글리세리드, 지방 및 지용성 비타민을 포함하나 이에 제한되지 않는 지질 접합체를 포함할 수 있다. 지질은 또한 디리놀레일메틸-4-디메틸아미노부티레이트(MC3) 및 MC3-유사 분자를 포함할 수 있다.

[0167] 용어 "지질 나노입자" 또는 "LNP"는 리포솜으로도 지칭되는 수성 내부를 둘러싸는 지질 함유 막을 사용하여 형성된 소포 유사 구조를 포함한다. 지질 나노입자는 계면활성제에 의해 안정화된 고체 지질 코어를 갖는 지질-기반 조성물을 포함한다. 코어 지질은 지방산, 아실글리세롤, 왁스 및 이러한 계면활성제의 혼합물일 수 있다. 인지질, 스핑고미엘린, 담즙염(나트륨 타우로콜레이트) 및 스테롤(콜레스테롤)과 같은 생물학적 막 지질이 안정화제로 사용될 수 있다. 지질 나노입자는 하나 이상의 양이온성, 음이온성 또는 중성 지질의 정의된 비율을 포함하지만 이에 제한되지 않는 상이한 지질 분자의 정의된 비율을 사용하여 형성될 수 있다. 지질 나노입자는 외부-막 껍질 내에서 분자를 캡슐화할 수 있고 후속적으로 표적 세포와 접촉하여 캡슐화된 분자를 숙주 세포 세포질에 전달할 수 있다. 지질 나노입자는 표면을 포함하여 비-지질 분자로 변형되거나 기능화될 수 있다. 지질 나노입자는 단일층(단층) 또는 다중층(다층)일 수 있다. 지질 나노입자는 핵산과 복합체화될 수 있다. 단층 지질 나노입자는 핵산과 복합체화될 수 있으며, 여기서 핵산은 수성 내부에 있다. 다층 지질 나노입자는 핵산과 복합체화될 수 있으며, 여기서 핵산은 수성 내부에 있거나 그 사이에 형성되거나 샌드위치된다.

[0168] 약어: MHC: 주요 조직적합성 복합체; HLA: 인간 백혈구 항원, 또는 인간 MHC 유전자 좌위; NGS: 차세대 시퀀싱; PPV: 양성 예측값; TSNA: 종양-특이적 신생항원; FFPE: 포르말린-고정된, 파라핀-임베딩된; NMD: 넌센스-매개 붕괴; NSCLC: 비-소세포 폐암; DC: 수지상 세포.

[0169] 명세서 및 첨부된 청구범위에서 사용된 바와 같이, 단수 형태 "하나", 및 "그"는 문맥이 명백하게 달리 지시하지 않는 한 복수 지시 대상을 포함한다는 점에 유의해야 한다.

[0170] 구체적으로 언급되지 않거나 문맥상 명백하지 않은 한, 본원에 사용된 용어 "약"은 당업계의 정상 허용 오차 범위 내, 예를 들어 평균의 2 표준 편차 내로 이해된다. 명시된 값의 약 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05% 또는 0.01% 이내로 이해될 수 있다. 문맥상 명백하지 않는 한, 여기에 제공된 모든 수치는 약이라는 용어로 수정된다.

[0171] 여기에서 직접 정의되지 않은 모든 용어는 본 발명의 기술 분야 내에서 이해되는 바와 같이 일반적으로 관련된 의미를 갖는 것으로 이해되어야 한다. 특정 용어는 본 발명의 측면의 조성물, 장치, 방법 등, 및 이들을 제조 또는 사용하는 방법을 설명함에 있어서 의사에게 추가적인 지침을 제공하기 위해 본원에서 논의된다. 동일한 것이 하나 이상의 방식으로 말해질 수 있음을 이해할 것이다. 결과적으로, 대체 언어 및 동의어가 여기에서 논의된 용어 중 하나 이상에 대해 사용될 수 있다. 용어가 여기에서 정교화되거나 논의되는지 여부는 중요하지 않다. 일부 동의어 또는 대체 가능한 방법, 물질 등이 제공된다. 하나 또는 몇 개의 동의어 또는 등가물을 언급하는 것은 명시적으로 언급되지 않는 한 다른 동의어 또는 등가물의 사용을 배제하지 않는다. 용어의 예를 포함하는 실시예의 사용은 단지 예시를 위한 것이며 본원에서 본 발명의 측면의 범위 및 의미를 제한하지 않는다.

[0172] 명세서의 본문 내에서 인용된 모든 참조, 발행된 특허 및 특허 출원은 모든 목적을 위해 그 전체가 참고로 여기에 포함된다.

[0173] **Ⅱ. 항원 확인**

[0174] 종양 및 정상 엑솜 및 전사체의 NGS 분석을 위한 연구 방법이 설명되고 항원 확인 공간에 적용되었다.[6,14,15] 임상 환경에서 항원 확인을 위한 민감도와 특이성을 높이기 위한 특정 최적화를 고려할 수 있다. 이러한 최적화는 실험실 과정와 관련된 영역 및 NGS 데이터 분석과 관련된 영역의 두 가지 영역으로 그룹화할 수 있다. 설명된 연구 방법은 또한 감염성 질병 유기체, 대상체의 감염 또는 대상체의 감염된 세포에서 항원을 확인하는 것과 같은 다른 설정에서 항원을 확인하는 데 적용될 수 있다. 최적화의 예는, 예를 들어 각각의 전체 내용이 모든 목적을 위해 참조로 여기에 포함되는 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1, 국제 특허 출원 공개 WO/2018/195357 및 WO/2018/208856, US 출원 번호 16/606,577 및 국제 특허 출원 PCT/US2020/021508에 상세히 기술된 방법과 같이 당업자에게 알려져 있다.

[0175] 항원(예를 들어, 감염성 질병 유기체로부터 유래된 항원)을 확인하는 방법에는 세포 표면 상에 제시될 가능성(예를 들어, 수지상 세포와 같은 전문 항원 제시를 포함하여 감염된 세포 또는 면역 세포 상에 MHC에 의해 제시)이 있는 및/또는 면역원성일 가능성이 있는 항원을 확인하는 것이 포함된다. 예로서, 이러한 방법 중 하나는 다음 단계를 포함한다: 감염된 세포 또는 감염성 질병 유기체로부터 엑솜, 전사체 또는 전체 게놈 뉴클레오티드 시퀀싱 및/또는 발현 데이터 중 적어도 하나를 수득하는 단계, 여기서 대상체의 감염된 세포로부터의 뉴클레오티드 시퀀싱 데이터 및/또는 발현 데이터, 감염성 질병 유기체 뉴클레오티드 시퀀싱 및/또는 발현 데이터가 각각의 항원 세트(예를 들어, 감염성 질병 유기체로부터 유래된 항원)의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고; 각각의 항원이 대상체의 감염된 세포 또는 대상체에 존재하는 세포의 세포 표면 상의 하나 이상의 MHC 대립유전자에 의해 제시될 수 있는 일련의 수치적 가능성을 생성하기 위해 각 항원의 펩티드 서열을 하나 이상의 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 선택된 항원 세트를 생성하기 위해 수치적 가능성 세트에 기초하여 항원 세트의 서브세트를 선택하는 단계.

[0176] 제시 모델은 상응하는 표지 세트를 포함하는 참조 데이터 세트(트레이닝 데이터 세트라고도 한다)에 대해 트레이닝된 통계적 회귀 또는 기계 학습(예를 들어, 딥 러닝) 모델을 포함할 수 있으며, 여기서 참조 데이터 세트는 임의로 일부 대상체가 감염을 가질 수 있는 다수의 별개 대상체 각각으로부터 수득되며, 참조 데이터 세트는 감염된 조직으로부터의 엑솜 뉴클레오티드 서열을 나타내는 데이터, 정상 조직으로부터의 엑솜 뉴클레오티드 서열을 나타내는 데이터, 감염된 조직의 전사체 뉴클레오티드 서열을 나타내는 데이터, 감염된 조직의 프로테옴 서열을 나타내는 데이터, 및 감염된 조직의 MHC 펩티돔 서열을 나타내는 데이터, 및 정상 조직의 MHC 펩티옴 서열을 나타내는 데이터 중 적어도 하나를 포함한다. 참조 데이터는 질량 분석 데이터, 시퀀싱 데이터, RNA 시퀀싱 데이터, 발현 프로파일링 데이터 및 합성 단백질, 정상 및 종양 인간 세포주 및 신선 및 냉동 1차 샘플, 및 T 세포 검정(예를 들어, ELISpot)에 후속적으로 노출되는 미리 결정된 MHC 대립유전자를 발현하도록 조작된 단일-대립유전자 세포주에 대한 프로테오믹스 데이터를 추가로 포함할 수 있다. 특정 측면에서, 참조 데이터 세트는 참조 데이터의 각 형태를 포함한다.

[0177] 제시 모델은 참조 데이터 세트로부터 적어도 부분적으로 유래된 특징 세트를 포함할 수 있고, 여기서 특징 세트는 대립유전자 의존적-특징 및 대립유전자-독립적 특징 중 적어도 하나를 포함한다. 특정 측면에서 각 기능이 포함된다.

[0178] 공유 항원을 확인하는 방법은 또한 감염된 세포의 표면 상에 제시될 가능성이 있는 대상체의 하나 이상의 세포로부터 하나 이상의 항원을 확인함으로써 개인화된 암 백신을 구성하기 위한 출력을 생성하는 것을 포함한다. 예로서, 이러한 방법 중 하나는 다음 단계를 포함한다: 대상체의 감염된 세포 및 정상 세포로부터 엑솜, 전사체 또는 전체 게놈 뉴클레오티드 시퀀싱 및/또는 발현 데이터 중 적어도 하나를 수득하는 단계, 여기서 뉴클레오티드 시퀀싱 및/또는 발현 데이터는 감염된 세포로부터의 뉴클레오티드 시퀀싱 및/또는 발현 데이터와 정상 세포로부터의 뉴클레오티드 시퀀싱 및/또는 발현 데이터, 대상체의 정상 세포로부터 확인된 펩티드 서열을 비교함으로써 확인된 항원 세트 각각의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되며; 각각의 항원의 펩티드 서열을 상응하는 수치 벡터로 코딩하는 단계, 각각의 수치 벡터는 펩티드 서열을 구성하는 다수의 아미노산 및 펩티드 서열 내의 아미노산의 위치 세트에 관한 정보를 포함하고; 항원 세트에 대한 제시 가능성 세트를 생성하기 위해 컴퓨터 프로세서를 사용하여 수치 벡터를 딥 러닝 제시 모델에 입력하는 단계, 세트에서 제시 가능성은 상응하는 항원이 대상체의 감염된 세포 표면 상에 하나 이상의 클래스 II MHC 대립유전자에 의해 제시될

가능성, 딥 러닝 제시 모델을 나타내고; 선택된 항원 세트를 생성하기 위해 제시 가능성 세트에 기초하여 항원 세트의 서브세트를 선택하는 단계; 및 선택된 항원 세트에 기초하여 개인화 암 백신을 구성하기 위한 출력을 생성하는 단계.

[0179] 항원(예를 들어, 감염성 질병 유기체 유래 항원)을 확인하기 위한 특정 방법은 당업자에게 공지되어 있으며, 예를 들어 국제 특허 출원 공개 WO/2017/106638, WO/2018/195357 및 WO/2018/208856에 보다 상세히 기재되어 있으며, 각각은 모든 목적을 위해 그 전체가 참조로 본원에 포함된다.

[0180] 감염이 있는 대상체을 치료하는 방법이 본원에 기술되어 있으며, 이 방법은본원에 설명된 임의의 항원 확인 방법의 단계를 수행하고, 선택된 항원 세트를 포함하는 감염성 질병 백신을 수득하는 단계, 및 감염성 질병 백신을 대상체에게 투여하는 단계를 추가로 포함한다.

[0181] 본원에 개시된 방법은 또한 서브세트의 항원 중 적어도 하나에 대해 항원-특이적인 하나 이상의 T 세포를 확인하는 것을 포함할 수 있다. 일부 구현양태에서, 확인은 하나 이상의 항원-특이적 T 세포를 확장시키는 조건 하에 서브세트 내의 하나 이상의 항원과 하나 이상의 T 세포를 공동-배양하는 것을 포함한다. 추가 구현양태에서, 확인은 T 세포와 사량체 사이의 결합을 허용하는 조건 하에 서브세트 내의 하나 이상의 항원을 포함하는 사량체와 하나 이상의 T 세포를 접촉시키는 것을 포함한다. 또 다른 구현양태에서, 본원에 개시된 방법은 또한 하나 이상의 확인된 T 세포의 하나 이상의 T 세포 수용체(TCR)를 확인하는 것을 포함할 수 있다. 특정 구현양태에서, 하나 이상의 T 세포 수용체를 확인하는 것은 하나 이상의 확인된 T 세포의 T 세포 수용체 서열을 시퀀싱하는 것을 포함한다. 본원에 개시된 방법은 하나 이상의 확인된 T 세포 수용체 중 적어도 하나를 발현하도록 다수의 T 세포를 유전적으로 조작하는 단계; 다수의 T 세포를 확장시키는 조건 하에 다수의 T 세포를 배양하는 단계; 및 확장된 T 세포를 대상체에 주입하는 단계를 포함한다. 일부 구현양태에서, 하나 이상의 확인된 T 세포 수용체 중 적어도 하나를 발현하도록 다수의 T 세포를 유전적으로 조작하는 것은 하나 이상의 확인된 T 세포의 T 세포 수용체 서열을 발현 벡터로 클로닝하는 단계; 및 다수의 T 세포 각각을 발현 벡터로 형질주입시키는 단계를 포함한다. 일부 구현양태에서, 본원에 개시된 방법은 하나 이상의 확인된 T 세포를 확장시키는 조건 하에 하나 이상의 확인된 T 세포를 배양하는 단계; 및 확장된 T 세포를 대상체에 주입하는 단계를 포함한다.

[0182] 서브세트에서 적어도 하나의 선택된 항원에 대해 항원-특이적인 분리된 T 세포가 또한 본원에 개시되어 있다.

[0183] 또한, 하기 단계를 포함하여 감염성 질병 백신을 제조하는 방법이 제공된다: 대상체의 감염된 세포로부터 엑솜, 전사체 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 및/또는 발현 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 및/또는 발현 데이터는 각각의 항원 세트의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고(예를 들어, 펩티드가 정상 세포 또는 조직과 비교하여 감염된 세포 또는 감염된 조직에서 발현이 변경된 것으로 알려져 있거나 발견된 임의의 폴리펩티드에서 유래된 경우); 각각의 항원이 대상체의 감염된 세포의 세포 표면 상의 하나 이상의 MHC 대립유전자에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각각 항원의 펩티드 서열을 하나 이상의 제시 모델에 입력하는 단계, 수치적 가능성의 세트 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 선택된 항원 세트를 생성하기 위해 수치적 가능성 세트에 기초하여 항원 세트의 서브세트를 선택하는 단계; 및 선택된 항원 세트를 포함하는 감염성 질병 백신을 생산하거나 생산한 단계.

[0184] 하기 단계를 포함하는 방법을 수행함으로써 선택된 선택된 항원 세트를 포함하는 감염성 질병 백신이 또한 본원에 제공된다: 대상체의 감염된 세포로부터 엑솜, 전사체 또는 전체 게놈 감염성 질병 유기체 뉴클레오티드 시퀀싱 및/또는 발현 데이터 중 적어도 하나를 수득하는 단계, 여기서 감염성 질병 유기체 뉴클레오티드 시퀀싱 및/또는 발현 데이터는 각각의 항원 세트의 펩티드 서열을 나타내는 데이터를 수득하는 데 사용되고(예를 들어, 펩티드가 정상 세포 또는 조직과 비교하여 감염된 세포 또는 감염된 조직에서 발현이 변경된 것으로 알려져 있거나 변경이 발견된 임의의 폴리펩티드에서 유래된 경우); 각각의 항원이 대상체의 감염된 세포의 세포 표면 상의 하나 이상의 MHC 대립유전자에 의해 제시되는 수치적 가능성의 세트를 생성하기 위해 각각 항원의 펩티드 서열을 하나 이상의 제시 모델에 입력하는 단계, 수치적 가능성의 세트는 적어도 수신된 질량 분석 데이터에 기초하여 확인되고; 및 선택된 항원 세트를 생성하기 위해 수치적 가능성 세트에 기초하여 항원 세트의 서브세트를 선택하는 단계; 및 선택된 항원 세트를 포함하는 감염성 질병 백신을 생산하거나 생산한 단계.

[0185] 백신은 뉴클레오티드 서열, 폴리펩티드 서열, RNA, DNA, 세포, 플라스미드 또는 벡터 중 하나 이상을 포함할 수 있다.

[0186] 백신은 감염된 세포 표면 상에 제시되는 하나 이상의 항원을 포함할 수 있다.

[0187] 감염성 질병 백신은 대상체에서 면역원성인 하나 이상의 항원을 포함할 수 있다.

[0188] 감염성 질병 백신은 대상체의 정상 조직에 대한 자가면역 반응을 유도하는 하나 이상의 항원을 포함하지 않을 수 있다.

[0189] 감염성 질병 백신은 보조제를 포함할 수 있다.

[0190] 감염성 질병 백신은 부형제를 포함할 수 있다.

[0191] 본원에 개시된 방법은 또한 제시 모델에 기초하여 선택되지 않은 항원에 비해 감염된 세포 표면 상에 제시될 가능성이 증가된 항원을 선택하는 것을 포함할 수 있다.

[0192] 본원에 개시된 방법은 또한 제시 모델에 기초하여 선택되지 않은 항원에 비해 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도할 수 있는 가능성이 증가된 항원을 선택하는 것을 포함할 수 있다.

[0193] 본원에 개시된 방법은 또한 제시 모델을 기초하여 선택되지 않은 항원에 비해 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가된 항원을 선택하는 것을 포함할 수 있으며, 임의로 여기서 APC는 수지상 세포(DC)이다.

[0194] 본원에 개시된 방법은 또한 제시 모델에 기초하여 선택되지 않은 항원에 비해 중추 또는 말초 내성을 통해 억제될 가능성이 감소된 항원을 선택하는 것을 포함할 수 있다.

[0195] 본원에 개시된 방법은 또한 제시 모델에 기초하여 선택되지 않은 항원에 비해 대상체의 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소된 항원을 선택하는 것을 포함할 수 있다.

[0196] 엑솜 또는 전사체 뉴클레오티드 시퀀싱 및/또는 발현 데이터는 감염된 조직에서 시퀀싱을 수행하여 얻을 수 있다.

[0197] 시퀀싱은 차세대 시퀀싱(NGS) 또는 임의의 대규모 병렬 시퀀싱 접근법일 수 있다.

[0198] 수치적 가능성의 세트는 다음 중 적어도 하나를 포함하는 적어도 MHC-대립유전자 상호작용 특징에 의해 추가로 확인될 수 있다: MHC 대립유전자 및 항원 코딩된 펩티드가 결합하는 예측된 친화도; 항원 코딩된 펩티드-MHC 복합체의 예측된 안정성; 항원 코딩된 펩티드의 서열 및 길이; 질량 분석 프로테오믹스 또는 기타 수단에 의해 평가된 특정 MHC 대립유전자를 발현하는 다른 개체로부터 세포에서 유사한 서열을 갖는 항원 코딩된 펩티드의 제시 확률; 문제의 대상체에서 특정 MHC 대립유전자의 발현 수준(예를 들어, RNA-seq 또는 질량 분석법에 의해 측정); 특정 MHC 대립유전자를 발현하는 다른 별개의 대상체에서 특정 MHC 대립유전자에 의한 제시의 전체 항원 코딩된 펩티드-서열-독립적 확률; 다른 별개의 대상체에서 분자의 동일한 계열(예를 들어, HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP)에서 MHC 대립유전자에 의한 제시의 전체 항원 코딩된 펩티드-서열-독립적 확률.

[0199] 수치적 가능성의 세트는 다음 중 적어도 하나를 포함하는 적어도 MHC-대립유전자 비상호작용 특징에 의해 추가로 확인된다: 이의 소스 단백질 서열 내에서 항원 코딩된 펩티드를 플랭킹하는 C- 및 N-말단 서열; 임의로 감염된 세포에서 상응하는 프로테아제의 발현에 따라 가중된(RNA-seq 또는 질량 분석법에 의해 측정), 항원 코딩된 펩티드에서 프로테아제 절단 모티프의 존재; 적절한 세포 유형에서 측정된 소스 단백질의 회전율; 임의로 RNA-seq 또는 프로테옴 질량 분석법에 의해 측정되거나 DNA 또는 RNA 서열 데이터에서 검출된 생식계열 또는 체세포 스플라이싱 돌연변이의 주석으로부터 예측된 바와 같이 감염된 세포에서 가장 높게 발현되는 특정 스플라이스 변이체("이소폼")를 고려한 소스 단백질의 길이; 감염된 세포에서 프로테아솜, 면역프로테아솜, 티모프로테아솜 또는 기타 프로테아제의 발현 수준(이는 RNA-seq, 프로테옴 질량 분석법 또는 면역조직화학에 의해 측정될 수 있다); 항원 코딩된 펩티드의 소스 유전자의 발현(예를 들어, RNA-seq 또는 질량 분석법에 의해 측정); 세포 주기의 다양한 단계 동안 항원 코딩된 펩티드의 소스 유전자의 전형적인 조직-특이적 발현; 예를 들어, uniProt 또는 PDB (http://www.rcsb.org/pdb/home/home.do)에서 찾을 수 있는 소스 단백질 및/또는 해당 도메인의 특징에 대한 포괄적인 카탈로그; 펩티드, 예를 들어 2차 또는 3차 구조(예를 들어, 알파 나선 대 베타 시트)를 함유하는 소스 단백질의 도메인 특성을 설명하는 특징; 대체 스플라이싱; 다른 별개의 대상체에서 문제의 항원 코딩 펩티드의 소스 단백질로부터 펩티드가 제시될 확률; 펩티드가 기술적 편향으로 인해 질량 분석법에 의해 검출되지 않거나 과대-표현되지 않을 확률; RNASeq(펩티드의 소스 단백질을 함유할 필요는 없다)에 의해 측정된 감염된 세포, 기질 또는 감염된 조직의 상태에 대한 정보를 제공하는 다양한 유전자 모듈/경로의 발현; 감염된 세포에서 항원 코딩된 펩티드의 소스 유전자의 카피 수; 펩티드가 TAP에 결합할 확률 또는 펩티드의 TAP에 대한 측정 또는 예측된 결합 친화도; 감염된 세포에서 TAP의 발현 수준(이는 RNA-seq, 프로테옴 질량 분석법, 면역조직화학에 의해 측정될 수 있다); EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3과

같은 알려진 암 유발 유전자에서 및 항원 제시 기구에 관여하는 단백질 코딩 유전자에서(예를 들어, B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 또는 프로테아솜 또는 면역프로테아솜의 구성요소를 코딩하는 임의의 유전자) 드라이버 돌연변이를 포함하나 이로 제한되지 않는 종양 돌연변이의 존재 또는 부재. 제시가 종양에서 기능-상실 돌연변이에 영향을 받는 항원-제시 기구의 구성요소에 의존하는 펩티드는 제시 가능성을 감소시켰다; 다음을 포함하지만 이에 제한되지 않는 기능적 생식계열 다형성의 존재 또는 부재: 항원 제시 기구에 관여하는 단백질을 코딩하는 유전자에서(예를 들어, B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 또는 프로테아솜 또는 면역프로테아솜의 구성요소를 코딩하는 임의의 유전자); 감염 유형(예를 들어, 병원체 감염, 바이러스 감염, 박테리아 감염, 진균 감염 및 기생충 감염); 임상 감염 하위유형(예를 들어, HIV 감염, HPV 감염, SARS 감염, SARS-CoV-2 감염, 에볼라 감염, HBV 감염, 인플루엔자 감염, HCV 감염, CMV 감염, 치쿤구니야 바이러스 감염, RSV 감염, 뎅기열 바이러스 감염, 오르티믹소바이러스과의 바이러스 감염, 및 결핵 감염); 흡연 이력; 관련 감염 유형 또는 임상 하위 유형에서 펩티드의 소스 유전자의 전형적인 발현.

[0200] 본원에 개시된 방법은 또한 선택된 항원(예를 들어, 감염성 질병 유기체 유래 항원) 또는 이의 서브세트를 포함하는 감염성 질병 백신을 수득하는 것을 포함할 수 있으며, 임의로 감염성 질병 백신을 대상체에게 투여하는 것을 추가로 포함한다.

[0201] 선택된 항원 세트에서 항원(예를 들어, 감염성 질병 유기체 유래 항원) 중 적어도 하나는 폴리펩티드 형태일 때 다음 중 적어도 하나를 포함할 수 있다: 1000 nM 미만의 IC50 값을 갖는 MHC와의 결합 친화도, MHC 클래스 I 폴리펩티드의 경우 8-15, 8, 9, 10, 11, 12, 13, 14 또는 15개의 아미노산 길이, MHC 클래스 II 폴리펩티드의 경우 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개의 아미노산, 프로테아좀 절단을 촉진하는 모 단백질 서열의 폴리펩티드 내부 또는 근처에 있는 서열 모티프의 존재, 및 TAP 수송을 촉진하는 서열 모티프의 존재. MHC 클래스 II의 경우, 세포외 또는 리소좀 프로테아제(예를 들어, 카텝신) 또는 HLA-DM 촉매된 HLA 결합에 의한 절단을 촉진하는 펩티드 내부 또는 근처에 서열 모티프의 존재.

[0202] 감염된 세포의 세포 표면 상에 제시될 가능성이 있는 하나 이상의 항원(예를 들어, 감염성 질병 유기체 유래 항원)을 확인하는 방법이 본원에 개시되어 있으며, 이 방법은 다음 단계를 실행하는 단계를 포함한다: 다수의 신선 또는 냉동 샘플로부터 유래된 주요 조직적합성 복합체(MHC)로부터 용리된 다수의 분리된 펩티드와 관련된 데이터를 포함하는 질량 분석 데이터 수신하는 단계; 샘플에 존재하고 각각의 트레이닝 펩티드 서열과 관련된 하나 이상의 MHC 대립유전자에 제시된 트레이닝 펩티드 서열의 세트를 적어도 확인함으로써 트레이닝 데이터 세트를 수득하는 단계; 트레이닝 펩티드 서열에 기초하여 트레이닝 단백질 서열 세트를 수득하는 단계; 및 트레이닝 단백질 서열 및 트레이닝 펩티드 서열을 사용하여 제시 모델의 수치적 매개변수 세트를 트레이닝하는 단계; 여기서, 제시 모델은 감염된 세포로부터의 펩티드 서열이 감염된 세포 표면 상의 하나 이상의 MHC 대립유전자에 의해 제시될 다수의 수치적 가능성을 제공한다.

[0203] 제시 모델은 다음 사이의 의존성을 나타낼 수 있다: 펩티드 서열의 특정 위치에서 MHC 대립유전자 중 특정 하나 및 특정 아미노산의 쌍의 존재; 및 특정 위치에 특정 아미노산을 포함하는 이러한 펩티드 서열이 쌍의 MHC 대립유전자 중 특정 하나에 의해 감염된 세포 표면에 제시될 가능성.

[0204] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 단계를 포함할 수 있으며, 여기서 항원의 서브세트는 하나 이상의 별개의 항원과 비교하여 각각이 감염된 세포의 세포 표면에 제시될 가능성이 증가하기 때문에 선택된다.

[0205] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 것을 포함할 수 있으며, 여기서 항원의 서브세트는 하나 이상의 별개의 항원과 비교하여 대상체에서 각각이 질병-특이적 면역 반응을 유도할 수 있는 가능성이 증가하기 때문에 선택된다.

[0206] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 것을 포함할 수 있으며, 여기서 항원의 서브세트는 하나 이상의 별개의 항원과 비교하여 각각이 전문 항원 제시 세포(APC)에 의해 나이브 T 세포에 제시될 수 있는 가능성이 증가하기 때문에 선택되고, 임의로 여기서 APC는 수지상 세포(DC)이다.

[0207] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 것을 포함할 수 있으며, 여기서 항원의 서브세트는 하나 이상의 별개의 항원과 비교하여 각각이 중추 또는 말초 내성을 통해 억제될 가능성이 감소하기 때문에 선택된다.

[0208] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 것을 포함할 수 있으며, 여기서 항원의 서브세트는 하나 이상의 별개의 항원과 비교하여 대상체에서 각각이 정상 조직에 대한 자가면역 반응을 유도할 수 있는 가능성이 감소하기 때문에 선택된다.

[0209] 본원에 개시된 방법은 또한 항원의 서브세트(예를 들어, 감염성 질병 유기체 유래 항원)를 선택하는 것을 포함할 수 있으며, 여기서 항원의 서브세트는 각각 감염된 세포 대 APC에서 번역 후 차등적으로 변형될 가능성이 감소하기 때문에 선택되고, 임의로 여기서 APC는 수지상 세포(DC)이다.

[0210] 본원에서 방법의 실행은 달리 명시되지 않는 한, 당업계의 기술 내에서 단백질 화학, 생화학, 재조합 DNA 기술 및 약리학의 통상적인 방법을 사용할 것이다. 이러한 기술은 문헌에 완전히 설명되어 있다(T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, 등, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick 및 N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992)).

[0211] **III. 항원**

[0212] 항원은 뉴클레오티드 또는 폴리펩티드를 포함할 수 있다. 예를 들어, 항원은 폴리펩티드 서열을 코딩하는 RNA 서열일 수 있다. 따라서 백신에 유용한 항원은 뉴클레오티드 서열 또는 폴리펩티드 서열을 포함할 수 있다. 항원은 감염성 질병 유기체의 뉴클레오티드 서열 또는 폴리펩티드 서열로부터 유래될 수 있다. 감염성 질병 유기체의 폴리펩티드 서열은 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및/또는 기생충-유래 펩티드를 포함하지만 이에 제한되지 않는다. 감염성 질병 유기체에는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵이 포함되지만 이에 제한되지는 않는다.

[0213] 본원에 개시된 방법에 의해 확인된 감염성 질병 유기체 특이적 항원 또는 에피토프를 포함하는 분리된 펩티드, 공지된 감염성 질병 유기체 특이적 항원 또는 에피토프를 포함하는 펩티드, 및 본원에 개시된 방법에 의해 확인된 돌연변이체 폴리펩티드 또는 이의 단편이 개시된다. 항원이 관련된 폴리펩티드 서열을 코딩하는 뉴클레오티드 서열(예를 들어, DNA 또는 RNA)을 포함하는 이들의 코딩 서열의 맥락에서 항원 펩티드가 설명될 수 있다.

[0214] 백신에 혼입될 수 있는 항원(예를 들어, 카세트에 코딩되어)은 인간 및 인간이 아닌 척추동물을 감염시키는 병원성 바이러스와 같은 바이러스에 대해 인간 또는 인간이 아닌 동물을 면역화하는 데 유용한 면역원을 포함한다. 항원은 다양한 바이러스 과에서 선택될 수 있다. 면역 반응이 바람직한 바람직한 바이러스 과의 예는 감기의 약 50%를 차지하는 리노바이러스 속을 포함하는 피코르나바이러스 과; 폴리오바이러스, 콕사키바이러스, 에코바이러스 및 A형 간염 바이러스와 같은 인간 엔테로바이러스를 포함하는 엔테로바이러스 속; 및 주로 인간이 아닌 동물에서 구제역을 담당하는 압토바이러스 속을 포함한다. 피코르나바이러스 과의 바이러스 내에서 표적 항원은 VP1, VP2, VP3, VP4 및 VPG를 포함한다. 또 다른 바이러스 과에는 유행성 위장염의 중요한 원인 인자인 노워크 군의 바이러스를 포함하는 칼시바이러스 과가 있다. 인간 및 비인간 동물에서 면역 반응을 유도하기 위해 항원을 표적화하는데 사용하기에 바람직한 또 다른 바이러스 과는 신드비스 바이러스, 로스리버 바이러스, 및 베네수엘라, 동부 및 서부 말 뇌염을 포함하는 알파바이러스 속및 풍진 바이러스를 포함한 루비바이러스를 포함하는 토가바이러스 과이다. 플라비바이러스과(Flaviviridae) 과에는 뎅기열, 황열병, 일본 뇌염, 세인트루이스 뇌염 및 진드기 매개 뇌염 바이러스가 포함된다. 다른 표적 항원은 C형 간염 또는 코로나바이러스 과에서 생성될 수 있으며, 이는 감염성 기관지염 바이러스(가금류), 돼지 전염성 위장 바이러스(돼지), 돼지 혈구응집성 뇌척수염 바이러스(돼지), 고양이 감염성 복막염 바이러스(고양이), 고양이 장 코로나바이러스(고양이), 개 코로나바이러스(개)와 같은 많은 비인간 바이러스 및 일반적인 감기 및/또는 비-A, B 또는 C형 간염을 유발할 수 있는 인간 호흡기 코로나바이러스를 포함한다. 코로나바이러스 과 내에서 표적 항원에는 E1(M 또는 기질 단

백질이라고도 한다), E2(S 또는 스파이크 단백질이라고도 한다), E3(HE 또는 헤마글루틴-엘테로스라고도 한다) 당단백질(모든 코로나바이러스에 존재하지 않는) 또는 N(뉴클레오캡시드)이 포함된다. 또 다른 항원은 베시큘로바이러스 속(예를 들어, 수포성 구내염 바이러스) 및 일반 리사바이러스(예를 들어, 광견병)를 포함하는 랍도바이러스 과에 대해 표적화될 수 있다. 랍도바이러스 과 내에서, 적합한 항원은 G 단백질 또는 N 단백질로부터 유래될 수 있다. 마르부르크(Marburg) 및 에볼라 바이러스와 같은 출혈열 바이러스를 포함하는 필로바이러스과는 항원의 적합한 소스일 수 있다. 파라믹소바이러스 과에는 파라인플루엔자 바이러스 유형 1, 파라인플루엔자 바이러스 유형 3, 소 파라인플루엔자 바이러스 유형 3, 풍진 바이러스(볼거리 바이러스), 파라인플루엔자 바이러스 유형 2, 파라인플루엔자 바이러스 유형 4, 뉴캐슬병 바이러스(닭), 린더페스트, 홍역과 개 홍역을 포함하는 모르빌리바이러스 및 호흡기 세포융합 바이러스를 포함하는 뉴모바이러스(예를 들어, 글리코-(G) 단백질 및 융합(F) 단백질, 서열은 GenBank에서 입수가능)가 포함된다. 인플루엔자 바이러스는 오르토믹소바이러스 과로 분류되며 항원의 적합한 소스(예를 들어, HA 단백질, N1 단백질)가 될 수 있다. 부냐바이러스 과에는 부냐바이러스 속(캘리포니아 뇌염, 라크로스), 플레보바이러스(리프트 밸리 열), 한타바이러스(퓨어말라는 헤마긴열 바이러스이다), 나이로바이러스(나이로비 양 질병) 및 다양한 지정되지 않은 붕가바이러스가 포함된다. 아레나바이러스 과는 LCM 및 라사열 바이러스에 대한 항원 소스을 제공한다. 레오바이러스과에는 레오바이러스, 로타바이러스(소아에서 급성 위장염을 일으킨다), 오르비바이러스 및 컬티바이러스(콜로라도 진드기열, 레봄보(인간), 말 뇌증, 푸른 혀) 속이 포함된다. 레트로바이러스 과는 고양이 백혈병 바이러스, HTLVI 및 HTLVII, 렌티바이러스(인간 면역결핍 바이러스(HIV), 원숭이 면역결핍 바이러스(SIV) 포함), 고양이 면역결핍 바이러스(FIV), 말 감염성 빈혈 바이러스 및 스푸마바이러스를 포함한다)와 같은 인간 및 수의학적 질병을 포함하는 아과 온코리바이러스를 포함한다. 렌티바이러스 중에서 적합한 항원이 많이 기술되어 있으며 쉽게 선택할 수 있다. 적합한 HIV 및 SIV 항원의 예는 gag, pol, Vif, Vpx, VPR, Env, Tat, Nef 및 Rev 단백질 뿐만 아니라 이의 다양한 단편을 포함하나 이에 제한되지 않는다. 예를 들어, Env 단백질의 적합한 단편은 gp120, gp160, gp41 또는 이의 더 작은 단편, 예를 들어 약 적어도 8개의 아미노산 길이와 같은 이의 서브유닛 중 임의의 것을 포함할 수 있다. 유사하게, tat 단백질의 단편이 선택될 수 있다(U.S. Pat. No. 5,891,994, U.S. Pat. No. 6,193,981; the HIV and SIV proteins, D. H. Barouch 등, J. Virol., 75(5):2462-2467 (March 2001), and R. R. Amara, 등, *Science,* 292:69-74 (6 Apr. 2001)). 다른 예에서, HIV 및/또는 SIV 면역원성 단백질 또는 펩티드는 융합 단백질 또는 다른 면역원성 분자를 형성하기 위해 사용될 수 있다. 예를 들어, 2001년 8월 2일 공개된 WO 01/54719 및 1999년 4월 8일 공개된 WO 99/16884에 기재된 HIV-1 Tat 및/또는 Nef 융합 단백질 및 면역화 요법을 참조할 수 있다. 본 발명은 본원에 기재된 HIV 및/또는 SIV 면역원성 단백질 또는 펩티드에 제한되지 않는다. 또한, 이들 단백질에 대한 다양한 변형이 기술되어 있거나 당업자에 의해 용이하게 이루어질 수 있고, 변형된 gag단백질이 U.S. 특허 제5,972,596호에 기재되어 있다. 추가로, 임의의 원하는 HIV 및/또는 SIV 면역원은 단독으로 또는 조합하여 전달될 수 있다. 이러한 조합은 단일 벡터 또는 다중 벡터로부터의 발현을 포함할 수 있다. 파포바바이러스 과에는 아과 폴리오마바이러스(BKU 및 JCU 바이러스) 및 아과 유두종바이러스(유두종의 악성 진행 또는 암과 관련)가 포함된다. 아데노바이러스 과에는 호흡기 질병 및/또는 장염을 유발하는 바이러스(EX, AD7, ARD, O.B.)가 포함된다. 파보바이러스 과 고양이 파보바이러스(고양이 장염), 고양이 범백혈구감소증바이러스, 개 파보바이러스 및 돼지 파보바이러스. 헤르페스바이러스과에는 단순바이러스(HSVI, HSVII) 속, 수두바이러스(가성광견병, 수두 대상포진)를 포함하는 알파헤르페스바이러스 아과 및 거대세포바이러스(인간 CMV) 속, 무로메갈로바이러스 속을 포함하는 베타헤르페스바이러스 아과 및 림포크립토바이러스, EBV(버킷 림프종), 감염성 비강기관지염, 마렉병 바이러스 및 라디노바이러스 속을 포함하는 감마헤르페스바이러스 아과가 포함된다. 폭스바이러스 과에는 오르토폭스바이러스(바리올라(천연두) 및 백시니아 (우두)), 파라폭스바이러스, 아비폭스바이러스, 카프리폭스바이러스, 레포리폭스바이러스, 수이폭스바이러스 및 아과엔토폭시이리나에를 포함하는 아과 코르도폭시이리나에가 포함된다. 헤파드나바이러스 과에는 B형 간염 바이러스가 포함된다. 항원의 적절한 소스가 될 수 있는 분류되지 않은 바이러스 중 하나는 델타 간염 바이러스이다. 또 다른 바이러스 소스는 조류 감염성 윤활낭 질병 바이러스 및 돼지 호흡기 및 생식 증후군 바이러스를 포함할 수 있다. 알파바이러스 과에는 말 동맥염 바이러스 및 다양한 뇌염 바이러스가 포함된다.

[0215] 백신에 혼입될 수 있는 항원(예를 들어, 카세트에 코딩되어)은 인간 및 비-인간 척추동물을 감염시키는 박테리아, 진균, 기생 미생물 또는 다세포 기생충을 포함하는 병원체에 대해 인간 또는 비인간 동물을 면역화하는 데 유용한 면역원도 또한 포함된다. 박테리아 병원체의 예로는 병원성 그람-양성 구균이 포함되며, 폐렴구균; 포도구균; 및 연쇄상 구균이 포함된다. 병원성 그람-음성 구균에는 수막 구균; 임균이 포함된다. 병원성 장내 그람-음성 간균은 장내세균과; 슈도모나스(*pseudomonas*), 아시네토박테리아 및 에이케넬라(*eikenella*); 멜리오이드증; 살모넬라균(*salmonella*); 이질균(*shigella*); 헤모필루스(*haemophilus*)(헤모필루스 인플루엔자

(*Haemophilus influenzae*), 헤모필루스 솜누스(*Haemophilus somnus*)); 모락셀라(*moraxella*); H. 듀크레이(*H. ducreyi*)(연골종을 유발); 브루셀라(*brucella*); 프러니셀라 툴라렌시스(*Franisella tularensis*)(야토병을 유발); 여시니아(*yersinia*)(파스퇴렐라(*pasteurella*)); 스트렙토바실러스 모닐리포르미스(*streptobacillus moniliformis*) 및 스피릴룸(*spirillum*)을 포함한다. 그람-양성 간균은 리스테리아 모노사이토제네스(*listeria monocytogenes*); 에리시펠로트릭스 류시오패티에(*erysipelothrix rhusiopathiae*); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*)(디프테리아); 콜레라; B. 탄저균(*anthracis*)(anthrax); 도노바노시스(사타구니 육아종); 및 바르토넬라증이 포함된다. 병원성 혐기성 박테리아에 의해 유발되는 질병에는 파상풍; 보툴리누스 중독; 다른 클로스트리듐; 결핵; 나병; 및 기타 마이코박테리아가 포함된다. 특정 박테리아 종의 예는 제한없이 스트렙토코커스 뉴모니애(*Streptococcus pneumoniae*), 스트렙토코커스 피오게네스(*Streptococcus pyogenes*), 스트렙토코커스 아갈락티아에(*Streptococcus agalactiae*), 스트렙토코커스 파에칼리스(*Streptococcus faecalis*), 모락셀라 카타랄리스(*Moraxella catarrhalis*), 헬리코박터 파일로리균(*Helicobacter pylori*), 나이세리아 수막염(*Neisseria meningitidis*), 나이세리아 임질(*Neisseria gonorrhoeae*), 클라미디아 트라코마티스(*Chlamydia trachomatis*), 클라미디아 뉴모니아(*Chlamydia pneumoniae*), 클라미디아 시타시(*Chlamydia psittaci*), 보르데텔라 백일해(*Bordetella pertussis*), 살모넬라 티피(*Salmonella typhi*), 살모넬라 티피무리움(*Salmonella typhimurium*), 살모넬라 콜레라수이스(*Salmonella choleraesuis*), 대장균(*Escherichia coli*), 시겔라(*Shigella*), 비브리오 콜레라(*Vibrio cholerae*), 코리네박테리움 디프테리아(*Corynebacterium diphtheriae*), 결핵균(*Mycobacterium tuberculosis*), 마이코박테리움 아비움(*Mycobacterium avium*), 마이코박테리움 세포내 세균복합체(*Mycobacterium intracellulare* complex), 프로테우스 미라빌리스(*Proteus mirabilis*), 프로테우스 불가리스(*Proteus vulgaris*), 황색포도상구균(*Staphylococcus aureus*), 클로스트리디움 테타니(*Clostridium tetani*), 렙토스피라 인터로간(*Leptospira interrogans*), 보렐리아 부르그도르페리(*Borrelia burgdorferi*), 파스퇴렐라 헤몰리티카(*Pasteurella haemolytica*), 파스퇴렐라 멀토시다(*Pasteurella multocida*), 방선균 흉막폐렴(*Actinobacillus pleuropneumoniae*) 및 마이코플라스마 갈리셉티쿰(*Mycoplasma gallisepticum*)이다. 병원성 스피로헤타 질병에는 매독; 트레포네마토시스: 매종(yaws), 핀타 및 풍토성 매독; 및 렙토스피라증이 포함된다. 고병원성 박테리아 및 병원성 진균에 의해 유발되는 기타 감염에는 방선균증; 노심증; 크립토코커스(*Cryptococcus*), 분구균증(*Blastomyces*), 히스토플라스마증(*Histoplasma*) 및 콕시디오이데스진균증(*Coccidiodes*); 칸디다증(*Candida*), 아스페르길루스증(*Aspergillis*), 및 점막진균증; 스포로트리코증; 파라콕시디오도진균증, 페트리엘리디움증, 토룰롭증, 균종 및 염색체 균증; 및 피부사상균증이 포함된다. 리케차 감염에는 장티푸스열, 로키산반점열, Q열 및 리케차두가 포함된다. 마이코플라스마 및 클라미디아 감염의 예는 다음을 포함한다: 마이코플라스마 폐렴(*mycoplasma pneumoniae*); 성성 림프육아종; 시타코증; 및 주산기 클라미디아 감염. 병원성 진핵생물은 병원성 원생동물 및 기생충을 포함하며 이에 의해 생성되는 감염은 아메바증; 말라리아; 리슈만편모충증(예를 들어, 주요 리슈만편모충증(*Leishmania major*)에 의해 유발); 트리파노소마증; 톡소플라스마증(예를 들어, 톡소플라스마 곤디(*Toxoplasma gondii*)에 의해 유발); 뉴모시스티스 카리니이(*Pneumocystis carinii*); 트라이칸스; 톡소플라스마 곤디(*Toxoplasma gondii*); 바베시오증; 편모충증(예를 들어, 편모충(*Giardia*)에 의해 유발); 선모충증(예를 들어, 트리코모나스(*Trichomonas*)에 의해 유발); 사상충증; 주혈흡충증(예를 들어, 주혈흡충(*Schistosoma*)에 의해 유발); 선충류; 흡충류 또는 흡충류; 및 촌충(촌충) 감염. 다른 기생충 감염은 특히 아스카리스(*Ascaris*), 트리츄리스(*Trichuris*), 크립토스포리디움(*Cryptosporidium*) 및 폐포자충(*Pneumocystis carinii*)에 의해 발생할 수 있다.

[0216] 또한 감염성 질병 유기체, 대상체의 감염, 또는 대상체의 감염된 세포와 관련된 임의의 폴리펩티드로부터 유래된 펩티드가 본원에 개시되어 있다. 항원은 감염성 질병 유기체의 핵산 서열 또는 폴리펩티드 서열로부터 유래될 수 있다. 감염성 질병 유기체의 폴리펩티드 서열은 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및/또는 기생충-유래 펩티드를 포함하지만 이에 제한되지 않는다. 감염성 질병 유기체에는 중증 급성 호흡기 증후군 관련 코로나바이러스(SARS), 중증 급성 호흡기 증후군 코로나바이러스 2(SARS-CoV-2), 에볼라, HIV, B형 간염 바이러스(HBV), 인플루엔자, C 형 간염 바이러스(HCV), 인유두종바이러스(HPV), 거대세포바이러스(CMV), 치쿤구니야 바이러스, 호흡기 세포융합 바이러스(RSV), 뎅기열 바이러스, 오르티믹소바이러스과 바이러스 및 결핵이 포함되지만 이에 제한되지는 않는다.

[0217] 수지상 세포와 같은 전문적인 항원 제시 세포를 포함하여, 감염된 세포 또는 면역 세포와 같은 세포의 세포 표면 상에 제시될 것으로 예측되는 항원이 선택될 수 있다. 면역원성으로 예측되는 항원이 선택될 수 있다.

[0218] 항원 뉴클레오티드 서열에 의해 코딩되는 하나 이상의 폴리펩티드는 1000 nM 미만의 IC50 값을 갖는 MHC와의 결합 친화도, MHC 클래스 I 펩티드의 경우 길이 8-15, 8, 9, 10, 11, 12, 13, 14 또는 15개의 아미노산, 프로테

아좀 절단을 촉진하는 펩티드 내부 또는 근처에 서열 모티프의 존재, 및 TAP 수송을 촉진하는 서열 모티프의 존재 중 적어도 하나를 포함할 수 있다. MHC 클래스 II 펩티드의 경우 길이는 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 아미노산, 세포외 또는 리소좀 프로테아제(예를들어, 카텝신) 또는 HLA-DM 촉매된 HLA 결합에 의한 절단을 촉진하는 펩티드 내부 또는 근처에 서열 모티프의 존재.

[0219] 하나 이상의 항원이 감염된 세포의 표면 상에 제시될 수 있다.

[0220] 하나 이상의 항원은, 예를 들어 대상체에서 T 세포 반응 및/또는 B 세포 반응을 유발할 수 있는 감염이 있거나 감염이 의심되는 대상체에서 면역원성일 수 있다. 하나 이상의 항원은 감염 감염의 위험이 있는 대상체에서 면역원성일 수 있으며, 예를 들어 기억 T 세포, 기억 B 세포 및/또는 감염에 특이적인 항체의 생성을 자극하여 감염에 대한 면역학적 보호(즉, 면역)를 제공하는 대상체에서 T 세포 반응 및/또는 B 세포 반응을 유발할 수 있다.

[0221] 하나 이상의 항원은 하나 이상의 항원을 인식하는 항체(예를 들어, 감염성 질병 항원을 인식하는 항체)의 생성과 같은 B 세포 반응을 유발할 수 있다. 항체는 선형 폴리펩티드 서열을 인식하거나 2차 및 3차 구조를 인식할 수 있다. 따라서, B 세포 항원은 전장 단백질, 단백질 서브유닛, 단백질 도메인, 또는 2차 및 3차 구조를 가질 것으로 알려져 있거나 예측되는 임의의 폴리펩티드 서열을 포함하나 이에 제한되지 않는 선형 폴리펩티드 서열 또는 2차 및 3차 구조를 갖는 폴리펩티드를 포함할 수 있다. 일반적으로, 감염에 대한 B 세포 반응을 유발할 수 있는 항원은 감염성 질병 유기체의 표면 상에서 발견되는 항원이다.

[0222] 하나 이상의 항원은 T 세포 반응을 유도할 수 있는 항원(예를 들어, 예측된 T 세포 에피토프 서열을 포함하는 펩티드) 및 T 세포 반응을 유도할 수 있는 별개의 항원(예를 들어, 전장 단백질, 단백질 서브유닛, 단백질 도메인)의 조합을 포함할 수 있다.

[0223] 대상체에서 자가면역 반응을 유도하는 하나 이상의 항원은 대상체에 대한 백신 생성과 관련하여 고려에서 제외될 수 있다.

[0224] 적어도 하나의 항원성 펩티드 분자(예를 들어, 에피토프 서열)의 크기는 약 5, 약 6, 약 7, 약 8, 약 9, 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 26, 약 27, 약 28, 약 29, 약 30, 약 31, 약 32, 약 33, 약 34, 약 35, 약 36, 약 37, 약 38, 약 39, 약 40, 약 41, 약 42, 약 43, 약 44, 약 45, 약 46, 약 47, 약 48, 약 49, 약 50, 약 60, 약 70, 약 80, 약 90, 약 100, 약 110, 약 120개 이상의 아미노 분자 잔기, 및 그 안에서 유도가능한 임의의 범위를 포함할 수 있다. 특정 구현양태에서 항원성 펩티드 분자는 50개 이하의 아미노산이다.

[0225] 항원성 펩티드 및 폴리펩티드는 MHC 클래스 I의 경우 길이가 15개 이하이고 일반적으로 약 8개 내지 약 11개 잔기, 특히 9개 또는 10개 잔기로 구성되고; MHC 클래스 II의 경우 6-30개 잔기를 포함한다.

[0226] 원하는 경우, 더 긴 펩티드가 여러 방식으로 설계될 수 있다. 한 경우에, HLA 대립유전자에 대한 펩티드의 제시 가능성이 예측되거나 알려져 있을 때, 더 긴 펩티드는 다음 중 하나로 구성될 수 있다: (1) 각각의 상응하는 유전자 생성물의 N- 및 C-말단을 향해 2-5개 아미노산의 확장을 갖는 개별 제시 펩티드; (2) 각각에 대해 확장된 서열을 갖는 제시된 펩티드의 일부 또는 전부의 연속. 또 다른 경우, 시퀀싱에서 긴(>10개 잔기) 에피토프 서열이 존재하는 것으로 밝혀지면, 더 긴 펩티드는 다음으로 구성된다: (3) 새로운 감염성 질병-특이적 아미노산의 전체 스트레치 -- 따라서 가장 강력한 HLA 제시 짧은 펩티드의 컴퓨터 또는 시험관 내 테스트-기반 선택에 대한 필요성 우회. 두 경우 모두, 더 긴 펩티드를 사용하면 환자 세포에 의한 내인성 처리가 가능하고 보다 효과적인 항원 제시 및 T 세포 반응 유도로 이어질 수 있다. 더 긴 펩티드는 또한 감염성 질병 유기체에서 발현되는 펩티드의 전장 단백질, 단백질 서브유닛, 단백질 도메인, 및 이들의 조합일 수 있다.

[0227] 항원성 펩티드 및 폴리펩티드는 HLA 단백질 상에 제시될 수 있다. 일부 측면에서 항원성 펩티드 및 폴리펩티드는 야생형 펩티드보다 더 큰 친화도로 HLA 단백질에 제시된다. 일부 측면에서, 항원성 펩티드 또는 폴리펩티드는 적어도 5000 nM 미만, 적어도 1000 nM 미만, 적어도 500 nM 미만, 적어도 250 nM 미만, 적어도 200 nM 미만, 적어도 150nM 미만, 적어도 100nM 미만, 적어도 50nM 미만 또는 그 미만의 IC50을 가질 수 있다.

[0228] 일부 측면에서, 항원성 펩티드 및 폴리펩티드는 대상체에게 투여될 때 자가면역 반응을 유도하지 않거나 및/또는 면역학적 내성을 유발하지 않는다.

[0229] 또한, 적어도 2개 이상의 항원성 펩티드를 포함하는 조성물이 제공된다. 일부 구현양태에서, 조성물은 적어도 2

개의 별개의 펩티드를 함유한다. 적어도 2개의 별개의 펩티드가 동일한 폴리펩티드로부터 유래될 수 있다. 별개의 폴리펩티드는 펩티드가 길이, 아미노산 서열, 또는 둘 다에 따라 변한다는 것을 의미한다. 펩티드는 감염성 질병 유기체에 대해 알려졌거나 관련이 있는 것으로 의심되는 임의의 폴리펩티드, 또는 정상 세포 또는 조직과 비교하여 감염된 세포에서 변경된 발현을 갖는 것으로 알려져 있거나 발견된 임의의 폴리펩티드로부터 유래된 펩티드일 수 있다(예를 들어, 감염성 질병 폴리뉴클레오티드 또는 숙주 세포에 제한된 발현을 갖는 폴리펩티드를 포함하는 감염성 질병 폴리뉴클레오티드 또는 폴리펩티드).

[0230] 원하는 활성 또는 특성을 갖는 항원성 펩티드 및 폴리펩티드는 원하는 MHC 분자에 결합하고 적절한 T 세포를 활성화시키기 위해 비변형된 펩티드의 생물학적 활성을 증가시키거나 적어도 실질적으로 모두 유지하면서, 특정 원하는 속성, 예를 들어 개선된 약리학적 특성을 제공하도록 변형될 수 있다. 예를 들어, 항원성 펩티드 및 폴리펩티드는 보존적 또는 비보존적 치환과 같은 다양한 변화를 받을 수 있으며, 이러한 변화는 개선된 MHC 결합, 안정성 또는 제시와 같은 사용상의 특정 이점을 제공할 수 있다. 보존적 치환은 아미노산 잔기를 생물학적 및/또는 화학적으로 유사한 다른 것으로, 예를 들어 하나의 소수성 잔기를 다른 것으로, 또는 하나의 극성 잔기를 다른 것으로 대체하는 것을 의미한다. 치환은 Gly, Ala; Val, Ile, Leu, Met; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; 및 Phe, Tyr와 같은 조합을 포함한다. 단일 아미노산 치환의 효과는 D-아미노산을 사용하여 조사할 수도 있다. 이러한 변형은, 예를 들어 문헌에 기재된 잘 공지된 합성 방법을 사용하여 만들 수 있다(Merrifield, Science 232:341-347 (1986), Barany & Merrifield, The Peptides, Gross & Meienhofer, eds. (N.Y., Academic Press), pp. 1-284 (1979); and Stewart & Young, Solid Phase Peptide Synthesis, (Rockford, Ill., Pierce), 2d Ed. (1984)).

[0231] 다양한 아미노산 모방체 또는 비천연 아미노산을 사용한 펩티드 및 폴리펩티드의 변형은 생체내 펩티드 및 폴리펩티드의 안정성을 증가시키는데 특히 유용할 수 있다. 안정성은 여러 가지 방법으로 검정할 수 있다. 예를 들어, 펩티다아제 및 인간 혈장 및 혈청과 같은 다양한 생물학적 배지를 사용하여 안정성을 테스트했다(Verhoef 등, Eur. J. Drug Metab Pharmacokin. 11:291-302 (1986)). 펩티드의 반감기는 25% 인간 혈청(v/v) 검정을 사용하여 편리하게 결정할 수 있다. 프로토콜은 일반적으로 다음과 같다. 풀링된 인간 혈청(유형 AB, 비-열 비활성화)은 사용 전에 원심분리에 의해 탈지된다. 그런 다음 혈청을 RPMI 조직 배양 배지로 25%로 희석하고 펩티드 안정성을 테스트하는 데 사용한다. 소정의 시간 간격으로 소량의 반응 용액을 제거하고 6% 트리클로로아세트산 수용액 또는 에탄올에 첨가한다. 탁한 반응 샘플을 15분 동안 냉각(4℃)한 다음 회전시켜 침전된 혈청 단백질을 펠릿화한다. 펩티드의 존재는 그 다음 안정성-특이적 크로마토그래피 조건을 사용하여 역상 HPLC에 의해 결정된다.

[0232] 펩티드 및 폴리펩티드는 개선된 혈청 반감기 이외의 원하는 속성을 제공하도록 변형될 수 있다. 예를 들어, CTL 활성을 유도하는 펩티드의 능력은 T 헬퍼 세포 반응을 유도할 수 있는 적어도 하나의 에피토프를 함유하는 서열에 대한 연결에 의해 향상될 수 있다. 면역원성 펩티드/T 헬퍼 접합체는 스페이서 분자에 의해 연결될 수 있다. 스페이서는 일반적으로 생리학적 조건에서 실질적으로 전하를 띠지 않는 아미노산 또는 아미노산 모방체와 같은 비교적 작은 중성 분자로 구성된다. 스페이서는 전형적으로, 예를 들어 Ala, Gly, 또는 비극성 아미노산 또는 중성 극성 아미노산의 다른 중성 스페이서로부터 선택된다. 임의로 존재하는 스페이서는 동일한 잔기로 구성될 필요가 없고 따라서 헤테로- 또는 호모-올리고머일 수 있음이 이해될 것이다. 존재하는 경우, 스페이서는 일반적으로 적어도 1개 또는 2개의 잔기, 보다 일반적으로 3개 내지 6개의 잔기일 것이다. 대안적으로, 펩티드는 스페이서 없이 T 헬퍼 펩티드에 연결될 수 있다.

[0233] 항원성 펩티드는 펩티드의 아미노 또는 카르복시 말단에서 직접적으로 또는 스페이서를 통해 T 헬퍼 펩티드에 연결될 수 있다. 항원성 펩티드 또는 T 헬퍼 펩티드의 아미노 말단은 아실화될 수 있다. 예시적인 T 헬퍼 펩티드는 파상풍 톡소이드 830-843, 인플루엔자 307-319, 말라리아 환포자소체 382-398 및 378-389를 포함한다.

[0234] 단백질 또는 펩티드는 표준 분자 생물학적 기술을 통한 단백질, 폴리펩티드 또는 펩티드의 발현, 천연 소스으로부터 단백질 또는 펩티드의 분리, 또는 단백질 또는 펩티드의 화학적 합성을 포함하는 당업자에게 공지된 임의의 기술에 의해 제조될 수 있다. 다양한 유전자에 상응하는 뉴클레오티드 및 단백질, 폴리펩티드 및 펩티드 서열은 이전에 개시되었으며, 당해 분야의 숙련가에게 공지된 컴퓨터화된 데이터베이스에서 찾을 수 있다. 그러한 데이터베이스 중 하나는 국립 보건원(National Institutes of Health) 웹사이트에 있는 국립 생명 공학 정보 센터(National Center for Biotechnology Information)의 젠뱅크(Genbank) 및 젠펩트GpenPept) 데이터베이스이다. 공지된 유전자에 대한 코딩 영역은 본원에 개시된 기술을 사용하여 또는 당업자에게 공지된 바와 같이 증폭 및/또는 발현될 수 있다. 대안적으로, 단백질, 폴리펩티드 및 펩티드의 다양한 상업적 제제가

당업자에게 공지되어 있다.

[0235] 추가 측면에서, 항원은 항원성 펩티드 또는 이의 일부를 코딩하는 핵산(예를 들어, 폴리뉴클레오티드)을 포함한다. 폴리뉴클레오티드는, 예를 들어 DNA, cDNA, PNA, CNA, RNA(예를 들어, mRNA), 단일-가닥 및/또는 이중-가닥, 또는 천연 또는 안정화된 형태의 폴리뉴클레오티드, 예를 들어 포스포로티에이트 백본을 갖는 폴리뉴클레오티드, 또는 이들의 조합이며 인트론을 포함하거나 포함하지 않을 수 있다. 항원을 코딩하는 폴리뉴클레오티드 서열은, 예를 들어 전사, 번역, 전사후 처리 및/또는 RNA 안정성의 개선을 통해 발현을 개선하도록 서열 최적화될 수 있다. 예를 들어, 항원을 코딩하는 폴리뉴클레오티드 서열은 코돈-최적화될 수 있다. 본원에서 "코돈-최적화"는 주어진 유기체의 코돈 편향과 관련하여 자주 사용되지 않는 코돈을 자주 사용되는 동의(synonymous) 코돈으로 대체하는 것을 의미한다. 폴리뉴클레오티드 서열은 전사 후 처리를 개선하도록 최적화될 수 있으며, 예를 들어 스플라이싱 모티프(예를 들어, 표준 및/또는 비밀/비표준 스플라이스 도너, 분지 및/또는 수용체 서열)의 제거 및/또는 편향된 스플라이싱 이벤트에 대한 외인성 스플라이싱 모티프(예를 들어, 스플라이스 도너, 분지 및/또는 수용체 서열)의 도입을 통해 의도하지 않은 스플라이싱을 줄이기 위해 최적화될 수 있다. 외인성 인트론 서열은 SV40(예를 들어, SV40 미니-인트론)에서 유래되고 면역글로불린(예를 들어, 인간 β-글로빈 유전자)에서 유래된 것을 포함하지만 이에 제한되지 않는다. 외인성 인트론 서열은 프로모터/인핸서 서열 및 항원(들) 서열 사이에 혼입될 수 있다. 발현 벡터에 사용하기 위한 외인성 인트론 서열은 모든 목적을 위해 본원에 참고로 포함되는 문헌에 기재되어 있다(Callendret 등, Virology. 2007 Jul 5; 363(2): 288-302). 폴리뉴클레오티드 서열은, 예를 들어 RNA 불안정성 모티프(예를 들어, AU-풍부 요소 및 3' UTR 모티프) 및/또는 반복적인 뉴클레오티드 서열의 제거를 통해 전사 안정성을 개선하도록 최적화될 수 있다. 폴리뉴클레오티드 서열은, 예를 들어 비밀 전사 개시자 및/또는 종결자의 제거를 통해 정확한 전사를 개선하도록 최적화될 수 있다. 폴리뉴클레오티드 서열은, 예를 들어 비밀 AUG 개시 코돈, 미성숙 폴리A 서열 및/또는 2차 구조 모티프의 제거를 통해 번역 및 번역 정확도를 개선하도록 최적화될 수 있다. 폴리뉴클레오티드 서열은 구성 수송 요소(CTE), RNA 수송 요소(RTE) 또는 우드척 전사후 조절 요소(WPRE)의 추가를 통해 전사체의 핵 엑스포트를 개선하도록 최적화될 수 있다. 발현 벡터에 사용하기 위한 핵 엑스포트 신호는 모든 목적을 위해 본원에 참고로 포함되는 문헌에 기재되어 있다(Callendret 등, Virology. 2007 Jul 5; 363(2): 288-302). 폴리뉴클레오티드 서열은, 예를 들어 주어진 유기체의 평균 GC 함량을 반영하기 위해 GC 함량과 관련하여 최적화될 수 있다. 서열 최적화는 전사, 번역, 전사 후 처리 및/또는 RNA 안정성과 같은 하나 이상의 서열 특성의 균형을 맞출 수 있다. 서열 최적화는 전사, 번역, 전사 후 처리 및 RNA 안정성 각각의 균형을 유지하는 최적의 서열을 생성할 수 있다. 서열 최적화 알고리즘은 진아트(GeneArt)(Thermo Fisher), 코돈 최적화 툴(Codon Optimization Tool)(IDT), 쿨 툴(Cool Tool)(University of Singapore), SGI-DNA(La Jolla California)와 같은 당업자에게 알려져 있다. 항원-코딩 단백질의 하나 이상의 영역은 별도로 서열 최적화될 수 있다.

[0236] 또 다른 측면은 폴리펩티드 또는 이의 일부를 발현할 수 있는 발현 벡터를 제공한다. 상이한 세포 유형에 대한 발현 벡터는 당업계에 잘 알려져 있고 과도한 실험 없이 선택될 수 있다. 일반적으로 DNA는 플라스미드와 같은 발현 벡터에 적절한 방향과 발현을 위한 정확한 해독틀로 삽입된다. 필요한 경우, DNA는 원하는 숙주에 의해 인식되는 적절한 전사 및 번역 조절 제어 뉴클레오티드 서열에 연결될 수 있지만, 이러한 제어는 일반적으로 발현 벡터에서 이용가능하다. 그런 다음 벡터는 표준 기술을 통해 숙주에 도입된다. 지침은 문헌에서 찾을 수 있다(Sambrook 등 (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y).

[0237] **IV. 백신 조성물**

[0238] 또한 특이적 면역 반응, 예를 들어 감염성 질병 유기체-특이적 면역 반응을 일으킬 수 있는 면역원성 조성물, 예를 들어 백신 조성물이 본원에 개시되어 있다. 백신 조성물은 전형적으로, 예를 들어 본원에 기재된 방법을 사용하여 선택되거나 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드, 및/또는 기생충 유래 펩티드로부터 선택된 하나 또는 다수의 항원을 포함한다. 백신 조성물은 또한 백신으로 지칭될 수 있다.

[0239] 백신은 1 내지 30개의 펩티드, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개의 상이한 펩티드, 6, 7, 8, 9, 10, 11, 12, 13, 또는 14개의 상이한 펩티드, 또는 12, 13 또는 14 상이한 펩티드를 함유할 수 있다. 펩티드는 번역 후 변형을 포함할 수 있다. 백신은 1 내지 100개 이상의 뉴클레오티드 서열, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67,

68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 그 이상의 상이한 뉴클레오티드 서열, 6, 7, 8, 9, 10 11, 12, 13, 또는 14개의 상이한 뉴클레오티드 서열, 또는 12, 13 또는 14개의 상이한 뉴클레오티드 서열을 포함할 수 있다. 백신에는 1 내지 30개의 항원 서열, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 그 이상의 상이한 항원 서열, 6, 7, 8, 9, 10, 11, 12, 13, 또는 14개의 상이한 항원 서열, 또는 12, 13 또는 14개의 상이한 항원 서열이 포함될 수 있다.

[0240] 백신은 1 내지 30개의 항원-코딩 핵산 서열, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 그 이상의 상이한항원-코딩 핵산 서열, 6, 7, 8, 9, 10 11, 12, 13 또는 14개의 상이한 항원-코딩 핵산 서열, 또는 12, 13 또는 14개의 상이한 항원-코딩 핵산 서열을 포함할 수 있다. 항원-코딩 핵산 서열은 항원 "카세트"의 항원-코딩 부분을 지칭할 수 있다. 항원 카세트의 특징이 본원에 설명되어 있다. 항원-코딩 핵산 서열은 하나 이상의 에피토프-코딩 핵산 서열(예를 들어, 연결된 T 세포 에피토프를 코딩하는 항원-코딩 핵산 서열)을 함유할 수 있다.

[0241] 백신은 1 내지 30개의 별개의 에피토프-코딩 핵산 서열, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 그 이상의 별개의 에피토프-코딩 핵산 서열, 6, 7, 8, 9, 10 11, 12, 13, 또는 14개의 별개의 에피토프-코딩 핵산 서열, 또는 12, 13 또는 14개의 별개의 에피토프-코딩 핵산 서열을 함유할 수 있다. 에피토프-코딩 핵산 서열은 연결된 T 세포 에피토프를 코딩하는 항원-코딩 핵산 서열에서 각각의 T 세포 에피토프와 같은 개별 에피토프 서열에 대한 서열을 지칭할 수 있다.

[0242] 백신은 에피토프-코딩 핵산 서열의 적어도 2개의 반복을 포함할 수 있다. 본원에 사용된 "반복"은 항원-코딩 핵산 서열내에서 동일한 핵산 에피토프-코딩 핵산 서열(본원에 기재된 임의의 5' 링커 서열 및/또는 임의의 3' 링커 서열 포함)의 2회 이상의 반복을 의미한다. 한 예에서, 카세트의 항원-코딩 핵산 서열 부분은 에피토프-코딩 핵산 서열의 적어도 2개의 반복부를 코딩한다. 추가의 비제한적 예에서, 카세트의 항원-코딩 핵산 서열 부분은 하나 이상의 별개의 에피토프를 코딩하고, 별개의 에피토프 중 적어도 하나는 별개의 에피토프를 코딩하는 핵산 서열의 적어도 2개의 반복부에 의해 코딩된다(즉, 적어도 2개의 별개의 에피토프-코딩 핵산 서열). 예시적인 비제한적 예에서, 항원-코딩 핵산 서열은 에피토프-코딩 핵산 서열 에피토프-코딩 서열 A(E_A), 에피토프-코딩 서열 B(E_B) 및 에피토프-코딩 서열C(E_C)에 의해 코딩된 에피토프 A, B 및 C를 코딩하고, 및 별개의 에피토프 중 적어도 하나의 반복을 갖는 예시적인 항원-코딩 핵산 서열은 하기 식에 의해 예시되지만 이에 제한되지는 않는다:

[0243] - 하나의 별개의 에피토프의 반복(에피토프 A의 반복):

[0244] E_A-E_B-E_C-E_A; 또는

[0245] E_A-E_A-E_B-E_C

[0246] - 다중 별개의 에피토프의 반복(에피토프 A, B 및 C의 반복):

[0247] E_A-E_B-E_C-E_A-E_B-E_C; 또는

[0248] E_A-E_A-E_B-E_B-E_C-E_C

[0249] - 다중 별개의 에피토프의 다중 반복(에피토프 A, B 및 C의 반복):

[0250] E_A-E_B-E_C-E_A-E_B-E_C-E_A-E_B-E_C; 또는

[0251] E_A-E_A-E_A-E_B-E_B-E_B-E_C-E_C-E_C

[0252] 상기 예는 제한적이지 않으며, 별개의 에피토프 중 적어도 하나의 반복을 갖는 항원-코딩 핵산 서열은 임의의 순서 또는 빈도로 별개의 에피토프 각각을 코딩할 수 있다. 예를 들어, 순서와 빈도는 식 E_A-E_B-E_C-E_C-E_A-E_B-E_A-E_C-E_A-E_C-E_C-E_B에 의한, 예를 들어 에피토프 A, B 및 C가 있는 예에서 별개의 에피토프의 무작위 배열일 수 있다.

[0253] 또한 본원은 5'에서 3'으로 하기 식으로 기술된 적어도 하나의 항원-코딩 핵산 서열을 갖는 항원-코딩 카세트를 제공한다:

[0254] $(E_x-(E^N_n)_y)_z$

[0255] 여기서, E는 적어도 하나의 별개의 에피토프-코딩 핵산 서열 중 적어도 하나를 포함하는 뉴클레오티드 서열을 나타내고,

[0256] n은 분리된 별개의 에피토프-코딩 핵산 서열의 수를 나타내고 0을 포함하는 임의의 정수이고,

[0257] E^N은 각각의 상응하는 n에 대한 분리된 별개의 에피토프-코딩 핵산 서열을 포함하는 뉴클레오티드 서열을 나타내고,

[0258] z의 각 반복에 대해 x = 0 또는 1, 각 n에 대해 y = 0 또는 1, x 또는 y 중 적어도 하나는 = 1, 및 z = 2 또는 이상, 여기서 항원-코딩 핵산 서열은 E, 주어진 E^N 또는 이들의 조합의 2회 이상의 반복을 포함한다.

[0259] 각각의 E 또는 E^N은 독립적으로 본원에 기재된 임의의 에피토프-코딩 핵산 서열(예를 들어, 감염성 질병 T 세포 에피토프를 코딩하는 펩티드)을 포함할 수 있다. 예를 들어, 각각의 E 또는 E^N은 식 ($L5_b$-N_c-$L3_d$)에 의해 5'에서 3'으로 기술된 뉴클레오티드 서열을 독립적으로 포함할 수 있으며, 여기서 N은 각 E 또는 E^N과 관련된 별개의 에피토프-코딩 핵산 서열을 포함하고, 여기서 c = 1, L5는 5' 링커 서열을 포함하고, 여기서 b= 0 또는 1이고, L3은 3' 링커 서열을 포함하고, 여기서 d= 0 또는 1이다. 사용될 수 있는 에피토프 및 링커는 본원에 추가로 기재되어 있다.

[0260] 에피토프-코딩 핵산 서열(임의의 5' 링커 서열 및/또는 임의의 3' 링커 서열 포함)의 반복은 서로 선형으로 직접 연결될 수 있다(예를 들어, 위에 예시된 바와 같이 E_A-E_A-…). 에피토프-코딩 핵산 서열의 반복은 하나 이상의 추가 뉴클레오티드 서열에 의해 분리될 수 있다. 일반적으로, 에피토프-코딩 핵산 서열의 반복은 본원에 기재된 조성물에 적용가능한 임의의 크기의 뉴클레오티드 서열에 의해 분리될 수 있다. 한 예에서, 에피토프-코딩 핵산 서열의 반복은 분리된 별개의 에피토프-코딩 핵산 서열(예를 들어, 상기 예시된 바와 같이 E_A-E_B-E_C-E_A…)에 의해 분리될 수 있다. 반복이 단일 분리된 별개의 에피토프-코딩 핵산 서열에 의해 분리되고 각각의 에피토프-코딩 핵산 서열(임의로 5' 링커 서열 및/또는 임의로 3' 링커 서열 포함)이 25개 아미노산 길이의 펩티드를 코딩하는 예에서, 반복은 E_A-E_B-E_A...로 표시되는 항원-코딩 핵산과 같이 75개 뉴클레오티드로 분리될 수 있고, E_A는 75개 뉴클레오티드에 의해 분리된다. 예시적인 실시예에서, 25mer 항원 Trp1 (VTNTEMFVTAPDNLGYMYEVQWPGQ) 및 Trp2(TQPQIANCSVYDFFVWLHYYSVRDT)의 서열 VTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDTVTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDT 코딩 반복부를 갖는 항원-코딩 핵산에서, Trp1의 반복부는 25mer Trp2에 의해 분리되고, 따라서 Trp1 에피토프-코딩 핵산 서열의 반복부는 75개 뉴클레오티드 Trp2 에피토프-코딩 핵산 서열로 분리된다. 반복이 2, 3, 4, 5, 6, 7, 8 또는 9개의 분리된 별개의 에피토프-코딩 핵산 서열에 의해 분리되는 예에서, 각각의 에피토프-코딩 핵산 서열(임의의 5' 링커 서열 및/또는 또는 임의의 3' 링커 서열 포함)은 25개 아미노산 길이의 펩티드를 코딩하며, 반복는 각각 150, 225, 300, 375, 450, 525, 600 또는 675개의 뉴클레오티드로 분리될 수 있다.

[0261] 한 구현양태에서, 펩티드 및/또는 폴리펩티드가 상이한 MHC 클래스 I 분자 및/또는 상이한 MHC 클래스 II분자와 같은 상이한 MHC 분자와 관련될 수 있도록 상이한 펩티드 및/또는 폴리펩티드 또는 이를 코딩하는 뉴클레오티드 서열이 선택된다. 일부 측면에서, 하나의 백신 조성물은 가장 빈번하게 발생하는 MHC 클래스 I 분자 및/또는 상이한 MHC 클래스 II 분자와 관련될 수 있는 펩티드 및/또는 폴리펩티드에 대한 코딩 서열을 포함한다. 따라서, 백신 조성물은 적어도 2개의 바람직한, 적어도 3개의 바람직한, 또는 적어도 4개의 바람직한 MHC 클래스 I 분자 및/또는 상이한 MHC 클래스 II 분자와 관련될 수 있는 상이한 단편을 포함할 수 있다.

[0262] 백신 조성물은 특이적 세포독성 T-세포 반응 및/또는 특이적 헬퍼 T-세포 반응을 일으킬 수 있다.

[0263] 백신 조성물은 특이적 B-세포 반응(예를 들어, 항체 반응)을 일으킬 수 있다.

[0264] 백신 조성물은 특이적 세포독성 T-세포 반응, 특이적 헬퍼 T-세포 반응, 및/또는 특이적 B-세포 반응(예를 들어, 항체 반응)을 일으킬 수 있다. 백신 조성물은 특이적 세포독성 T-세포 반응, 특이적 헬퍼 T-세포 반응, 및 특이적 B-세포 반응(예를 들어, 항체 반응)을 일으킬 수 있다.

[0265] 백신 조성물은 보조제 및/또는 담체를 추가로 포함할 수 있다. 유용한 보조제 및 담체의 예는 하기에 제공된다. 조성물은 예를 들어, T-세포에 펩티드를 제시할 수 있는 수지상 세포(DC)와 같은 단백질 또는 항원-제시 세포와 같은 담체와 관련될 수 있다.

[0266] 보조제는 백신 조성물로의 이의 혼합이 항원에 대한 면역 반응을 증가시키거나 달리 변형시키는 임의의 물질이다. 담체는 항원이 관련될 수 있는 스캐폴드 구조, 예를 들어 폴리펩티드 또는 다당류일 수 있다. 임의로, 보조제는 공유적으로 또는 비공유적으로 접합된다.

[0267] 항원에 대한 면역 반응을 증가시키는 보조제의 능력은 일반적으로 면역-매개 반응의 상당하거나 실질적인 증가 또는 질병 증상의 감소에 의해 나타난다. 예를 들어, 체액성 면역의 증가는 전형적으로 항원에 대해 증가된 항체 역가의 상당한 증가에 의해 나타나고, T-세포 활성의 증가는 전형적으로 증가된 세포 증식, 또는 세포 세포독성 또는 사이토카인 분비로 나타난다. 보조제는 또한, 예를 들어 주로 체액성 또는 Th 반응을 주로 세포성 또는 Th 반응으로 변경함으로써 면역 반응을 변경할 수 있다.

[0268] 적합한 보조제는 1018 ISS, 명반, 알루미늄 염, 앰플리백스(Amplivax), AS15, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, 이미퀴모드(Imiquimod), 이뮤팩트(ImuFact)IMP321, IS 패치, ISS, ISCOMATRIX, 쥬브이뮨(JuvImmune), 리포백(LipoVac), MF59, 모노포스포릴 지질 A, 몬타니드(Montanide) IMS 1312, 몬타니드 ISA 206, 몬타니드 ISA 50V, 몬타니드 ISA-51, OK-432, OM-174, OM-197-MP-EC, ONTAK, PepTel 벡터 시스템, PLG 마이크로입자, 레시퀴모드, SRL172, 비로솜(Virosomes) 및 기타 바이러스-유사 입자, YF-17D, VEGF 트랩, R848, 베타-글루칸, Pam3Cys, 사포닌, 마이코박테리아 추출물 및 합성 박테리아 세포벽 모방체에서 유래된 아퀼라(Aquila's) QS21 스티뮬론(Aquila Biotech, Worcester, Mass., USA), 및 리비 데톡스(Ribi's Detox), 퀼(Quil) 또는 슈퍼포스(Superfos)와 같은 기타 독점 보조제를 포함하나 이에 제한되지는 않는다. 불완전 프로인트 또는 GM-CSF와 같은 보조제가 유용하다. 수지상 세포 및 이의 제조에 특이적인 여러 면역학적 보조제(예를 들어, MF59)는 이전에 기술된 바 있다(Dupuis M, 등, Cell Immunol. 1998; 186(1):18-27; Allison A C; Dev Biol Stand. 1998; 92:3-11). 또한 사이토카인을 사용할 수 있다. 여러 사이토카인은 수지상 세포가 림프 조직(예를 들어, TNF-알파)으로 이동하는 데 영향을 미치도록 직접적으로 연결되어 수지상 세포가 T-림프구(예를 들어, GM-CSF, IL-1 및 IL-4) (전체가 본원에 참조로 포함되는 U.S.특허 5,849,589)에 대한 효율적인 항원-제시 세포로 성숙하는 것을 가속화하고 면역보조제(예를 들어, IL-12)로서 작용한다(Gabrilovich D I, 등, J Immunother Emphasis Tumor Immunol. 1996 (6):414-418).

[0269] CpG 면역자극성 올리고뉴클레오티드는 또한 백신 환경에서 보조제의 효과를 향상시키는 것으로 보고되었다. RNA 결합 TLR 7, TLR 8 및/또는 TLR 9와 같은 다른 TLR 결합 분자가 또한 사용될 수 있다.

[0270] 유용한 보조제의 다른 예에는 화학적으로 변형된 CpG(예를 들어, CpR, Idera), 폴리(I:C)(예를 들어, polyi:CI2U), 비-CpG 박테리아 DNA 또는 RNA 및 면역활성 작은 분자 및 사이클로포스파미드, 수니티닙, 베바시주맙, 셀레브렉스, NCX-4016, 실데나필, 타다라필, 바르데나필, 소라피닙, XL-999, CP-547632, 파조파닙, ZD2171, AZD2171, 이필리무맙, 트레멜리무맙 및 SC58175와 같은 항체를 포함하지만 이로 제한되지 않으며, 치료적으로 및/또는 보조제로서 작용할 수 있다. 보조제 및 첨가제의 양 및 농도는 과도한 실험 없이 숙련된 기술자에 의해 용이하게 결정될 수 있다. 추가 보조제는 과립구 대식세포 집락 자극 인자(GM-CSF, 사그라모스팀)와 같은 집락-자극 인자를 포함한다.

[0271] 백신 조성물은 하나 이상의 상이한 보조제를 포함할 수 있다. 또한, 치료 조성물은 상기 중 임의의 것 또는 이들의 조합을 포함하는 임의의 보조제 물질을 포함할 수 있다. 백신 및 보조제가 함께 또는 임의의 적절한 순서로 별도로 투여될 수 있음이 또한 고려된다.

[0272] 담체(또는 부형제)는 보조제와 독립적으로 존재할 수 있다. 담체의 기능은, 예를 들어 활성 또는 면역원성을 증가시키거나, 안정성을 부여하거나, 생물학적 활성을 증가시키거나, 혈청 반감기를 증가시키기 위해 특정 돌연변이체의 분자량을 증가시키는 것일 수 있다. 또한, 담체는 펩티드를 T 세포에 제시하는 데 도움이 될 수 있다.

담체는 당업자에게 공지된 임의의 적합한 담체, 예를 들어 단백질 또는 항원 제시 세포일 수 있다. 담체 단백질은 키홀 림펫 헤모시아닌, 트랜스페린과 같은 혈청 단백질, 소 혈청 알부민, 인간 혈청 알부민, 티로글로불린 또는 오브알부민, 면역글로불린, 또는 호르몬, 예컨대 인슐린 또는 팔미트산일 수 있지만 이에 제한되지는 않는다. 인간의 면역화를 위해, 담체는 일반적으로 인간에게 허용되고 안전한 생리학적으로 허용되는 담체이다. 그러나 파상풍 톡소이드 및/또는 디프테리아 톡소이드가 적합한 담체이다. 대안적으로, 담체는 덱스트란, 예를 들어 세파로스일 수 있다.

[0273] 세포독성 T-세포(CTL)는 온전한 외래 항원 자체보다는 MHC 분자에 결합된 펩티드 형태의 항원을 인식한다. MHC 분자 자체는 항원 제시 세포의 세포 표면에 위치한다. 따라서, 펩티드 항원, MHC 분자 및 APC의 삼량체 복합체가 존재하는 경우, CTL의 활성화가 가능하다. 상응하게, 펩티드가 CTL의 활성화를 위해 사용될 뿐만 아니라 추가로 각각의 MHC 분자와 함께 APC가 추가되는 경우, 면역 반응을 향상시킬 수 있다. 따라서, 일부 구현양태에서 백신 조성물은 적어도 하나의 항원 제시 세포를 추가로 함유한다.

[0274] 항원은 또한 백시니아, 계두, 자가-복제 알파바이러스, 마라바바이러스, 아데노바이러스(Tatsis 등, Adenoviruses, *Molecular Therapy* (2004) 10, 616-629) 또는 2세대, 3세대 또는 하이브리드 2세대/3세대 렌티바이러스 및 특정 세포 유형 또는 수용체를 표적으로 하도록 설계된 모든 세대의 재조합 렌티바이러스를 포함하지만 이에 제한되지 않는 렌티바이러스와 같은 바이러스 벡터-기반 백신 플랫폼에 포함될 수 있다(Hu 등, Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, *Immunol Rev.* (2011) 239(1): 45-61, Sakuma 등, Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper 등, Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey 등, Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880)). 위에서 언급한 바이러스 벡터-기반 백신 플랫폼의 패키징 능력에 따라, 이 접근법은 하나 이상의 항원 펩티드를 코딩하는 하나 이상의 뉴클레오티드 서열을 전달할 수 있다. 서열은 돌연변이되지 않은 서열에 의해 플랭킹될 수 있거나, 링커에 의해 분리될 수 있거나, 세포내 구획을 표적화하는 하나 이상의 서열이 선행될 수 있다(Gros 등, Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen 등, Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu 등, Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10). 숙주에 도입되면, 감염된 세포는 항원을 발현하고, 이에 의해 펩티드(들)에 대한 숙주 면역(예를 들어, CTL) 반응을 유도한다. 면역화 프로토콜에 유용한 백시니아 벡터 및 방법은, 예를 들어 U.S. 특허 제4,722,848호에 기재되어 있다. 또 다른 벡터는 BCG(Bacille Calmette Guerin)이다. BCG 벡터는 문헌에 기재되어 있다(Stover 등, Nature 351:456-460 (1991)). 항원의 치료적 투여 또는 면역화에 유용한 매우 다양한 다른 백신 벡터, 예를 들어 살모넬라 타이피 벡터 등은 본원의 설명으로부터 당업자에게 명백할 것이다.

[0275] **IV.A. 항원 카세트**

[0276] 하나 이상의 항원의 선택, "카세트"의 클로닝 및 구성 및 바이러스 벡터로의 삽입에 사용되는 방법은 본원에 제공된 교시가 주어진 당업계의 기술 범위 내에 있다. "항원 카세트" 또는 "카세트"는 선택된 항원 또는 다수의 항원(예를 들어, 항원-코딩 핵산 서열) 및 항원(들)을 전사하고 전사된 생성물을 발현하는 데 필요한 다른 조절 요소의 조합을 의미한다. 선택된 항원 또는 다수의 항원은 별개의 에피토프 서열을 지칭할 수 있으며, 예를 들어 카세트 내의 항원-코딩 핵산 서열은 에피토프가 전사되고 발현되도록 에피토프-코딩 핵산 서열(또는 다수의 에피토프-코딩 핵산 서열)을 코딩할 수 있다. 항원 또는 다수의 항원은 전사를 허용하는 방식으로 조절 구성요소에 작동가능하게 연결될 수 있다. 이러한 구성요소는 바이러스 벡터로 형질주입된 세포에서 항원(들)의 발현을 유도할 수 있는 통상적인 조절 요소를 포함한다. 따라서 항원 카세트는 또한 항원(들)에 연결되고 다른 임의의 조절 요소와 함께 재조합 벡터의 선택된 바이러스 서열 내에 위치하는 선택된 프로모터를 함유할 수 있다. 카세트는 하나 이상의 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드, 및/또는 기생충-유래 펩티드와 같은 하나 이상의 항원을 포함할 수 있다. 카세트는 분리된 프로모터에 각각 독립적으로 작동가능하게 연결되고 및/또는 2A 리보솜 스키핑 서열 요소(예를 들어, E2A, P2A, F2A, 또는 T2A 서열) 또는 내부 리보솜 진입 부위(IRES) 서열 요소와 같은 다른 멀티시스트론 시스템을 사용하여 함께 연결되는 다중 항원-코딩 핵산 서열을 함유하는 카세트와 같은 하나 이상의 항원-코딩 핵산 서열을 가질 수 있다. 링커는 또한 TEV 또는 푸린 절단 부위와 같은 절단 부위를 가질 수 있다. 절단 부위가 있는 링커는 멀티시스트론 시스

템과 같은 다른 요소와 조합하여 사용할 수 있다. 비제한적인 예시적인 예에서, 푸린 프로테아제 절단 부위는 푸린 프로테아제 절단 부위가 번역 후 2A 서열의 제거를 용이하게 하도록 구성되도록 2A 리보솜 스키핑 서열 요소와 함께 사용될 수 있다. 하나 이상의 항원-코딩 핵산 서열을 함유하는 카세트에서, 각각의 항원-코딩 핵산 서열은 하나 이상의 에피토프-코딩 핵산 서열(예를 들어, 연결된 T 세포 에피토프를 코딩하는 항원-코딩 핵산 서열)을 함유할 수 있다.

[0277] 유용한 프로모터는 구성적 프로모터 또는 조절된(유도성) 프로모터일 수 있으며, 이는 발현될 항원(들)의 양을 제어할 수 있게 한다. 예를 들어, 바람직한 프로모터는 거대세포바이러스 즉시 초기 프로모터/인핸서의 프로모터이다(Boshart 등, Cell, 41:521-530 (1985)). 또 다른 바람직한 프로모터는 라우스 육종 바이러스 LTR 프로모터/인핸서를 포함한다. 또 다른 프로모터/인핸서 서열은 닭 세포질 베타-액틴 프로모터이다(T. A. Kost 등, Nucl. Acids Res., 11(23):8287 (1983)). 다른 적합하거나 바람직한 프로모터는 당업자에 의해 선택될 수 있다.

[0278] 항원 카세트는 또한 전사체(폴리(A), 폴리-A 또는 pA)의 효율적인 폴리아데닐화를 위한 신호를 제공하는 서열 및 기능적 스플라이스 공여자 및 수용체 부위를 갖는 인트론을 포함하는 바이러스 벡터 서열에 이종성인 핵산 서열을 포함할 수 있다. 본 발명의 예시적인 벡터에 사용되는 일반적인 폴리-A 서열은 파포바바이러스 SV-40에서 유래된 것이다. 폴리-A 서열은 일반적으로 항원-기반 서열 다음에 및 바이러스 벡터 서열 전에 카세트에 삽입될 수 있다. 일반적인 인트론 서열도 SV-40에서 유래될 수 있으며 SV-40 T 인트론 서열로 지칭된다. 항원 카세트는 또한 프로모터/인핸서 서열 및 항원(들) 사이에 위치한 이러한 인트론을 함유할 수 있다. 이들 및 기타 공통 벡터 요소의 선택은 통상적이고(Sambrook 등, "Molecular Cloning. A Laboratory Manual.", 2d edit., Cold Spring Harbor Laboratory, New York (1989) 및 그 안에 인용된 참고 문헌 참조) 및 많은 이러한 서열은 젠뱅크(Genbank) 뿐만 아니라 상업적 및 산업적 소스로부터 입수가능하다.

[0279] 항원 카세트는 하나 이상의 항원을 가질 수 있다. 예를 들어, 주어진 카세트는 1-10, 1-20, 1-30, 10-20, 15-25, 15-20, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 그 이상의 항원을 포함할 수 있다. 항원은 서로 직접 연결될 수 있다. 항원은 또한 링커로 서로 연결될 수 있다. 항원은 N에서 C 또는 C에서 N을 포함하여 서로에 대해 임의의 방향일 수 있다.

[0280] 다른 곳에서 기술된 바와 같이, 항원 카세트는 ChAd-기반 벡터의 E1 유전자 영역 결실 또는 E3 유전자 영역 결실 부위 또는 VEE 백본의 결실된 구조 단백질에서와 같은 바이러스 벡터 백본에서 임의의 선택된 결실 부위에 위치할 수 있다.

[0281] 항원 카세트는 각 요소의 정렬된 서열을 설명하기 위해 5'에서 3'로 다음 식을 사용하여 설명될 수 있다:

[0282] $(P_a\text{-}(L5_b\text{-}N_c\text{-}L3_d)_X)_Z\text{-}(P2_h\text{-}(G5_e\text{-}U_f)_Y)_W\text{-}G3_g$

[0283] 여기서 P 및 P2는 프로모터 뉴클레오티드 서열을 포함하고, N은 MHC 클래스 I 에피토프 코딩 핵산 서열을 포함하고, L5는 5' 링커 서열을 포함하고, L3은 3' 링커 서열을 포함하고, G5는 아미노산 링커를 코딩하는 핵산 서열을 포함하고, G3은 아미노산 링커를 코딩하는 적어도 하나의 핵산 서열 중 하나를 포함하고, U는 MHC 클래스 II 항원-코딩 핵산 서열을 포함하고, 여기서 각각의 X에 대해 상응하는 N_c는 에피토프 코딩 핵산 서열이고, 여기서 각각의 Y에 대해 상응하는 U_f는 보편적인 MHC 클래스 II 에피토프-코딩 핵산 서열이고, 임의로 적어도 하나의 보편적인 서열은 파상풍 톡소이드 및 PADRE 중 적어도 하나를 포함한다. 구성 및 정렬된 서열은 존재하는 요소의 수를 선택하여 추가로 정의할 수 있다. 예를 들어 a = 0 또는 1, b = 0 또는 1, c = 1, d = 0 또는 1, e = 0 또는 1, f = 1, g = 0 또는 1, h = 0 또는 1, X = 1 내지 400, Y = 0, 1, 2, 3, 4 또는 5, Z = 1 내지 400, W = 0, 1, 2, 3, 4 또는 5이다.

[0284] 한 예에서, 존재하는 요소는 a = 0, b = 1, d = 1, e = 1, g = 1, h = 0, X = 10, Y = 2, Z = 1, 및 W = 1을 포함하고, 추가 프로모터가 존재하지 않는 경우(예를 들어, RNA 알파바이러스 백본과 같은 벡터 백본에 의해 제공된 프로모터 뉴클레오티드 서열만이 존재), 10개의 MHC 클래스 I 에피토프가 존재하고, 각각에 N에 대해 5' 링커가 존재하고, 각각의 N에 대해 3' 링커가 존재하고, 2개의 MHC 클래스 II 에피토프가 존재하고, 2개의 MHC 클래스 II 에피토프를 연결하는 링커가 존재하고, 2개의 MHC 클래스 II 에피토프의 5' 말단을 최종 MHC 클래스 I 에피토프의 3' 링커에 연결하는 링커가 존재하고 및 2개의 MHC 클래스 II 에피토프의 3' 말단을 벡터 백본에 연결하는 링커가 존재한다. 항원 카세트의 3' 말단을 벡터 백본에 연결하는 예는 3' 19-nt CSE와 같은 벡터 백본에 의해 제공되는 3' UTR 요소에 직접 연결하는 것을 포함한다. 항원 카세트의 5' 말단을 벡터 백본에 연결하는 예는 26S 프로모터 서열, 알파바이러스 5' UTR, 51-nt CSE 또는 24-nt CSE와 같은 벡터 백본의 프로모터 또

는 5' UTR 요소에 직접 연결하는 것을 포함한다.

[0285] 다른 예는 다음을 포함한다: 여기서 a = 1은 벡터 백본에 의해 제공되는 프로모터 뉴클레오티드 서열 이외의 프로모터가 존재하는 위치를 설명하고; 여기서 a = 1이고 Z는 1보다 크며, 여기서 벡터 백본에 의해 제공되는 프로모터 뉴클레오티드 서열 이외의 다중 프로모터가 각각 존재하여 1개 이상의 별개의 MHC 클래스 I 에피토프 코딩 핵산 서열의 발현을 유도하고; 여기서 h = 1은 분리된 프로모터가 MHC 클래스 II 에피토프-코딩 핵산 서열의 발현을 유도하기 위해 존재하는 것을 설명하고; 및 여기서 g = 0은 존재하는 경우, MHC 클래스 II 에피토프-코딩 핵산 서열이 벡터 백본에 직접 연결됨을 설명한다.

[0286] 다른 예는 존재하는 존재하는 각각의 MHC 클래스 I 에피토프는 5' 링커, 3' 링커, 둘 다 갖거나 또는 둘 다를 가질 수 없는 경우를 포함한다. 하나 이상의 MHC 클래스 I 에피토프가 동일한 항원 카세트에 존재하는 예에서, 일부 MHC 클래스 I 에피토프는 5' 링커 및 3' 링커를 모두 가질 수 있는 반면, 다른 MHC 클래스 I 에피토프는 5' 링커, 3' 링커 중 하나, 또는 둘 다 가질 수 없다. 하나 이상의 MHC 클래스 I 에피토프가 동일한 항원 카세트에 존재하는 다른 예에서, 일부 MHC 클래스 I 에피토프는 5' 링커 또는 3' 링커를 가질 수 있는 반면, 다른 MHC 클래스 I 에피토프는 5' 링커, 3' 링커를 갖거나, 또는 둘 다를 가질 수 없다.

[0287] 다른 예는 존재하는 각각의 항원이 5' 링커, 3' 링커, 둘 다 또는 둘 다를 가질 수 없는 경우를 포함한다. 하나 이상의 항원이 동일한 항원 카세트에 존재하는 예에서, 일부 항원은 5' 링커 및 3' 링커를 모두 가질 수 있는 반면, 다른 항원은 5' 링커, 3' 링커를 갖거나 또는 둘 다 가질 수 없다. 하나 이상의 항원이 동일한 항원 카세트에 존재하는 다른 예에서, 일부 항원은 5' 링커 또는 3' 링커를 가질 수 있는 반면, 다른 항원은 5' 링커, 3' 링커를 갖거나 또는 둘 다 가질 수 없다.

[0288] 하나 이상의 MHC 클래스 II 에피토프가 동일한 항원 카세트에 존재하는 예에서, 일부 MHC 클래스 II 에피토프는 5' 링커 및 3' 링커를 모두 가질 수 있는 반면, 다른 MHC 클래스 II 에피토프는 5' 링커, 3' 링커를 갖거나 또는 둘 다가질 수 없다. 하나 이상의 MHC 클래스 II 에피토프가 동일한 항원 카세트에 존재하는 다른 예에서, 일부 MHC 클래스 II 에피토프는 5' 링커 또는 3' 링커를 가질 수 있는 반면, 다른 MHC 클래스 II 에피토프는 5' 링커, 3' 링커를 갖거나, 또는 둘 다 가질 수 없다.

[0289] 프로모터 뉴클레오티드 서열 P 및/또는 P2는 RNA 알파바이러스 백본과 같은 벡터 백본에 의해 제공되는 프로모터 뉴클레오티드 서열과 동일할 수 있다. 예를 들어, 벡터 백본, Pn 및 P2에 의해 제공되는 프로모터 서열은 각각 26S 서브게놈 프로모터 또는 CMV 프로모터를 포함할 수 있다. 프로모터 뉴클레오티드 서열 P 및/또는 P2는 벡터 백본에 의해 제공되는 프로모터 뉴클레오티드 서열과 상이할 수 있을 뿐만 아니라 서로 상이할 수 있다.

[0290] 5' 링커 L5는 천연 서열 또는 비-천연 서열일 수 있다. 비천연 서열은 AAY, RR 및 DPP를 포함하지만 이에 제한되지는 않는다. 3' 링커 L3은 또한 천연 서열 또는 비-천연 서열일 수 있다. 추가로, L5 및 L3은 둘 다 천연 서열일 수 있고, 둘 다 비-천연 서열이거나, 하나는 천연이고 다른 하나는 비-천연일 수 있다. 각각의 X에 대해 아미노산 링커는 길이가 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 이상의 아미노산일 수 있다. 각각의 X에 대해, 아미노산 링커는 또한 길이가 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 21개, 적어도 22개, 적어도 23개, 적어도 24개, 적어도 25개, 적어도 26개, 적어도 27개, 적어도 28개, 적어도 29개, 또는 적어도 30개 아미노산일 수 있다.

[0291] 각각의 Y에 대해 아미노산 링커 G5는 길이가 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 이상의 아미노산일 수 있다. 각각의 Y에 대해, 아미노산 링커는 또한 길이가 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 21개, 적어도 22개, 적어도 23개, 적어도 24개, 적어도 25개, 적어도 26개, 적어도 27개, 적어도 28개, 적어도 29개, 또는 적어도 30개 아미노산일 수 있다.

[0292] 아미노산 링커 G3은 길이가 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 또는 이상의 아미노산일 수 있다. G3은 또한 길이가 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 21개, 적어도 22개, 적어도 23개, 적어도 24개, 적어도 25개, 적어도 26개, 적어도 27개, 적어도 28개, 적어도 29개, 또는 적어도 30개 아미노산일 수 있다.

[0293] 각각의 X에 대해, 각각의 N은 MHC 클래스 I 에피토프, MHC 클래스 II 에피토프, B 세포 반응을 자극할 수 있는 에피토프, 또는 이들의 조합을 코딩할 수 있다. 각각의 X에 대해, N은 MHC 클래스 I 에피토프, MHC 클래스 II 에피토프, 및 B 세포 반응을 자극할 수 있는 에피토프의 조합을 코딩할 수 있다. 각각의 X에 대해, N은 MHC 클래스 I 에피토프 및 MHC 클래스 II 에피토프의 조합을 코딩할 수 있다. 각각의 X에 대해 N은 MHC 클래스 I 에피토프 및 B 세포 반응을 자극할 수 있는 에피토프의 조합을 코딩할 수 있다. 각각의 X에 대해 N은 MHC 클래스 II 에피토프 및 B 세포 반응을 자극할 수 있는 에피토프의 조합을 코딩할 수 있다. 각각의 X에 대해, 각각의 N은 7-15개 아미노산 길이의 MHC 클래스 I 에피토프를 코딩할 수 있다. 각각의 X에 대해, 각각의 N은 또한 아미노산 길이가 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개인 MHC 클래스 I 에피토프를 코딩할 수 있다. 각각의 X에 대해 각각의 N은 아미노산 길이가 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 21개, 적어도 22개, 적어도 23개, 적어도 24개, 적어도 25개, 적어도 26개, 적어도 27개, 적어도 28개, 적어도 29개, 또는 적어도 30개인 MHC 클래스 I 에피토프를 코딩할 수도 있다. 각각의 X에 대해 각각의 N은 MHC 클래스 에피토프를 코딩할 수 있다. 각각의 X에 대해 각각의 N은 B 세포 반응을 자극할 수 있는 에피토프를 코딩할 수 있다.

[0294] 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있다. 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있고 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다(예를 들어, 면역원성 폴리펩티드를 코딩하는 2개의 별개의 SARS-CoV-2 유래 핵산 서열을 코딩한다). 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있고 적어도 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있고 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있고 적어도 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 700개 이하의 뉴클레오티드일 수 있고 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 항원을 포함할 수 있다.

[0295] 하나 이상의 항원을 코딩하는 카세트는 길이가 375-700개 뉴클레오티드일 수 있다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-700개 뉴클레오티드일 수 있고 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 인코딩하는 카세트는 길이가 375-700개 뉴클레오티드일 수 있고 적어도 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-700개 뉴클레오티드일 수 있고 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-700개 뉴클레오티드 사이이고 적어도 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-700개의 뉴클레오티드일 수 있고 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 항원을 포함한다.

[0296] 하나 이상의 항원을 코딩하는 카세트는 길이가 600, 500, 400, 300, 200 또는 100개 뉴클레오티드 이하일 수 있다. 하나 이상의 항원을 코딩하는 카세트는 길이가 600, 500, 400, 300, 200 또는 100개 이하의 뉴클레오티드일 수 있고 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 600, 500, 400, 300, 200, 또는 100개 이하의 뉴클레오티드일 수 있고 적어도 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 600, 500, 400, 300, 200, 또는 100개 이하의 뉴클레오티드일 수 있고 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 600, 500, 400, 300, 200, 또는 100개 이하의 뉴클레오티드일 수 있고 적어도 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 인코딩하는 카세트는 길이가 600, 500, 400, 300, 200 또는 100개 이하의 뉴클레오티드일 수 있고 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 항원을 포함

한다.

[0297] 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있고 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있고 적어도 2개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있고 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있고 적어도 3개의 별개의 에피토프-코딩 핵산 서열을 코딩한다. 하나 이상의 항원을 코딩하는 카세트는 길이가 375-600개, 375-500개, 또는 375-400개 뉴클레오티드일 수 있고 1-10, 1-5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 항원을 포함한다.

[0298] **IV.B. 면역 관문**

[0299] 본원에 기재된 C68 벡터 또는 본원에 기재된 알파바이러스 벡터와 같은 본원에 기재된 벡터는 적어도 하나의 항원을 코딩하는 핵산을 포함할 수 있고, 동일하거나 분리된 벡터는 면역 관문 분자에 결합하고 그 활성을 차단하는 적어도 하나의 면역 조절자(예를 들어, scFv같은 항체)를 코딩하는 핵산을 포함한다. 벡터는 항원 카세트 및 관문 억제제를 코딩하는 하나 이상의 핵산 분자를 포함할 수 있다.

[0300] 차단 또는 억제를 위해 표적화될 수 있는 예시적인 면역 관문 분자에는 CTLA-4, 4-1BB (CD137), 4-1BBL (CD137L), PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4(CD2 분자 계열에 속하며 모든 NK, γδ 및 기억 CD8+(αβ) T 세포 상에서 발현), CD160(BY55라고도 한다) 및 CGEN-15049가 포함되지만 이에 제한되지 않는다. 면역 관문 억제제는 CTLA-4, PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4, CD160, 및 CGEN-15049 중 하나 이상에 결합하여 이들의 활성을 차단하거나 억제하는 항체 또는 이의 항원 결합 단편, 또는 기타 결합 단백질을 포함한다. 예시적인 면역 관문 억제제에는 트레멜리무맙(CTLA-4 차단 항체), 항-OX40, PD-L1 모노클로날 항체(항-B7-H1; MEDI4736), 이필리무맙, MK-3475(PD-1 차단제), 니볼루맙(항-PD1 항체), CT-011(항-PD1 항체), BY55 모노클로날 항체, AMP224(항-PDL1 항체), BMS-936559(항-PDL1 항체), MPLDL3280A(항-PDL1 항체), MSB0010718C(항-PDL1 항체) 및 여보이/이필리무맙(항-CTLA-4 관문 억제제)를 포함한다. 항체-코딩 서열은 해당 분야의 통상적인 기술을 사용하여 C68과 같은 벡터 내로 조작될 수 있다. 예시적인 방법은 모든 목적을 위해 참조로 여기에 포함되는 문헌에 설명되어 있다(Fang 등, Stable antibody expression at therapeutic levels using the 2A peptide. *Nat Biotechnol.* 2005 May;23(5):584-90. Epub 2005 Apr 17).

[0301] **IV.C. 백신 설계 및 제조에 대한 추가 고려 사항**

[0302] **IV.C.1. 모든 종양 서브클론을 커버하는 펩티드 세트의 결정**

[0303] 모든 또는 대부분의 서브클론에 의해 제시되는 것을 의미하는 줄기 펩티드는 백신에 포함시키기 위해 우선 순위를 지정할 수 있다.[53] 임의로 제시될 것으로 예측되고 높은 확률로 면역원성이 있는 줄기 펩티드가 없는 경우, 또는 제시될 것으로 예측되고 높은 확률로 면역원성이 있는 것으로 예측된 줄기 펩티드의 수가 충분히 작아서 백신에 추가 비-줄기 펩티드가 포함될 수 있는 경우, 추가로 펩티드는 서브클론의 수와 정체성을 추정하고 백신에 의해 커버되는 서브클론의 수를 최대화하기 위해 펩티드를 선택함으로써 우선순위를 지정할 수 있다.[54]

[0304] **IV.C.2. 항원 우선순위 지정**

[0305] 위의 모든 항원 필터가 적용된 후, 백신 기술이 지원할 수 있는 것보다 더 많은 후보 항원이 백신 포함에 여전히 이용가능할 수 있다. 또한, 항원 분석의 다양한 측면에 대한 불확실성이 남아 있을 수 있으며 후보 백신 항원의 서로 다른 특성 간에 트레이드오프가 존재할 수 있다. 따라서, 선택 과정의 각 단계에서 미리 결정된 필터 대신에, 적어도 다음 축이 있는 공간에 후보 항원을 배치하고 통합 접근 방식을 사용하여 선택을 최적화하는 통합 다차원 모델을 고려할 수 있다.

[0306] 1. 자가-면역 또는 내성의 위험(생식계열의 위험)(일반적으로 자가-면역의 위험이 낮은 것이 바람직하다)

[0307] 2. 시퀀싱 인공물의 확률(일반적으로 인공물의 낮은 확률이 바람직하다)

[0308] 3. 면역원성의 확률(일반적으로 면역원성의 확률이 높을수록 바람직하다)

[0309] 4. 제시 확률(일반적으로 제시 확률이 높을수록 바람직하다)

[0310] 5. 유전자 발현(일반적으로 더 높은 발현이 바람직하다)

[0311] 6. HLA 유전자의 적용범위(항원 세트의 제시와 관련된 더 많은 수의 HLA 분자는 HLA 분자의 하향조절 또는 돌연변이를 통해 감염된 세포가 면역 공격을 피할 확률을 낮출 수 있다)

[0312] 7. HLA 클래스의 적용범위(HLA-I 및 HLA-II를 모두 포함하면 치료 반응 확률이 증가하고 감염성 질병 도피 확률이 감소할 수 있다)

[0313] 추가로, 임의로, 항원은 환자의 감염된 세포의 전부 또는 일부에서 손실되거나 비활성화된 HLA 대립유전자에 의해 제시될 것으로 예측되는 경우 백신접종에서 우선순위를 낮출 수 있다(예를 들어, 제외). HLA 대립유전자 손실은 체세포 돌연변이, 이형접합의 손실 또는 유전자좌의 동형접합 결실에 의해 발생할 수 있다. HLA 대립유전자 체세포 돌연변이의 검출 방법은 당업계에 잘 알려져 있다(Shukla 등, 2015). 체세포 LOH 및 동형접합 결실(HLA 유전자좌 포함)의 검출 방법도 마찬가지로 잘 설명되어 있다(Carter 등, 2012; McGranahan 등, 2017; Van Loo 등, 2010). 또한 질량 분석 데이터가 예측된 항원이 예측된 HLA 대립 유전자에 의해 제시되지 않는다는 것을 나타내는 경우 항원의 우선 순위를 낮출 수 있다.

[0314] **IV.D. 자가-증폭 RNA 벡터**

[0315] 일반적으로, 모든 자가-증폭 RNA(SAM) 벡터는 자가-복제 바이러스에서 유래한 자가-증폭 백본을 포함한다. "자가-증폭 백본"이라는 용어는 바이러스 게놈의 자가-복제를 허용하는 자가-복제 바이러스의 최소 서열(들)을 의미한다. 예를 들어, 알파바이러스의 자가-복제를 허용하는 최소 서열은 비구조적 단백질-매개 증폭을 위한 보존된 서열(예를 들어, 비구조적 단백질 1(nsP1) 유전자, nsP2 유전자, nsP3 유전자, nsP4 유전자 및/또는 또는 폴리A 서열)을 포함한다. 자가-증폭 백본은 또한 서브게놈 바이러스 RNA(예를 들어, 알파바이러스에 대한 26S 프로모터 요소)의 발현을 위한 서열을 포함할 수 있다. SAM 벡터는 포지티브-센스 또는 네거티브-센스 자가-복제 바이러스로부터 유래된 백본을 갖는 벡터와 같은 포지티브-센스 RNA 폴리뉴클레오티드 또는 네거티브-센스 RNA 폴리뉴클레오티드일 수 있다. 자가- 복제 바이러스에는 알파바이러스, 플라비바이러스(예를 들어, 건진(Kunjin) 바이러스), 홍역 바이러스 및 랍도바이러스(예를 들어, 광견병 바이러스 및 수포성 구내염 바이러스)가 포함되지만 이에 제한되지 않는다. 자가-복제 바이러스로부터 유래된 SAM 벡터 시스템의 예는 룬드스트롬(Lundstrom)(Molecules. 2018 Dec 13;23(12). pii: E3310. doi: 10.3390/molecules23123310)에 기술되어 있으며, 이는 모든 목적을 위해 본원에 참조로 포함된다.

[0316] **IV.D.1. 알파바이러스 생물학**

[0317] 알파바이러스는 토가비리과(*Togaviridae*)과의 구성원이며 포지티브-센스 단일 가닥 RNA 바이러스이다. 구성원은 일반적으로 신드비스, 로스 리버, 마야로, 치쿤구니야 및 셈리키 포레스트 바이러스와 같은 구세계 또는 동부 말 뇌염, 아우라,포트 모건 또는 베네수엘라 말 뇌염 바이러스 및 이의 파생 균주 TC-83(Strauss Microbrial Review 1994)와 같은 신세계로 분류된다. 천연 알파바이러스 게놈의 길이는 일반적으로 약 12 kb이며, 처음 2/3는 바이러스 게놈의 자가-복제를 위한 RNA 복제 복합체를 형성하는 비구조 단백질(nsP)을 코딩하는 유전자를 함유하고 마지막 3분의 1은 비리온 생산을 위한 구조 단백질을 코딩하는 서브게놈 발현 카세트를 함유한다(Frolov RNA 2001).

[0318] 알파바이러스의 모델 수명주기는 몇 가지 별개의 단계를 포함한다(Strauss Microbrial Review 1994, Jose Future Microbiol 2009). 숙주 세포에 바이러스가 부착된 후, 비리온은 세포내이입 구획 내의 막과 융합하여 결국 게놈 RNA를 세포질로 방출한다. 플러스-가닥 방향이고 5' 메틸구아닐레이트 캡 및 3' 폴리A 테일을 포함하는 게놈 RNA는 복제 복합체를 형성하는 비구조 단백질 nsP1-4를 생성하도록 번역된다. 감염 초기에 플러스-가닥은 복합체에 의해 마이너스-가닥 주형으로 복제된다. 현재 모델에서, 복제 복합체는 감염이 진행됨에 따라 추가 처리되며, 생성된 처리된 복합체는 전장의 포지티브-가닥 게놈 RNA와 구조 유전자를 함유하는 26S 서브게놈 포지티브-가닥 RNA로 마이너스-가닥의 전사로 전환된다. 알파바이러스의 여러 보존 서열 요소(CSE)는 다음을 포함한 다양한 RNA 복제 단계에서 잠재적으로 역할을 하는 것으로 확인되었다; 마이너스-가닥 주형에서 플러스-가닥 RNA 복제에서 5' UTR의 보체, 게놈 주형에서 마이너스-가닥 합성 복제에서 51-nt CSE, 마이너스 -가닥의 서브게놈 RNA 전사에서 nsP 및 26S RNA 사이의 접합 영역에 있는 24-nt CSE, 및 플러스-가닥 주형에서 마이너스-가닥 합성에 있는 3'19-nt CSE.

[0319] 다양한 RNA 종의 복제 후, 바이러스 입자는 일반적으로 바이러스의 자연적 수명 주기에서 조립된다. 26S RNA는

번역되고 생성된 단백질은 캡시드 단백질, 당단백질 E1 및 E2, 및 두 개의 작은 폴리펩티드 E3 및 6K를 포함하는 구조적 단백질을 생산하기 위해 추가 처리된다(Strauss 1994). 바이러스 RNA의 캡시드(encapsidation)가 발생하며 일반적으로 게놈 RNA에만 특이적인 캡시드 단백질이 팩키징된 후 비리온 조립 및 막 표면의 출아가 뒤따른다.

[0320] **IV.D.2. 전달 벡터로서의 알파바이러스**

[0321] 알파바이러스(알파바이러스 서열, 특징 및 기타 요소 포함)를 사용하여 알파바이러스-기반 전달 벡터(알파바이러스 벡터, 알파바이러스 바이러스 벡터, 알파바이러스 백신 벡터, 자가-복제 RNA(srRNA) 벡터, 또는 자가-증폭 RNA(samRNA) 벡터라고도 한다)를 생성할 수 있다. 알파바이러스는 이전에 발현 벡터 시스템으로 사용하도록 조작되었다(Pushko 1997, Rheme 2004). 알파바이러스는 특히 이종 항원 발현이 요구될 수 있는 백신 환경에서 몇 가지 이점을 제공한다. 숙주 세포질에서 자가-복제하는 능력으로 인해, 알파바이러스 벡터는 일반적으로 높은 수준의 이종 항원 생산을 초래하는 세포 내에서 높은 카피 수의 발현 카세트를 생산할 수 있다. 또한, 벡터는 일반적으로 일시적이어서 생물학적 안전성이 향상되고 벡터에 대한 면역학적 내성의 유도가 감소한다. 일반 대중은 또한 인간 아데노바이러스와 같은 다른 표준 바이러스 벡터와 비교하여 알파바이러스 벡터에 대한 기존 면역성이 부족하다. 알파바이러스 기반 벡터는 또한 일반적으로 감염된 세포에 대한 세포독성 반응을 초래한다. 세포독성은 발현된 이종 항원에 대한 면역 반응을 적절하게 불법화하는 백신 환경에서 어느 정도 중요할 수 있다. 그러나 원하는 세포독성의 정도는 균형을 이루는 작용이 될 수 있으므로 VEE의 TC-83 균주를 포함하여 여러 약독화된 알파바이러스가 개발되었다. 따라서, 본원에 기술된 항원 발현 벡터의 예는 높은 수준의 항원 발현을 허용하고, 항원에 대한 강력한 면역 반응을 유도하고, 벡터 자체에 대한 면역 반응을 유도하지 않고, 안전한 방식으로 사용될 수 있는 알파바이러스 백본을 이용할 수 있다. 더욱이, 항원 발현 카세트는 VEE 또는 이의 약독화된 유도체 TC-83으로부터 유래된 서열을 포함하지만 이에 제한되지 않는 벡터가 사용하는 알파바이러스 서열의 최적화를 통해 면역 반응의 상이한 수준을 유도하도록 설계될 수 있다.

[0322] 알파바이러스 서열을 사용하여 여러 발현 벡터 디자인 전략이 조작되었다(Pushko 1997). 한 전략에서, 알파바이러스 벡터 디자인은 구조 단백질 유전자의 다운스트림에 26S 프로모터 서열 요소의 두 번째 카피를 삽입한 다음 이종 유전자를 삽입하는 것을 포함한다(Frolov 1993). 따라서, 천연 비-구조 및 구조 단백질에 더하여, 이종 단백질을 발현하는 추가적인 서브게놈 RNA가 생성된다. 이 시스템에는 감염성 비리온 생산을 위한 모든 요소가 존재하므로 비-감염 세포에서 발현 벡터의 반복적인 감염이 발생할 수 있다.

[0323] 다른 발현 벡터 디자인은 헬퍼 바이러스 시스템을 사용한다(Pushko 1997). 이 전략에서 구조 단백질은 이종 유전자로 대체된다. 따라서, 여전히 온전한 비-구조적 유전자에 의해 매개되는 바이러스 RNA의 자가-복제 후, 26S 서브게놈 RNA는 이종 단백질의 발현을 제공한다. 전통적으로, 구조 단백질을 발현하는 추가 벡터는 감염성 바이러스를 생산하기 위해 세포주의 동시-형질주입과 같이 트랜스(*in trans*)로 공급된다. 시스템은 USPN 8,093,021에 자세히 설명되어 있으며, 이는 모든 목적을 위해 전체가 참조로 본원에 포함된다. 헬퍼 벡터 시스템은 감염성 입자 형성 가능성을 제한하는 이점을 제공하여 생물학적 안전성을 향상시킨다. 또한, 헬퍼 벡터 시스템은 전체 벡터 길이를 줄여 잠재적으로 복제 및 발현 효율을 향상시킨다. 따라서, 본원에 기술된 항원 발현 벡터의 예는 구조 단백질이 항원 카세트로 대체된 알파바이러스 백본을 이용할 수 있으며, 생성된 벡터는 생물학적 안전성 문제를 감소시키는 동시에 전체 발현 벡터 크기의 감소로 인한 효율적인 발현을 촉진한다.

[0324] **IV.D.3. 시험관내 자가-증폭 바이러스 생산**

[0325] RNA 생산을 위한 당업계에 잘 알려진 편리한 기술은 시험관내 전사(IVT)이다. 이 기술에서, 원하는 벡터의 DNA 주형은 클로닝, 제한 분해, 결찰, 유전자 합성 및 중합효소 연쇄 반응(PCR)과 같은 표준 분자 생물학 기술을 포함하는 당업자에게 잘 알려진 기술에 의해 먼저 생성된다.

[0326] DNA 주형은 RNA(예를 들어, SAM)로 전사되기를 원하는 서열의 5' 말단에 RNA 중합효소 프로모터를 포함한다. 프로모터는 T3, T7, K11 또는 SP6과 같은 박테리오파지 중합효소 프로모터를 포함하지만 이에 제한되지 않는다. 선택한 특정 RNA 중합효소 프로모터 서열에 따라, 원하는 서열 외에 추가 5' 뉴클레오티드가 전사될 수 있다. 예를 들어, 표준 T7 프로모터는 TAATACGACTCACTATAGG 서열로 지칭될 수 있으며, 여기서 원하는 서열 N의 생성을 위해 DNA 주형 TAATACGACTCACTATAGGN을 사용하는 IVT 반응은 mRNA 서열 GG-N을 생성할 것이다. 일반적으로 이론에 얽매이지 않고, T7 중합효소는 구아노신으로 시작하는 RNA 전사체를 보다 효율적으로 전사한다. 추가 5' 뉴클레오티드가 필요하지 않은 경우(예를 들어, 추가 GG 없음), DNA 주형에 함유된 RNA 중합효소 프로모터는, 예를 들어 SAM 벡터가 유래된 자가-복제 바이러스의 천연 5' 서열을 가지고 있는 SAM과 같은 원하는 서열의 5' 뉴클레오티드만 함유하는 전사체를 생성하는 서열일 수 있다. 예를 들어, 최소 T7 프로모터는

TAATACGACTCACTATA서열로 지칭될 수 있으며, 여기서 원하는 서열 N의 생성을 위해 DNA 주형 TAATACGACTCACTATAN을 사용하는 IVT 반응은 mRNA 서열 N을 생성할 것이다. 마찬가지로, 최소 SP6 프로모터는 ATTTAGGTGACACTATA 서열을 사용하여 추가 5' 뉴클레오티드 없이 전사체를 생성할 수 있다. 일반적인 IVT 반응에서, DNA 주형은 적절한 RNA 중합효소 효소, 완충제 및 뉴클레오티드(NTP)와 함께 배양된다.

[0327] 생성된 RNA 폴리뉴클레오티드는 7-메틸구아노신 또는 관련 구조와 같은 5' 캡 구조의 추가, 및 임의로 폴리아데닐레이트(폴리A) 테일을 포함하도록 3' 말단을 변형하는 것을 포함하지만 이에 제한되지 않는 추가로 변형될 수 있다. 변형된 IVT 반응에서, RNA는 IVT 동안 캡 유사체의 첨가를 통해 공동-전사적으로 5' 캡 구조로 캡핑된다. 캡 유사체는 디뉴클레오티드(m^7G-ppp-N) 캡 유사체 또는 트리뉴클레오티드(m^7G-ppp-N-N) 캡 유사체를 포함할 수 있으며, 여기서 N은 뉴클레오티드 또는 변형된 뉴클레오티드(예를 들어, 아데노신, 구아노신, 시티딘 및 우라딘을 포함하지만 이에 제한되지 않는 리보뉴클레오시드)를 나타낸다. 예시적인 캡 유사체 및 IVT 반응에서의 이들의 사용은 또한 모든 목적을 위해 본원에 참고로 포함되는 미국 특허 제10,519,189호에 기재되어 있다. 논의된 바와 같이, T7 중합효소는 구아노신으로 시작하는 RNA 전사체를 보다 효율적으로 전사한다. 구아노신으로 시작하지 않는 주형에서 전사 효율을 향상시키기 위해, 트리뉴클레오티드 캡 유사체(m^7G-ppp-N-N)를 사용할 수 있다. 트리뉴클레오티드 캡 유사체는 디뉴클레오티드 캡 유사체(m^7G-ppp-N)를 사용하여 IVT 반응과 비교하여 전사 효율을 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20배 이상 증가시킬 수 있다.

[0328] 5' 캡 구조는 또한 mRNA 2'-O-메틸트랜스퍼라제 및 S-아데노실 메티오닌을 함유하는 백시니아 캡핑 시스템(예를 들어, NEB 카탈로그 번호 M2080)을 사용하는 것과 같이 전사 후에 추가될 수 있다.

[0329] 생성된 RNA 폴리뉴클레오티드는 폴리아데닐레이트(폴리A) 테일을 포함하도록 3' 말단을 변형하는 것을 포함하지만 이에 제한되지 않는 기술된 캡핑 기술과 별도로 또는 추가로 선택적으로 추가로 변형될 수 있다.

[0330] 그 다음, RNA는 페놀-클로로포름 추출과 같은 분야에 잘 알려진 기술을 사용하여 정제될 수 있다.

[0331] **IV.D.4. 지질 나노입자를 통한 전달**

[0332] 백신 벡터 디자인에서 고려해야 할 중요한 측면은 벡터 자체에 대한 면역성이다(Riley 2017). 이것은 특정 인간 아데노바이러스 시스템과 같이 벡터 자체에 대한 기존 면역의 형태이거나 백신 투여 후 벡터에 대한 면역이 발달하는 형태일 수 있다. 후자는 분리된 프라이밍 및 부스팅 용량과 같이 동일한 백신의 다중 투여가 수행되거나 동일한 백신 벡터 시스템을 사용하여 다른 항원 카세트를 전달하는 경우 중요한 고려 사항이다.

[0333] 알파바이러스 벡터의 경우, 표준 전달 방법은 캡시드, E1 및 E2 단백질을 트랜스(*in trans*)로 제공하여 감염성 바이러스 입자를 생성하는 이전에 논의된 헬퍼 바이러스 시스템이다. 그러나 E1 및 E2 단백질이 종종 항체를 중화시키는 주요 표적이라는 점에 유의하는 것이 중요하다(Strauss 1994). 따라서, 감염성 입자가 중화 항체에 의해 표적화되는 경우, 표적 세포에 관심 항원을 전달하기 위해 알파바이러스 벡터를 사용하는 효능이 감소될 수 있다.

[0334] 바이러스 입자 매개 유전자 전달에 대한 대안은 발현 벡터를 전달하기 위해 나노물질을 사용하는 것이다(Riley 2017). 나노물질 비히클은 중요하게는 비-면역원성 물질로 만들어질 수 있으며 일반적으로 전달 벡터 자체에 대한 면역을 유발하는 것을 피할 수 있다. 이러한 물질은 지질, 무기 나노물질 및 기타 중합체 물질을 포함할 수 있지만 이에 제한되지는 않는다. 지질은 양이온성, 음이온성 또는 중성일 수 있다. 물질은 합성 또는 천연 유래일 수 있으며 어떤 경우에는 생분해성일 수 있다. 지질은 지방, 콜레스테롤, 인지질, 폴리에틸렌글리콜(PEG) 접합체(PEG화 지질), 왁스, 오일, 글리세라이드 및 지방 가용성 비타민을 포함하나 이에 제한되지 않는 지질 접합체를 포함할 수 있다.

[0335] 지질 나노입자(LNP)는 막 및 소포 유사 구조의 형성을 가능하게 하는 지질의 양친매성 특성으로 인해 매력적인 전달 시스템이다(Riley 2017). 일반적으로 이러한 소포는 표적 세포의 막으로 흡수되고 핵산을 세포질 내로 방출함으로써 발현 벡터를 전달한다. 또한, 특정 세포 유형의 표적화를 용이하게 하기 위해 LNP를 추가로 변형하거나 기능화할 수 있다. LNP 설계에서 또 다른 고려 사항은 표적화 효율성과 세포 독성 간의 균형이다. 지질 조성물은 일반적으로 양이온성, 중성, 음이온성 및 양친매성 지질의 정의된 혼합물을 포함한다. 일부 예에서, 특정 지질은 LNP 응집을 방지하고, 지질 산화를 방지하고, 또는 추가 모이어티의 부착을 용이하게 하는 기능적 화학기를 제공하기 위해 포함된다. 지질 조성은 전체 LNP 크기와 안정성에 영향을 줄 수 있다. 한 예에서, 지질 조성물은 디리놀레일메틸-4-디메틸아미노부티레이트(MC3) 또는 MC3-유사 분자를 포함한다. MC3 및 MC3-유사 지질 조성물은 PEG 또는 PEG-접합 지질, 스테롤, 또는 중성 지질과 같은 하나 이상의 다른 지질을 포함하도록 제

형화될 수 있다.

[0336] 혈청에 직접 노출된 발현 벡터와 같은 핵산 벡터는 혈청 뉴클레아제에 의한 핵산의 분해 또는 유리 핵산에 의한 면역계의 표적외 자극을 포함하여 몇 가지 바람직하지 않은 결과를 가질 수 있다. 따라서 알파바이러스 벡터의 캡슐화는 잠재적인 표적외 영향을 피하면서 분해를 방지하는 데 사용할 수 있다. 특정 예에서, 알파바이러스 벡터는 LNP의 수성 내부와 같은 전달 비히클 내에서 완전히 캡슐화된다. LNP 내의 알파바이러스 벡터의 캡슐화는 미세유체 액적 생성 장치에서 수행되는 미세유체 혼합 및 액적 생성과 같은 당업자에게 잘 알려진 기술에 의해 수행될 수 있다. 이러한 장치에는 표준 T-접합 장치 또는 흐름-초점 장치가 포함되지만 이에 제한되지는 않는다. 한 예에서, MC3 또는 MC3-유사 함유 조성물과 같은 원하는 지질 제형은 알파바이러스 전달 벡터 및 다른 원하는 제제와 병행하여 액적 생성 장치에 제공되어 전달 벡터 및 원하는 제제가 MC3 또는 MC3-유사 기반 LNP의 내부에 완전히 캡슐화되도록 한다. 한 예에서, 액적 생성 장치는 생성된 LNP의 크기 범위 및 크기 분포를 제어할 수 있다. 예를 들어, LNP는 직경이 1 내지 1000 나노미터, 예를 들어 1, 10, 50, 100, 500, 또는 1000 나노미터 범위의 크기를 가질 수 있다. 액적 생성 후, 발현 벡터를 캡슐화하는 전달 비히클은 투여를 위해 제조하기 위해 추가로 처리되거나 변형될 수 있다.

[0337] **IV.E. 침팬지 아데노바이러스(ChAd)**

[0338] **IV.E.1. 침팬지 아데노바이러스로 바이러스 전달**

[0339] 하나 이상의 항원을 전달하기 위한 백신 조성물(예를 들어, 항원 카세트를 통해 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드 및 기생충-유래 펩티드 포함)은 침팬지 기원의 아데노바이러스 뉴클레오티드 서열, 다양한 신규 벡터 및 침팬지 아데노바이러스 유전자를 발현하는 세포주를 제공함으로써 생성될 수 있다. 침팬지 C68 아데노바이러스(본원에서 ChAdV68로도 지칭된다)의 뉴클레오티드 서열은 항원 전달을 위한 백신 조성물에 사용될 수 있다(서열번호: 1 참조). C68 아데노바이러스 유래 벡터의 사용은 USPN 6,083,716에 더 자세히 설명되어 있으며, 이는 모든 목적을 위해 그 전체가 참고로 본원에 포함된다.

[0340] 추가 측면에서, C68과 같은 침팬지 아데노바이러스의 DNA 서열 및 이의 발현을 지시하는 조절 서열에 작동가능하게 연결된 항원 카세트를 포함하는 재조합 아데노바이러스가 본원에 제공된다. 재조합 바이러스는 포유동물, 바람직하게는 인간 세포를 감염시킬 수 있고 세포에서 항원 카세트 생성물을 발현할 수 있다. 이 벡터에서 천연 침팬지 E1 유전자 및/또는 E3 유전자 및/또는 E4 유전자가 결실될 수 있다. 항원 카세트는 이러한 유전자 결실 부위에 삽입될 수 있다. 항원 카세트는 프라이밍된 면역 반응이 요구되는 항원을 포함할 수 있다.

[0341] 또 다른 측면에서, C68과 같은 침팬지 아데노바이러스로 감염된 포유동물 세포가 본원에 제공된다.

[0342] 또 다른 측면에서, 침팬지 아데노바이러스 유전자(예를 들어, C68로부터의) 또는 이의 기능적 단편을 발현하는 신규한 포유동물 세포주가 제공된다.

[0343] 또 다른 측면에서, 본원에서는 항원 카세트를 발현하도록 조작된 C68과 같은 침팬지 아데노바이러스의 유효량을 세포 내로 도입하는 단계를 포함하는 포유동물 세포 내로 항원 카세트를 전달하는 방법을 제공한다.

[0344] 또 다른 측면은 포유동물 숙주에서 면역 반응을 유도하는 방법을 제공한다. 상기 방법은 면역 반응이 표적화되는 감염으로부터 하나 이상의 항원을 코딩하는 항원 카세트를 포함하는, C68과 같은 재조합 침팬지 아데노바이러스의 유효량을 숙주에 투여하는 단계를 포함할 수 있다.

[0345] 또 다른 측면은 감염성 질병과 같은 대상체의 질병을 치료 또는 예방하기 위해 포유동물 숙주에서 면역 반응을 유도하는 방법을 제공한다. 상기 방법은 면역 반응이 표적화되는 감염성 질병으로부터의 하나 이상의 항원을 코딩하는 항원 카세트를 포함하는, C68과 같은 재조합 침팬지 아데노바이러스의 유효량을 숙주에 투여하는 단계를 포함할 수 있다.

[0346] 서열번호: 1의 서열로부터 수득된 침팬지 아데노바이러스 유전자를 발현하는 비-유인종 포유동물 세포가 또한 개시된다. 유전자는 서열번호: 1의 아데노바이러스 E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5로 이루어진 군으로부터 선택될 수 있다.

[0347] 서열번호: 1의 서열로부터 수득된 유전자를 포함하는 침팬지 아데노바이러스 DNA 서열을 포함하는 핵산 분자가 또한 개시된다. 유전자는 서열번호: 1의 상기 침팬지 아데노바이러스 E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5로 이루어진 군으로부터 선택될 수 있다. 일부 측면에서 핵산 분자는 서열번호: 1을 포함한다. 일부 측면에서 핵산 분자는 서열번호: 1의 E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5 유전자로 이루어

진 군으로부터 선택된 적어도 하나의 유전자가 결여된 서열번호: 1을 포함한다.

[0348] 또한, 서열번호: 1로부터 수득된 침팬지 아데노바이러스 DNA 서열 및 이종 숙주 세포에서 카세트의 발현을 지시하는 하나 이상의 조절 서열에 작동가능하게 연결된 항원 카세트를 포함하는 벡터가 개시되며, 여기서 임의로 침팬지 아데노바이러스 DNA 서열은 복제 및 비리온 캡시드화에 필요한 적어도 시스(*cis*-)-요소, 항원 카세트 및 조절 서열을 플랭킹하는 시스-요소를 포함한다. 일부 측면에서, 침팬지 아데노바이러스 DNA 서열은 서열번호: 1의 E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5 유전자 서열로 이루어진 군으로부터 선택된 유전자를 포함한다. 일부측면에서, 벡터는 E1A 및/또는 E1B 유전자가 결여될 수 있다.

[0349] 또한 결실되거나 부분적으로 결실된 E4orf2 영역 및 결실되거나 부분적으로 결실된 E4orf3 영역, 및 임의로 결실되거나 부분적으로 결실된 E4orf4 영역을 포함하는 부분적으로 결실된 E4 유전자를 포함하는 아데노바이러스 벡터가 본원에 개시된다. 부분적으로 결실된 E4는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642의 E4 결실을 포함할 수 있고, 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함한다. 부분적으로 결실된 E4는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 34,942의 적어도 부분적 결실, 서열번호: 1에 제시된 서열의 뉴클레오티드 34,952 내지 35,305의 적어도 부분적 결실, 및 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642의 적어도 부분적 결실의 E4 결실을 포함하고, 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함한다. 부분적으로 결실된 E4는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516의 E4 결실을 포함할 수 있고, 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함한다. 부분적으로 결실된 E4는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,979 내지 36,642의 E4 결실을 포함할 수 있고, 여기서 벡터는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 2 내지 36,518을 포함한다. 부분적으로 결실된 E4는 E4Orf2의 적어도 부분적인 결실, 완전히 결실된 E4Orf3, 및 E4Orf4의 적어도 부분적인 결실의 E4 결실을 포함할 수 있다. 부분적으로 결실된 E4는 E4Orf2의 적어도 부분적인 결실, E4Orf3의 적어도 부분적인 결실, 및 E4Orf4의 적어도 부분적인 결실의 E4 결실을 포함할 수 있다. 부분적으로 결실된 E4는 E4Orf1의 적어도 부분적인 결실, 완전히 결실된 E4Orf2, 및 E4Orf3의 적어도 부분적인 결실의 E4 결실을 포함할 수 있다. 부분적으로 결실된 E4는 E4Orf2의 적어도 부분적인 결실 및 E4Orf3의 적어도 부분적인 결실의 E4 결실을 포함할 수 있다. 부분적으로 결실된 E4는 E4Orf1의 시작 부위와 E4Orf5의 시작 부위 사이의 E4 결실을 포함할 수 있다. 부분적으로 결실된 E4는 E4Orf1의 시작 부위에 인접한 E4 결실일 수 있다. 부분적으로 결실된 E4는 E4Orf2의 시작 사이트에 인접한 E4 결실일 수 있다. 부분적으로 결실된 E4는 E4Orf3의 시작 부위에 인접한 E4 결실일 수 있다. 부분적으로 결실된 E4는 E4Orf4의 시작 부위에 인접한 E4 결실일 수 있다. E4 결실은 적어도 50, 적어도 100, 적어도 200, 적어도 300, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100, 적어도 1200, 적어도 1300, 적어도 1400, 적어도 1500, 적어도 1600, 적어도 1700, 적어도 1800, 적어도 1900, 또는 적어도 2000개의 뉴클레오티드일 수 있다. E4 결실은 적어도 700개의 뉴클레오티드일 수 있다. E4 결실은 적어도 1500개의 뉴클레오티드일 수 있다. E4 결실은 50 이하, 100 이하, 200 이하, 300 이하, 400 이하, 500 이하, 600 이하, 700 이하, 800 이하, 900 이하, 1000 이하, 1100 이하, 1200 이하, 1300 이하, 1400 이하, 1500 이하, 1600 이하, 1700 이하, 1800 이하, 1900 이하, 또는 2000 이하의 뉴클레오티드일 수 있다. E4 결실은 750개 이하의 뉴클레오티드일 수 있다. E4 결실은 1550개 이상의 뉴클레오티드 이하일 수 있다.

[0350] 부분적으로 결실된 E4 유전자는 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 35,642가 결여된 서열번호: 1에 제시된 E4 유전자 서열일 수 있다. 부분적으로 결실된 E4 유전자는 서열번호: 1에 제시된 E4 유전자 서열이 결여되고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,916 내지 34,942, 뉴클레오티드 34,952 내지 35,305 및 서열번호: 1에 제시된 서열의 뉴클레오티드 35,302 내지 35,642가 결여된 서열번호: 1에 제시된 E4 유전자 서열일 수 있다. 부분적으로 결실된 E4 유전자는 서열번호: 1에 제시된 E4 유전자 서열일 수 있고 서열번호: 1에 제시된 서열의 적어도 뉴클레오티드 34,980 내지 36,516이 결여되어 있다. 부분적으로 결실된 E4 유전자는 서열번호: 1에 제시된 뉴클레오티드 E4 유전자 서열일 수 있고 서열번호: 1에 제시된 뉴클레오티드 서열의 적어도 뉴클레오티드 34,979 내지 35,642가 결여되어 있다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 카세트를 가질 수 있고, 여기서 카세트는 적어도 하나의 페이로드 핵산 서열을 포함하고, 카세트는 적어도 하나의 페이로드 핵산 서열에 작동가능하게 연결된 적어도 하나의 프로모터 서열을 포함한다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 서열번호: 1에 제시된 ChAdV68 서열의 하나 이상의 유전자 또는 조절 서열을 가질 수 있으며, 임의로 여기서 하나 이상의 유전자 또는 조절 서열은 서열번호: 1에 제시된 서열의 적어도 하나의 침팬지 아데노바이러스 역 말단 반복(ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 및 L5 유전자를 포함한다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 서

열번호: 1에 제시된 서열의 뉴클레오티드 2 내지 34,916을 가질 수 있고, 여기서 부분적으로 결실된 E4 유전자는 뉴클레오티드 2 내지 34,916의 3'이고, 임의로 뉴클레오티드 2 내지 34,916에는 E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403이 추가로 결여되어 있고 및/또는 E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825가 결여된다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 서열번호: 1에 제시된 서열의 뉴클레오티드 35,643 내지 36,518을 가질 수 있고, 여기서 부분적으로 결실된 E4 유전자는 뉴클레오티드 35,643 내지 36,518의 5'이다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 서열번호: 1에 제시된 서열의 뉴클레오티드 2 내지 34,916을 가질 수 있고, 여기서 부분적으로 결실된 E4 유전자는 뉴클레오티드 2 내지 34,916의 3'이고, 뉴클레오티드 2 내지 34,916에는 추가로 E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403 및 E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825가 결여되어 있다. 부분적으로 결실된 E4 유전자를 갖는 아데노바이러스 벡터는 서열번호: 1에 제시된 서열의 뉴클레오티드 2 내지 34,916을 가질 수 있고, 여기서 부분적으로 결실된 E4 유전자는 뉴클레오티드 2 내지 34,916의 3'이고, 뉴클레오티드 2 내지 34,916에는 추가로 E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403가 결여되고 및 E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 27,125 내지 31,825가 결여되고, 서열번호: 1에 제시된 서열의 뉴클레오티드 35,643 내지 36,518을 갖고, 부분적으로 결실된 E4 유전자는 뉴클레오티드 35,643 내지 36,518의 5'이다.

[0351] 부분적으로 결실된 E4 유전자는 서열번호: 1에 제시된 서열의 뉴클레오티드 34,916 내지 35,642, 서열번호: 1에 제시된 서열의 뉴클레오티드 2 내지 34,916이 결여된 서열번호: 1에 나타낸 E4 유전자 서열일 수 있고, 여기서 부분적으로 결실된 E4 유전자는 뉴클레오티드 2 내지 34,916의 3'이고, 뉴클레오티드 2 내지 34,916은 E1 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드 577 내지 3403이 추가로 결여되고 E3 결실에 상응하는 서열번호: 1에 제시된 서열의 뉴클레오티드가 27,125 내지 31,825가 결여되고 서열번호: 1에 제시된 서열의 뉴클레오티드 35,643 내지 36,518을 가지며, 부분적으로 결실된 E4 유전자는 뉴클레오티드 35,643 내지 36,518의 5'이다.

[0352] 또한, 항원 카세트를 발현하도록 조작된 C68 벡터와 같은 본원에 개시된 벡터로 형질주입된 숙주 세포가 본원에 개시되어 있다. 또한, 본원에 개시된 벡터를 세포 내로 도입함으로써 내부에 도입된 선택된 유전자를 발현하는 인간 세포가 본원에 개시된다

[0353] 또한, 항원 카세트를 발현하도록 조작된 C68 벡터와 같은 본원에 개시된 벡터의 유효량을 포유동물 세포에 도입하는 것을 포함하는, 항원 카세트를 포유동물 세포에 전달하는 방법이 본원에 개시되어 있다.

[0354] 또한, 본원에 개시된 벡터를 포유동물 세포에 도입하고, 적합한 조건 하에 세포를 배양하고, 항원을 생산하는 것을 포함하여, 항원을 생산하는 방법이 본원에 개시되어 있다.

[0355] **IV.E.2. E1-발현 상보성 세포주**

[0356] 본원에 기재된 임의의 유전자에서 결실된 재조합 침팬지 아데노바이러스(Ad)를 생성하기 위해, 바이러스의 복제 및 감염성에 필수적인 경우 결실된 유전자 영역의 기능은 헬퍼 바이러스 또는 세포주, 즉, 상보성 또는 팩키징 세포주에 의해 재조합 바이러스에 공급될 수 있다. 예를 들어, 복제-결함 침팬지 아데노바이러스 벡터를 생성하기 위해, 인간 또는 침팬지 아데노바이러스의 E1 유전자 생성물을 발현하는 세포주가 사용될 수 있고; 이러한 세포주는 HEK293 또는 이의 변이체를 포함할 수 있다. 침팬지 E1 유전자 생성물을 발현하는 세포주의 생성을 위한 프로토콜(USPN 6,083,716의 실시예 3 및 4)은 선택된 침팬지 아데노바이러스 유전자를 발현하는 세포주를 생성하기 위해 따를 수 있다.

[0357] AAV 증대 검정은 침팬지 아데노바이러스 E1-발현 세포주를 확인하는 데 사용될 수 있다. 이 검정은, 예를 들어 다른 종의 다른 특성화되지 않은 아데노바이러스의 E1 유전자를 사용하여 만든 세포주에서 E1 기능을 확인하는 데 유용하다. 그 검정은 USPN 6,083,716의 실시예 4B에 설명되어 있다.

[0358] 선택된 침팬지 아데노바이러스 유전자, 예를 들어 E1은 선택된 모 세포주에서의 발현을 위한 프로모터의 전사 제어 하에 있을 수 있다. 이러한 목적을 위해 유도성 또는 구성적 프로모터를 사용할 수 있다. 유도성 프로모터 중에는 아연에 의해 유도되는 양 메탈로티오닌 프로모터 또는 글루코코르티코이드, 특히 덱사메타손에 의해 유도되는 마우스 유방 종양 바이러스(MMTV) 프로모터가 포함된다. 본원에 참조로 포함된 국제 특허 출원 WO95/13392에서 확인된 것과 같은 다른 유도성 프로모터도 패키징 세포주의 생산에 사용될 수 있다. 침팬지 아데노바이러스 유전자의 발현을 제어하는 구성적 프로모터가 또한 사용될 수 있다.

[0359] 임의의 원하는 C68 유전자를 발현하는 신규 세포주의 생성을 위해 모 세포를 선택할 수 있다. 제한 없이, 이러한 모 세포주는 HeLa[ATCC 수탁 번호 CCL 2], A549[ATCC 수탁 번호 CCL 185], KB[CCL 17], 디트로이트[예를 들어, 디트로이트 510, CCL 72] 및 WI-38[CCL 75] 세포일 수 있다. 다른 적절한 모 세포주는 다른 출처에서 얻을 수 있다. 모 세포주는 CHO, HEK293 또는 이의 변이체, 911, HeLa, A549, LP-293, PER.C6 또는 AE1-2a를 포함할 수 있다.

[0360] E1-발현 세포주는 재조합 침팬지 아데노바이러스 E1 결실 벡터의 생성에 유용할 수 있다. 하나 이상의 다른 침팬지 아데노바이러스 유전자 생성물을 발현하는 본질적으로 동일한 절차를 사용하여 구성된 세포주는 그러한 생성물을 코딩하는 유전자에서 결실된 재조합 침팬지 아데노바이러스 벡터의 생성에 유용하다. 또한, 다른 인간 Ad E1 유전자 생성물을 발현하는 세포주는 침팬지 재조합 Ad를 생성하는 데에도 유용하다.

[0361] **IV.E.3. 벡터로서의 재조합 바이러스 입자**

[0362] 본원에 개시된 조성물은 하나 이상의 항원을 세포에 전달하는 바이러스 벡터를 포함할 수 있다. 이러한 벡터는 C68과 같은 침팬지 아데노바이러스 DNA 서열 및 카세트의 발현을 지시하는 조절 서열에 작동가능하게 연결된 항원 카세트를 포함한다. C68 벡터는 감염된 포유동물 세포에서 카세트를 발현할 수 있다. C68 벡터는 하나 이상의 바이러스 유전자에서 기능적으로 결실될 수 있다. 항원 카세트는 프로모터와 같은 하나 이상의 조절 서열의 제어 하에 하나 이상의 항원을 포함한다. 임의의 헬퍼 바이러스 및/또는 팩키징 세포주는 결실된 아데노바이러스 유전자의 필요한 생성물을 침팬지 바이러스 벡터에 공급할 수 있다.

[0363] "기능적으로 결실된"이라는 용어는 충분한 양의 유전자 영역이, 예를 들어 돌연변이 또는 변형에 의해 제거되거나 달리 변경되어 유전자 영역이 유전자 발현의 하나 이상의 기능적 생성물을 더 이상 생성할 수 없음을 의미한다. 기능적 결실을 초래할 수 있는 돌연변이 또는 변형에는 미성숙 종결 코돈의 도입 및 정규 및 비정규 개시 코돈의 제거와 같은 넌센스 돌연변이, mRNA 스플라이싱 또는 기타 전사 처리를 변경하는 돌연변이, 또는 이들의 조합이 포함되지만 이에 제한되지는 않는다. 원하는 경우 전체 유전자 영역이 제거될 수 있다.

[0364] 서열 결실, 삽입 및 기타 돌연변이를 포함하는 본원에 개시된 벡터를 형성하는 핵산 서열의 변형은 표준 분자생물학적 기술을 사용하여 생성될 수 있고 본 발명의 범위 내에 있다.

[0365] **IV.E.4. 바이러스 플라스미드 벡터의 구성**

[0366] 본 발명에 유용한 침팬지 아데노바이러스 C68 벡터는 재조합 결함 아데노바이러스, 즉 E1a 또는 E1b 유전자에서 기능적으로 결실된 침팬지 아데노바이러스 서열을 포함하고, 임의로 다른 돌연변이, 예를 들어 온도-민감성 돌연변이 또는 다른 유전자에서 결실을 보유한다. 이들 침팬지 서열은 또한 다른 아데노바이러스 및/또는 아데노-관련 바이러스 서열로부터 하이브리드 벡터를 형성하는데 유용할 것으로 예상된다. 인간 아데노바이러스로부터 제조된 상동성 아데노바이러스 벡터는 공개된 문헌에 기재되어 있다(위에서 인용되고 여기에 참고로 포함된 Kozarsky I and II, 및 U.S. Pat. No. 5,240,846).

[0367] 항원 카세트를 인간(또는 다른 포유동물) 세포로 전달하기 위한 유용한 침팬지 아데노바이러스 C68 벡터의 구성에서, 다양한 아데노바이러스 핵산 서열이 벡터에 사용될 수 있다. 최소 침팬지 C68 아데노바이러스 서열을 포함하는 벡터는 감염성 재조합 바이러스 입자를 생성하기 위해 헬퍼 바이러스와 함께 사용될 수 있다. 헬퍼 바이러스는 최소 침팬지 아데노바이러스 벡터의 바이러스 감염성 및 증식에 필요한 필수 유전자 산물을 제공한다. 다른 기능의 바이러스 벡터에서 침팬지 아데노바이러스 유전자의 하나 이상의 선택된 결실만 이루어진 경우, 결실된 유전자 생성물은 트랜스로 결실된 유전자 기능을 제공하는 선택된 패키징 세포주에서 바이러스를 증식시켜 바이러스 벡터 생산 공정에 공급될 수 있다.

[0368] **IV.E.5. 재조합 최소 아데노바이러스**

[0369] 최소 침팬지 Ad C68 바이러스는 복제 및 비리온 캡시드화에 필요한 아데노바이러스 시스-요소만을 함유하는 바이러스 입자이다. 즉, 벡터는 아데노바이러스의 시스-작용 5' 및 3' 역 말단 반복부(ITR) 서열(복제 기점으로 기능) 및 천연 5' 패키징/인핸서 도메인(E1 프로모터에 대한 Ad 게놈 및 인핸서 요소 선형 패키징에 필요한 서열을 함유)을 함유한다. 예를 들어, 본원에 참조로 포함되는 국제 특허 출원 WO96/13597에서 "최소" 인간 Ad 벡터의 제조에 대한 기술이 설명되어 있다.

[0370] **IV.E.6. 기타 결함 아데노바이러스**

[0371] 재조합 복제-결함 아데노바이러스는 또한 최소 침팬지 아데노바이러스 서열보다 많은 것을 함유할 수 있다. 이들 다른 Ad 벡터는 바이러스의 유전자 영역의 다양한 부분의 결실, 및 헬퍼 바이러스 및/또는 패키징 세포주의

임의의 사용에 의해 형성된 감염성 바이러스 입자를 특징으로 할 수 있다.

[0372] 한 예로서, 적절한 벡터는 정상적인 생물학적 기능을 제거하기 위해 C68 아데노바이러스 즉시 초기 유전자 E1a 및 지연된 초기 유전자 E1b의 전부 또는 충분한 부분을 결실시킴으로써 형성될 수 있다. 복제-결함 E1-결실된 바이러스는 해당 유전자 생성물을 트랜스로 제공하는 기능성 아데노바이러스 E1a 및 E1b 유전자를 함유하는 침팬지 아데노바이러스로 형질전환된 상보성 세포주에서 증식할 때 감염성 바이러스를 복제하고 생산할 수 있다. 알려진 아데노바이러스 서열에 대한 상동성에 기초하여, 당해 분야의 인간 재조합 E1-결실된 아데노바이러스에 대해 사실인 것처럼, 생성된 재조합 침팬지 아데노바이러스는 많은 세포 유형을 감염시킬 수 있고 항원(들)을 발현할 수 있지만, 침팬지 E1 영역 DNA를 가지고 있지 않은 대부분의 세포에서는 세포가 매우 높은 감염 다중도로 감염되지 않는 한 복제할 수 없다.

[0373] 또 다른 예로서, C68 아데노바이러스 지연 초기 유전자 E3의 전부 또는 일부는 재조합 바이러스의 일부를 형성하는 침팬지 아데노바이러스 서열로부터 제거될 수 있다.

[0374] E4 유전자가 결실된 침팬지 아데노바이러스 C68 벡터도 구성될 수 있다. 또 다른 벡터는 지연된 초기 유전자 E2a에 결실을 함유할 수 있다.

[0375] 침팬지 C68 아데노바이러스 게놈의 후기 유전자 L1 내지 L5 중 어느 것에서도 결실이 이루어질 수 있다. 유사하게, 중간 유전자 IX 및 IVa2의 결실은 일부 목적에 유용할 수 있다. 다른 구조적 또는 비구조적 아데노바이러스 유전자에서 다른 결실이 이루어질 수 있다.

[0376] 상기 논의된 결실은 개별적으로 사용될 수 있으며, 즉, 아데노바이러스 서열은 E1의 결실만을 함유할 수 있다. 대안적으로, 그들의 생물학적 활성을 파괴하거나 감소시키는데 효과적인 전체 유전자 또는 이의 부분의 결실은 임의의 조합으로 사용될 수 있다. 예를 들어, 하나의 예시적인 벡터에서, 아데노바이러스 C68 서열은 E3 결실과 함께 또는 없이 E1 유전자 및 E4 유전자 결실, 또는 E1, E2a 및 E3 유전자 결실, 또는 E1 및 E3 유전자 결실, 또는 E1, E2a 및 E4 유전자의 결실을 가질 수 있다. 상기 논의된 바와 같이, 이러한 결실은 원하는 결과를 달성하기 위해 온도-민감성 돌연변이와 같은 다른 돌연변이와 조합하여 사용될 수 있다.

[0377] 항원(들)을 포함하는 카세트는 임의로 침팬지 C68 Ad 바이러스의 임의의 결실된 영역에 삽입된다. 대안적으로, 카세트는 원하는 경우 해당 영역의 기능을 방해하기 위해 기존 유전자 영역에 삽입될 수 있다.

[0378] **IV.E.7. 헬퍼 바이러스**

[0379] 항원 카세트를 운반하기 위해 사용된 바이러스 벡터의 침팬지 아데노바이러스 유전자 함량에 따라, 카세트를 함유하는 감염성 재조합 바이러스 입자를 생성하기에 충분한 침팬지 아데노바이러스 유전자 서열을 제공하기 위해 헬퍼 아데노바이러스 또는 비-복제 바이러스 단편이 사용될 수 있다.

[0380] 유용한 헬퍼 바이러스는 아데노바이러스 벡터 구성물에 존재하지 않거나 및/또는 벡터가 형질주입된 패키징 세포주에 의해 발현되지 않는 선택된 아데노바이러스 유전자 서열을 함유한다. 헬퍼 바이러스는 복제-결함이 있을 수 있으며 위에서 설명한 서열 외에도 다양한 아데노바이러스 유전자를 함유한다. 헬퍼 바이러스는 본원에 기재된 E1-발현 세포주와 조합하여 사용될 수 있다.

[0381] C68의 경우, "헬퍼" 바이러스는 SspI를 갖는 C68 게놈의 C 말단을 클리핑하여 형성된 단편일 수 있으며, 이는 바이러스의 왼쪽 말단으로부터 약 1300 bp를 제거한다. 이 클리핑된 바이러스는 플라스미드 DNA와 함께 E1-발현 세포주로 공동-형질주입되어 플라스미드에서 C68 서열과 상동 재조합에 의해 재조합 바이러스를 형성한다.

[0382] 헬퍼 바이러스는 또한 문헌에 기재된 바와 같이 다중-양이온 접합체로 형성될 수 있다(Wu 등, J. Biol. Chem., 264:16985-16987 (1989); K. J. Fisher and J. M. Wilson, Biochem. J., 299:49 (Apr. 1, 1994)). 헬퍼 바이러스는 임의로 리포터 유전자를 함유할 수 있다. 다수의 그러한 리포터 유전자가 당업계에 공지되어 있다. 아데노바이러스 벡터 상의 항원 카세트와 다른 헬퍼 바이러스 상의 리포터 유전자의 존재는 Ad 벡터 및 헬퍼 바이러스 모두를 독립적으로 모니터링할 수 있게 한다. 이 두 번째 리포터는 정제 시 생성된 재조합 바이러스 및 헬퍼 바이러스를 분리하는 데 사용된다.

[0383] **IV.E.8. 바이러스 입자의 조립 및 세포주의 감염**

[0384] 아데노바이러스, 항원 카세트 및 기타 벡터 요소의 선택된 DNA 서열을 다양한 중간 플라스미드 및 셔틀 벡터로 조립하고 플라스미드 및 벡터를 사용하여 재조합 바이러스 입자를 생성하는 것은 모두 통상적인 기술을 사용하여 달성할 수 있다. 이러한 기술에는 시험관내 재조합 기술(예를 들어, 깁슨(Gibson) 어셈블리)에서, cDNA의 통

상적인 클로닝 기술, 아데노바이러스 게놈의 중첩 올리고뉴클레오티드 서열의 사용, 중합효소 연쇄 반응, 및 원하는 뉴클레오티드 서열을 제공하는 임의의 적합한 방법이 포함된다. 표준 형질주입 및 공동-형질주입 기술, 예를 들어 CaPO4 침전 기술 또는 리포펙타민과 같은 리포솜-매개 형질감염 방법이 사용된다. 사용되는 다른 통상적인 방법에는 바이러스 게놈의 상동 재조합, 한천 오버레이에서 바이러스 플라크, 신호 생성 측정 방법 등이 포함된다.

[0385] 예를 들어, 원하는 항원 카세트-함유 바이러스 벡터의 구성 및 조립 후, 벡터는 헬퍼 바이러스의 존재 하에 시험관내에서 패키징 세포주로 형질주입될 수 있다. 헬퍼 및 벡터 서열 사이에 상동 재조합이 일어나서 벡터의 아데노바이러스-항원 서열이 복제되고 비리온 캡시드로 패키징되어 재조합 바이러스 벡터 입자가 생성된다.

[0386] 생성된 재조합 침팬지 C68 아데노바이러스는 항원 카세트를 선택된 세포로 전달하는데 유용하다. 패키징 세포주에서 성장한 재조합 바이러스를 사용한 생체내 실험에서, E1-결실된 재조합 침팬지 아데노바이러스는 비-침팬지, 바람직하게는 인간, 세포로 카세트를 옮기는 데 유용함을 보여준다.

[0387] **IV.E.9. 재조합 바이러스 벡터의 사용**

[0388] 항원 카세트를 함유하는 생성된 재조합 침팬지 C68 아데노바이러스(상기 기재된 바와 같이 아데노바이러스 벡터 및 헬퍼 바이러스 또는 아데노바이러스 벡터 및 패키징 세포주의 협력에 의해 생성)는 따라서 항원(들)을 생체내 또는 생체외에서 대상체에게 전달할 수 있는 효율적인 유전자 전달 비히클을 제공한다.

[0389] 상기 기술된 재조합 벡터는 공개된 유전자 치료 방법에 따라 인간에게 투여된다. 항원 카세트를 보유하는 침팬지 바이러스 벡터는 바람직하게는 생물학적으로 적합한 용액 또는 약제학적으로 허용되는 전달 비히클에 현탁되어 환자에게 투여될 수 있다. 적합한 비히클에는 멸균 식염수가 포함된다. 다른 수성 및 비-수성 등장 멸균 주사 용액 및 약제학적으로 허용가능한 담체로 알려져 있고 당업자에게 잘 알려진 수성 및 비-수성 멸균 현탁액이 이러한 목적을 위해 사용될 수 있다.

[0390] 침팬지 아데노바이러스 벡터는 인간 세포를 형질도입하고 충분한 수준의 항원 전달 및 발현을 제공하여 과도한 유해 없이 또는 의학적으로 허용되는 생리학적 효과와 함께 치료적 이점을 제공하기에 충분한 양으로 투여되며, 이는 의료 부문에서 당업자에 의해 결정될 수 있다. 통상적이고 약제학상 허용되는 투여 경로는 간, 비강내, 정맥내, 근육내, 피하, 피내, 경구 및 기타 비경구 투여 경로로의 직접 전달을 포함하지만 이에 제한되지는 않는다. 원하는 경우 투여 경로를 조합할 수 있다.

[0391] 바이러스 벡터의 투여량은 주로 치료되는 상태, 환자의 연령, 체중 및 건강과 같은 요인에 의존할 것이며, 따라서 환자마다 다를 수 있다. 투여량은 임의의 부작용에 대한 치료적 이점의 균형을 맞추기 위해 조정될 것이며 이러한 투여량은 재조합 벡터가 사용되는 치료적 적용에 따라 달라질 수 있다. 항원(들)의 발현 수준은 투여량 투여의 빈도를 결정하기 위해 모니터링될 수 있다.

[0392] 재조합 복제 결함 아데노바이러스는 "약제학적 유효량", 즉 원하는 세포를 형질주입시키고 선택된 유전자의 충분한 발현 수준을 제공하기 위한 투여 경로에서 백신 접종 혜택, 즉 측정가능한 수준의 보호 면역을 제공하기 위해 효과적인 재조합 아데노바이러스의 양으로 투여될 수 있다. 항원 카세트를 포함하는 C68 벡터는 보조제와 함께 공동-투여될 수 있다. 보조제는 특히 보조제가 단백질인 경우 벡터와 분리되거나(예를 들어, 명반) 벡터 내에서 코딩될 수 있다. 보조제는 당업계에 잘 알려져 있다.

[0393] 통상적이고 약제학상 허용되는 투여 경로는 비강내, 근육내, 기관내, 피하, 피내, 직장, 경구 및 기타 비경구 투여 경로를 포함하지만 이에 제한되지는 않는다. 투여 경로는 원하는 경우 조합되거나 면역원 또는 질병에 따라 조정될 수 있다. 예를 들어, 광견병 예방에서는 피하, 기관 내 및 비강 경로가 선호된다. 투여 경로는 주로 치료되는 질병의 특성에 따라 달라진다.

[0394] 항원(들)에 대한 면역 수준을 모니터링하여 부스터에 대한 필요성이 있는 경우 이를 결정할 수 있다. 혈청 내 항체 역가 평가 후, 예를 들어 임의의 부스터 면역이 필요할 수 있다.

[0395] **V. 치료 및 제조 방법**

[0396] 또한, 본원에 개시된 방법을 사용하여 확인된 다수의 항원과 같은 하나 이상의 항원을 대상체에게 투여함으로써, 대상체에서 감염성 질병 유기체-특이적 면역 반응을 유도하고, 감염성 질병 유기체에 대해 백신을 접종하고, 대상체에서 감염성 질병 유기체와 관련된 감염의 증상을 치료 및/또는 완화하는 방법이 제공된다.

[0397] 일부 측면에서, 대상체는 감염 진단을 받았거나 감염 위험이 있거나, 예를 들어 연령, 지리적/여행, 및/또는 업

무-관련 감염 위험 증가 또는 감염의 소인, 또는 계절성 및/또는 새로운 질병 감염의 위험이 있는 것으로 진단되었다.

[0398] 항원은 CTL 반응을 유도하기에 충분한 양으로 투여될 수 있다. 항원은 T 세포 반응 반응을 유도하기에 충분한 양으로 투여될 수 있다. 항원은 B 세포 반응 반응을 유도하기에 충분한 양으로 투여될 수 있다.

[0399] 항원은 단독으로 또는 다른 치료제와 조합하여 투여될 수 있다. 치료제는 항바이러스제 또는 항생제와 같은 감염성 질병 유기체를 표적으로 하는 것들을 포함할 수 있다.

[0400] 또한, 대상체는 관문 억제제와 같은 항-면역억제제/면역자극제를 추가로 투여받을 수 있다. 예를 들어, 대상체는 항-CTLA 항체 또는 항-PD-1 또는 항-PD-L1을 추가로 투여받을 수 있다. 항체에 의한 CTLA-4 또는 PD-L1의 차단은 환자의 암 세포에 대한 면역 반응을 향상시킬 수 있다. 특히, CTLA-4 차단은 백신 접종 프로토콜을 따를 때 효과적인 것으로 나타났다.

[0401] 백신 조성물에 포함될 각 항원의 최적량 및 최적 투여 요법이 결정될 수 있다. 예를 들어, 항원 또는 이의 변이체는 정맥내(i.v.) 주사, 피하(s.c.) 주사, 피내(i.d.) 주사, 복강내(i.p.) 주사, 근육내(i.m.) 주사를 위해 제조될 수 있다. 주사 방법에는 s.c., i.d., i.p., i.m. 및 i.v.가 포함된다. DNA 또는 RNA 주입 방법에는 i.d., i.m., s.c., i.p. 및 i.v.가 포함된다. 백신 조성물의 다른 투여 방법은 당업자에게 공지되어 있다.

[0402] 백신은 조성물에 존재하는 항원의 선택, 수 및/또는 양이 조직, 감염성 질병 및/또는 환자-특이적이도록 편집될 수 있다. 예를 들어, 펩티드의 정확한 선택은 주어진 조직에서 모 단백질의 발현 패턴에 의해 안내되거나 환자의 돌연변이 또는 질병 상태에 의해 안내될 수 있다. 선택은 특정 감염성 질병(예를 들어, 대상체가 감염되었거나 감염 위험이 있는 특정 SARS-CoV-2 분리주), 질병 상태, 백신 접종 목표(예를 들어, 예방 또는 진행 중인 질병을 표적으로), 초기 치료 요법, 환자의 면역 상태, 및 물론 환자의 HLA-일배체형에 따라 달라질 수 있다. 또한 백신은 특정 환자의 개인적 필요에 따라 개별화된 구성요소를 포함할 수 있다. 예에는 특정 환자의 항원 발현에 따라 항원 선택을 변경하거나 1차 치료 또는 치료 계획에 따른 2차 치료에 대한 조정이 포함된다.

[0403] 다양한 진단 방법, 예를 들어 하기에 추가로 설명되는 환자 선택 방법을 사용하여 항원 백신의 투여를 위해 환자를 확인할 수 있다. 환자 선택은 하나 이상의 유전자의 돌연변이 또는 발현 패턴을 확인하는 것을 포함할 수 있다. 환자 선택은 진행 중인 감염의 감염성 질병을 확인하는 것을 포함할 수 있다. 환자 선택에는 전염성 질병에 의한 감염 위험을 확인하는 것이 포함될 수 있다. 어떤 경우에는 환자 선택에 환자의 일배체형 확인이 포함된다. 다양한 환자 선택 방법은 병렬로 수행될 수 있고, 예를 들어 시퀀싱 진단은 환자의 돌연변이 및 일배체형을 모두 확인할 수 있다. 다양한 환자 선택 방법은 순차적으로 수행될 수 있고, 예를 들어 하나의 진단 테스트는 돌연변이를 확인하고 별도의 진단 테스트는 환자의 일배체형을 확인하며, 여기서 각 테스트는 동일하거나(예를 들어, 고처리량 시퀀싱) 또는 상이한(예를 들어, 하나는 높은 처리량 시퀀싱 및 다른 하나는 생거 시퀀싱) 진단 방법일 수 있다.

[0404] 감염성 질병에 대한 백신으로 사용되는 조성물의 경우, 정상 조직에서 다량으로 발현되는 유사한 정상 자가-펩티드를 갖는 항원은 회피하거나 본원에 기재된 조성물에 소량으로 존재할 수 있다. 반면에, 환자의 감염된 세포가 특정 항원을 다량으로 발현하는 것으로 알려진 경우, 이 감염의 치료를 위한 각각의 약제학적 조성물은 다량으로 존재할 수 있고/있거나 이 특정 항원 또는 이 항원의 경로에 특이적인 하나 이상의 항원이 포함될 수 있다.

[0405] 항원을 포함하는 조성물은 이미 감염을 앓고 있는 개체에게 투여될 수 있다. 치료적 적용에서, 조성물은 감염성 질병 유기체 항원에 대한 효과적인 CTL 반응을 이끌어내고 증상 및/또는 합병증을 치료하거나 적어도 부분적으로 정지시키기에 충분한 양으로 환자에게 투여된다. 이를 달성하기에 적절한 양은 "치료 유효량"으로 정의된다. 이러한 용도에 효과적인 양은, 예를 들어, 조성, 투여 방식, 치료되는 질병의 단계 및 중증도, 환자의 체중 및 일반적인 건강 상태, 처방 의사의 판단에 따라 달라질 것이다. 조성물은 일반적으로 심각한 질병 상태, 즉 생명을 위협하거나 잠재적으로 생명을 위협하는 상황에서, 특히 감염성 질병 유기체가 장기 손상 및/또는 기타 면역병리를 유도한 경우에 사용될 수 있음을 염두에 두어야 한다. 그러한 경우에, 외래 물질의 최소화 및 항원의 상대적인 비독성 특성의 관점에서, 이러한 조성물의 실질적인 과량을 투여하는 것이 치료 의사에 의해 가능하고 바람직하다고 느낄 수 있다.

[0406] 치료적 사용을 위해, 투여는 감염의 검출 또는 치료에서 시작할 수 있다. 그 다음에는 적어도 증상이 실질적으로 완화될 때까지 및 그 후 일정 기간 동안 또는 면역이 제공되는 것으로 간주될 때까지(예를 들어, 기억 B 세포 또는 T 세포 집단, 또는 항원 특이적 B 세포 또는 항체가 생성되는) 부스팅 용량이 뒤따를 수 있다.

[0407] 치료적 치료를 위한 약제학적 조성물(예를 들어, 백신 조성물)은 비경구, 국소, 비강, 경구 또는 국부 투여를 위한 것이다. 약제학적 조성물은 비경구적으로, 예를 들어 정맥내로, 피하로, 피내로, 또는 근육내로 투여될 수 있다. 조성물은 대상체의 특정 감염된 조직 및/또는 세포를 표적화하기 위해 투여될 수 있다. 본원에는 항원의 용액을 포함하는 비경구 투여용 조성물 및 백신 조성물이 허용되는 담체, 예를 들어 수성 담체에 용해되거나 현탁되어 있는 조성물이 기재되어 있다. 다양한 수성 담체, 예를 들어 물, 완충수, 0.9% 식염수, 0.3% 글리신, 히알루론산 등이 사용될 수 있다. 이들 조성물은 통상적이고 잘 알려진 멸균 기술에 의해 멸균될 수 있거나 멸균 여과될 수 있다. 생성된 수용액은 그대로 사용하기 위해 포장되거나 동결건조될 수 있으며, 동결건조된 제제는 투여 전에 멸균 용액과 조합된다. 조성물은 pH 조절제 및 완충제, 긴장성 조절제, 습윤제 등과 같은 생리학적 조건을 비슷하게 하는 데 필요한 약제학상 허용되는 보조 물질, 예를 들어 나트륨 아세테이트, 나트륨 락테이트, 염화나트륨, 염화칼륨, 염화칼슘, 소르비탄 모노라우레이트, 트리에탄올아민 올레이트 등을 함유할 수 있다.

[0408] 항원은 또한 림프 조직과 같은 특정 세포 조직을 표적으로 하는 리포솜을 통해 투여될 수 있다. 리포솜은 또한 반감기를 증가시키는 데 유용하다. 리포솜은 에멀젼, 발포체, 미셀, 불용성 단일층, 액정, 인지질 분산액, 라멜라층 등을 포함한다. 이러한 제제에서 전달될 항원은 단독으로 또는 예를 들어, CD45 항원에 결합하는 모노클로날 항체와 같은 림프성 세포 사이에 널리 퍼져 있는 수용체에 결합하는 분자와 함께 또는 다른 치료 또는 면역원성 조성물과 함께 혼입될 수 있다. 따라서, 원하는 항원으로 채워진 리포솜은 림프 세포의 부위로 유도될 수 있고, 여기서 리포솜은 선택된 치료/면역원성 조성물을 전달한다. 리포솜은 일반적으로 중성 및 음으로 하전된 인지질과 콜레스테롤과 같은 스테롤을 포함하는 표준 소포 형성-지질로부터 형성될 수 있다. 지질의 선택은 일반적으로, 예를 들어 리포솜 크기, 산 불안정성 및 혈류에서 리포솜의 안정성을 고려하여 안내된다. 예를 들어, 문헌에 기술된 바와 같이 리포솜을 제조하기 위해 다양한 방법을 사용할 수 있다(Szoka 등, Ann. Rev. Biophys. Bioeng. 9; 467 (1980), U.S. Pat. Nos. 4,235,871, 4,501,728, 4,501,728, 4,837,028, 및 5,019,369).

[0409] 면역 세포를 표적화하기 위해, 리포솜에 혼입될 리간드는, 예를 들어 원하는 면역계 세포의 세포 표면 결정인자에 특이적인 항체 또는 이의 단편을 포함할 수 있다. 리포솜 현탁액은 특히 투여 방식, 전달되는 펩티드 및 치료되는 질병의 단계에 따라 달라지는 용량으로 정맥내, 국부적으로, 국소적으로 등으로 투여될 수 있다.

[0410] 치료 또는 면역화 목적을 위해, 펩티드 및 임의로 본원에 기재된 펩티드 중 하나 이상을 코딩하는 핵산이 또한 환자에게 투여될 수 있다. 핵산을 환자에게 전달하기 위해 다수의 방법이 편리하게 사용된다. 예를 들어, 핵산은 "네이키드(naked) DNA"로 직접 전달될 수 있다. 이러한 접근법은 문헌에 기술되어 있다(Wolff 등, Science 247: 1465-1468 (1990); U.S. Pat. Nos. 5,580,859 및 5,589,466). 핵산은 또한, 예를 들어 U.S. 특허 제5,204,253호에 기재된 바와 같이탄도 전달을 사용하여 투여할 수도 있다. DNA만으로 구성된 입자를 투여할 수 있다. 대안적으로, DNA는 금 입자와 같은 입자에 부착될 수 있다. 핵산 서열을 전달하기 위한 접근법은 전기천공이 있거나 없는 바이러스 벡터, mRNA 벡터 및 DNA 벡터를 포함할 수 있다.

[0411] 핵산은 또한 양이온성 지질과 같은 양이온성 화합물과 복합되어 전달될 수 있다. 지질-매개 유전자 전달 방법은 문헌에 기재되어 있다(9618372 WO AWO 96/18372; 9324640 WOAWO 93/24640; Mannino & Gould-Fogerite, BioTechniques 6(7): 682-691 (1988); U.S. Pat. No. 5,279,833 Rose U.S. Pat. No. 5,279,833; 9106309WOAWO 91/06309; 및 Felgner 등, Proc. Natl. Acad. Sci. USA 84: 7413-7414 (1987)).

[0412] 항원은 또한 백시니아, 계두, 자가-복제 알파바이러스, 마라바바이러스, 아데노바이러스(Tatsis 등, Adenoviruses, *Molecular Therapy* (2004) 10, 616-629) 또는 2세대, 3세대 또는 하이브리드 2세대/3세대 렌티바이러스 및 특정 세포 유형 또는 수용체를 표적으로 하도록 설계된 임의의 세대의 재조합 렌티바이러스(Hu 등, Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, *Immunol Rev.* (2011) 239(1): 45-61, Sakuma 등, Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper 등, Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey 등, Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880)를 포함하나 이에 제한되지 않는 렌티바이러스와 같은 바이러스 벡터 기반-백신 플랫폼에 포함될 수 있다. 위에서 언급한 바이러스 벡터-기반 백신 플랫폼의 패키징 능력에 따라, 이 접근법은 하나 이상의 항원 펩티드를 코딩하는 하나 이상의 뉴클레오티드 서열을 전달할 수 있다. 서열은 돌연변이되지 않은 서열에 의해 플랭킹될 수 있거나, 링커에 의해 분리될 수 있거나, 세포하 구획을 표적으로 하는 하나 이상의 서열이 선행될 수 있다(Gros 등, Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen 등, Targeting of cancer neoantigens with

donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu 등, Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10). 숙주에 도입되면, 감염된 세포는 항원을 발현하고, 이에 의해 펩티드(들)에 대한 숙주 면역(예를 들어, CTL) 반응을 유도한다. 면역화 프로토콜에 유용한 백시니아 벡터 및 방법은 예를 들어 U.S. 특허 제4,722,848호에 기재되어 있다. 또 다른 벡터는 BCG(Bacille Calmette Guerin)이다. BCG 벡터는 문헌에 기재되어 있다(Stover 등, Nature 351:456-460 (1991)). 항원의 치료적 투여 또는 면역화에 유용한 매우 다양한 다른 백신 벡터, 예를 들어 살모넬라 타이피 벡터 등은 본원의 설명으로부터 당업자에게 명백할 것이다.

[0413] 핵산을 투여하는 수단은 하나 또는 다중 에피토프를 코딩하는 미니유전자 구성물을 사용한다. 인간 세포에서의 발현을 위해 선택된 CTL 에피토프(미니유전자)를 코딩하는 DNA 서열을 생성하기 위해, 에피토프의 아미노산 서열은 역 번역된다. 인간 코돈 사용 표는 각 아미노산에 대한 코돈 선택을 안내하는 데 사용된다. 이러한 에피토프-코딩 DNA 서열은 직접 연결되어 연속적인 폴리펩티드 서열을 생성한다. 발현 및/또는 면역원성을 최적화하기 위해, 추가 요소를 미니유전자 설계에 혼입할 수 있다. 역 번역될 수 있고 미니유전자 서열에 포함될 수 있는 아미노산 서열의 예는 헬퍼 T 림프구, 에피토프, 리더(신호) 서열 및 소포체 보유 신호를 포함한다. 또한, CTL 에피토프의 MHC 제시는 CTL 에피토프에 인접한 합성(예를 들어, 폴리-알라닌) 또는 자연-발생 플랭킹 서열을 포함함으로써 개선될 수 있다. 미니유전자 서열은 미니유전자의 플러스 및 마이너스 가닥을 코딩하는 올리고뉴클레오티드를 조립함으로써 DNA로 전환된다. 중첩 올리고뉴클레오티드(30-100 염기 길이)는 잘 알려진 기술을 사용하여 적절한 조건에서 합성, 인산화, 정제 및 어닐링된다. 올리고뉴클레오티드의 끝은 T4 DNA 리가제를 사용하여 연결된다. CTL 에피토프 폴리펩티드를 코딩하는 이 합성 미니유전자는 원하는 발현 벡터로 클로닝될 수 있다.

[0414] 정제된 플라스미드 DNA는 다양한 제형을 사용하여 주사용으로 제조할 수 있다. 이들 중 가장 간단한 것은 멸균 인산염-완충 식염수(PBS)에서 동결건조된 DNA를 재구성하는 것이다. 다양한 방법이 설명되었으며 새로운 기술을 사용할 수 있다. 위에서 언급한 바와 같이, 핵산은 양이온성 지질과 함께 편리하게 제형화된다. 또한 당지질, 융합성 리포솜, 펩티드 및 집합적으로 보호, 상호작용, 비-응축(PINC)이라고 하는 화합물도 정제된 플라스미드 DNA와 복합체를 형성하여 안정성, 근육내 분산 또는 특정 장기 또는 세포 유형으로의 트래픽킹과 같은 변수에 영향을 미칠 수 있다.

[0415] 본원에 개시된 방법의 단계를 수행하고, 다수의 항원 또는 다수의 항원의 서브세트를 포함하는 감염성 질병 백신을 생산하는 것을 포함하여 감염성 질병 백신을 제조하는 방법이 또한 개시된다.

[0416] 본원에 개시된 항원은 당업계에 공지된 방법을 사용하여 제조할 수 있다. 예를 들어, 본원에 개시된 항원 또는 벡터(예를 들어, 하나 이상의 항원을 코딩하는 적어도 하나의 서열을 포함하는 벡터)를 생산하는 방법은 항원 또는 벡터를 발현하기에 적합한 조건 하에 숙주 세포를 배양하는 것을 포함할 수 있으며, 여기서 숙주 세포는 항원 또는 벡터를 코딩하는 적어도 하나의 폴리뉴클레오티드를 포함하고, 항원 또는 벡터를 정제하는 것을 포함한다. 표준 정제 방법에는 크로마토그래피 기술, 전기영동, 면역학, 침전, 투석, 여과, 농축 및 크로마토포커싱 기술이 포함된다.

[0417] 숙주 세포는 차이니즈 햄스터 난소(CHO) 세포, NSO 세포, 효모 또는 HEK293 세포를 포함할 수 있다. 숙주 세포는 본원에 개시된 항원 또는 벡터를 코딩하는 적어도 하나의 핵산 서열을 포함하는 하나 이상의 폴리뉴클레오티드로 형질전환될 수 있으며, 임의로 분리된 폴리뉴클레오티드는 항원 또는 벡터을 코딩하는 적어도 하나의 핵산 서열에 작동가능하게 연결된 프로모터 서열을 추가로 포함한다. 특정 구현양태에서, 분리된 폴리뉴클레오티드는 cDNA일 수 있다.

[0418] **VI. 항원 사용 및 투여**

[0419] 백신접종 프로토콜을 사용하여 대상체에게 하나 이상의 항원을 투여할 수 있다. 프라이밍 백신 및 부스팅 백신을 사용하여 대상체에게 투여할 수 있다. 프라이밍 백신은 C68(예를 들어, 서열번호: 1 또는 2에 제시된 서열) 또는 srRNA(예를 들어, 서열번호: 3 또는 4에 제시된 서열)를 기반으로 할 수 있고, 부스팅 백신은 C68(예를 들어, 서열번호: 1 또는 2에 제시된 서열) 또는 srRNA(예를 들어, 서열번호: 3 또는 4에 제시된 서열)를 기반으로 할 수 있다. 각 벡터는 일반적으로 항원을 포함하는 카세트를 포함한다. 카세트는 일반적으로 각 항원 또는 AAY와 같은 다른 비-천연 스페이서 서열을 둘러싸는 천연 서열과 같은 스페이서에 의해 분리된 약 20개의 항원을 포함할 수 있다. 카세트에는 또한 파상풍 톡소이드 항원 및 PADRE 항원과 같은 MHC II 항원이 포함될 수 있으며, 이는 보편적 클래스 II 항원으로 간주될 수 있다. 카세트는 또한 유비퀴틴 표적화 서열과 같은 표적화

서열을 포함할 수 있다. 또한, 각 백신 용량은 관문 억제제(CPI)와 함께(예를 들어, 동시에, 전에 또는 후에) 대상체에게 투여될 수 있다. CPI는 항체 또는 이의 항원 결합 부분과 같은 CTLA4, PD1 및/또는 PDL1을 억제하는 것을 포함할 수 있다. 이러한 항체에는 트레멜리무맙 또는 더발루맙이 포함될 수 있다.

[0420] 프라이밍 백신은 대상체에게 (예를 들어, 근육내로) 주사될 수 있다. 용량당 양측 주사를 사용할 수 있다. 예를 들어, ChAdV68(C68)을 1회 이상 주사(예를 들어, 총 용량 $1x10^{12}$ 바이러스 입자); 0.001 내지 1 ug RNA, 특히 0.1 또는 1 ug 범위로부터 선택된 낮은 백신 용량에서 자가-복제 RNA(srRNA)의 1회 이상의 주사; 또는 1 내지 100 ug RNA, 특히 10 또는 100 ug 범위에서 선택된 고 백신 용량에서 srRNA의 1회 이상의 주사가 사용될 수 있다.

[0421] 백신 부스트(부스팅 백신)는 프라임 백신 접종 후에 (예를 들어, 근육내로) 주사될 수 있다. 부스팅 백신은 프라임 후 약 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10주마다, 예를 들어 4주마다 및/또는 8주마다 투여될 수 있다. 용량 당 양측 주사를 사용할 수 있다. 예를 들어, ChAdV68(C68)을 1회 이상 주사(예를 들어, 총 용량 $1x10^{12}$ 바이러스 입자)가 사용될 수 있고; 0.001 내지 1 ug RNA, 특히 0.1 또는 1 ug 범위로부터 선택된 낮은 백신 용량에서 자가-복제 RNA(srRNA)의 1회 이상의 주사가 사용될 수 있고; 또는 1 내지 100 ug RNA, 특히 10 또는 100 ug 범위에서 선택된 고 백신 용량에서 srRNA의 1회 이상의 주사가 사용될 수 있다.

[0422] 항-CTLA-4(예를 들어, 트레멜리무맙)가 또한 대상체에게 투여될 수 있다. 예를 들어, 항-CTLA4는 동일한 림프절로의 배수를 보장하기 위해 근육내 백신 주사(ChAdV68 프라임 또는 srRNA 저용량) 부위 근처에 피하 투여될 수 있다. 트레멜리무맙은 CTLA-4의 선택적 인간 IgG2 mAb 억제제이다. 표적 항-CTLA-4(트레멜리무맙) 피하 용량은 전형적으로 70-75 mg(특히 75 mg)이며, 용량 범위는, 예를 들어 1-100 mg 또는 5-420 mg이다.

[0423] 특정 예에서 항-PD-L1 항체, 예컨대 더발루맙(MEDI 4736)이 사용될 수 있다. 더발루맙은 PD-1 및 CD80에 대한 PD-L1 결합을 차단하는 선택적 고친화성 인간 IgG1 mAb이다. 더발루맙은 일반적으로 4주마다 20 mg/kg i.v.로 투여된다.

[0424] 면역 모니터링은 백신 투여 전, 투여 중 및/또는 투여 후에 수행될 수 있다. 이러한 모니터링은 다른 매개변수 중에서 안전성과 효능에 대한 정보를 줄 수 있다.

[0425] 면역 모니터링을 수행하기 위해, PBMC가 일반적으로 사용된다. PBMC는 프라임 백신접종 전 및 프라임 백신접종 후(예를 들어, 4주 및 8주)로 분리할 수 있다. PBMC는 부스터 백신접종 직전과 각 부스터 백신접종 후(예를 들어, 4주 및 8주) 수확할 수 있다.

[0426] T 세포 반응 및 B 세포 반응과 같은 면역 반응은 면역 모니터링 프로토콜의 일부로 평가될 수 있다. 예를 들어, 면역 반응을 자극하는 본원에 기재된 백신 조성물의 능력이 모니터링 및/또는 평가될 수 있다. 본원에 사용된 바와 같이, "면역 반응을 자극하다"는 면역 반응의 개시(예를 들어, 나이브 대상체에서 면역 반응의 개시를 자극하는 프라이밍 백신) 또는 면역 반응의 향상(예를 들어, 프라이밍 백신에 의해 개시되는 기존 면역 반응과 같이 항원에 대한 기존 면역 반응을 갖는 대상체에서 면역 반응의 향상을 자극하는 부스팅 백신)과 같은 면역 반응에서 임의의 증가를 지칭한다.

[0427] T 세포 반응은 ELISpot, 세포내 사이토카인 염색, 사이토카인 분비 및 세포 표면 포획, T 세포 증식, MHC 다량체 염색 또는 세포독성 검정과 같은 당업계에 공지된 하나 이상의 방법을 사용하여 측정될 수 있다. 백신에 코딩된 에피토프에 대한 T 세포 반응은 ELISpot 검정을 사용하여 IFN-감마와 같은 사이토카인의 유도를 측정함으로써 PBMC에서 모니터링할 수 있다. 백신에 코딩된 에피토프에 대한 특정 CD4 또는 CD8 T 세포 반응은 유세포분석을 사용하여 IFN-감마와 같이 세포내 또는 세포외에서 포획된 사이토카인의 유도를 측정함으로써 PBMC에서 모니터링할 수 있다. 백신에 코딩된 에피토프에 대한 특정 CD4 또는 CD8 T 세포 반응은 MHC 다량체 염색을 사용하여 에피토프/MHC 클래스 I 복합체에 특이적인 T 세포 수용체를 발현하는 T 세포 집단을 측정함으로써 PBMC로부터 모니터링할 수 있다. 백신에 코딩된 에피토프에 대한 특정 CD4 또는 CD8 T 세포 반응은 3H-티미딘, 브로모데옥시우리딘 및 카르복시플루오레세인-디아세테이트-숙신이미딜에스테르(CFSE) 통합 후 T 세포 집단의 생체외 확장을 측정하여 PBMC에서 모니터링할 수 있다. 백신에 코딩된 에피토프에 특이적인 PBMC-유래 T 세포의 항원 인식 능력 및 용해 활성은 크롬 방출 검정 또는 대체 비색 세포독성 검정에 의해 기능적으로 평가될 수 있다.

[0428] B 세포 반응은 B 세포 분화(예를 들어, 형질 세포로의 분화), B 세포 또는 형질 세포 증식, B 세포 또는 형질 세포 활성화(예를 들어, CD80 또는 CD86과 같은 공동자극 마커의 상향조절), 항체 클래스 전환 및/또는 항체 생산(예를 들어, ELISA)를 결정하는 데 사용되는 검정과 같은 당업계에 알려진 하나 이상의 방법을 사용하여 측정

할 수 있다.

[0429] **VII. 항원 확인**

[0430] **VII.A. 항원 후보 확인**

[0431] NGS 분석 전사체에 대한 연구 방법이 설명되고 항원 확인 공간에 적용되었다.[6,14,15] 임상 환경에서 항원 확인을 위한 민감도와 특이성을 높이기 위한 특정 최적화를 고려할 수 있다. 이러한 최적화는 실험실 과정와 관련된 영역 및 NGS 데이터 분석과 관련된 영역의 두 가지 영역으로 그룹화할 수 있다. 최적화의 예는 당업자에게 공지되어 있으며, 예를 들어 모든 목적을 위해 전체적으로 각각이 본원에 참조로 포함되는 국제 특허 출원 공개 WO/2017/106638, WO/2018/195357, 및 WO/2018/208856에 보다 상세하게 기재된 방법이 있다.

[0432] **VII.B. HLA 펩티드의 분리 및 검출**

[0433] HLA-펩티드 분자의 분리는 조직 샘플(55-58)의 용해 및 가용화 후 고전적인 면역침전(IP) 방법을 사용하여 수행되었다. HLA 특이적 IP에 대해 정화된(clarified) 용해물을 사용했다.

[0434] 항체가 HLA 분자에 특이적인 비드에 커플링된 항체를 사용하여 면역침전을 수행하였다. 범-클래스 I HLA 면역침전의 경우, 범-클래스 I CR 항체가 사용되며 클래스 II HLA - DR의 경우, HLA-DR 항체가 사용된다. 항체는 밤새 배양하는 동안 NHS-세파로스 비드에 공유적으로 부착된다. 공유 부착 후, 비드를 세척하고 IP를 위해 분취했다. (59, 60) 면역침전은 비드에 공유 결합되지 않은 항체로 수행할 수도 있다. 일반적으로 이것은 항체를 컬럼에 고정하기 위해 단백질 A 및/또는 단백질 G로 코팅된 세파로스 또는 자기 비드를 사용하여 수행된다. MHC/펩티드 복합체를 선택적으로 강화하는 데 사용할 수 있는 일부 항체는 아래에 나열되어 있다.

[0435]

항체 명칭	특이성
W6/32	클래스 I HLA-A, B, C
L243	클래스 II – HLA-DR
Tu36	클래스 II – HLA-DR
LN3	클래스 II – HLA-DR
Tu39	클래스 II – HLA-DR, DP, DQ
SPVL3	클래스 II – HLA - DQ
B7/21	클래스 II – HLA - DP

[0436] 면역침전을 위해 정화된 조직 용해물을 항체 비드에 첨가한다. 면역 침전 후, 비드는 용해물에서 제거되고 용해물은 추가 IP를 포함한 추가 실험을 위해 저장된다. IP 비드를 세척하여 비-특이적 결합을 제거하고 HLA/펩티드 복합체는 표준 기술을 사용하여 비드에서 용출된다. 단백질 성분은 분자량 스핀 컬럼 또는 C18 분획을 사용하여 펩티드에서 제거된다. 생성된 펩티드는 스피드백(SpeedVac) 증발에 의해 건조되고 어떤 경우에는 MS 분석 전에 -20℃에서 저장된다. HLA IP는 필터 바닥이 함유된 플레이트를 사용하여 96웰 플레이트 형식으로도 수행할 수 있다. 플레이트를 사용하면 여러 IP를 동시에 수행할 수 있다.

[0437] 건조된 펩티드는 역상 크로마토그래피에 적합한 HPLC 완충액에서 재구성되고 퓨전 루모스(Fusion Lumos) 질량분석기(Thermo)에서 구배 용출을 위해 C-18 미세모세관 HPLC 컬럼에 로딩된다. 펩티드 질량/전하(m/z)의 MS1 스펙트럼을 오비트랩(Orbitrap) 검출기에서 고해상도로 수집한 다음 MS2 저해상도 스캔을 선택한 이온의 HCD 단편화 후 이온 트랩 검출기에서 수집했다. 또한 MS2 스펙트럼은 CID 또는 ETD 단편화 방법 또는 3가지 기술의 조합을 사용하여 얻을 수 있어서 펩티드의 더 큰 아미노산 범위를 수득할 수 있다. MS2 스펙트럼은 오비트랩 검출기에서 높은 분해능 질량 정확도로 측정될 수도 있다.

[0438] 각 분석으로부터의 MS2 스펙트럼은 코미트(Comet)(61, 62)를 사용하여 단백질 데이터베이스에 대해 검색되고 펩티드 확인은 퍼콜레이터(Percolator)(63-65)를 사용하여 점수화된다. 추가 시퀀싱은 PEAKS 스튜디오(Bioinformatics Solutions Inc.)를 사용하여 수행되며 스펙트럼 일치 및 새로운 시퀀싱을 포함하여(97) 다른 검색 엔진 또는 시퀀싱 방법을 사용할 수 있다.

[0439] **Ⅷ. 제시 모델**

[0440] 제시 모델은 환자에서 펩티드 제시의 가능성을 확인하는 데 사용될 수 있다. 다양한 제시 모델이, 예를 들어 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1 및 US2011/0293637, 및 국제 특허 출원 공개 WO/2018/195357, WO/2018/208856, WO/2016/187508에 상세히 기재되어 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0441] **IX. 트레이닝 모듈**

[0442] 트레이닝 모듈은 펩티드 서열이 펩티드 서열과 관련된 MHC 대립유전자에 의해 제시될 가능성을 생성하는 트레이닝 데이터 세트에 기초하여 하나 이상의 제시 모델을 구성하는데 사용될 수 있다. 다양한 트레이닝 모듈이, 예를 들어 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1, 및 국제 특허 출원 공개 WO/2018/195357, WO/2018/208856에 상세히 기재되어 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다. 트레이닝 모듈은 대립 유전자별로 펩티드의 제시 가능성을 예측하는 제시 모델을 구성할 수 있다. 트레이닝 모듈은 두 개 이상의 MHC 대립유전자가 존재하는 다중-대립유전자 설정에서 펩티드의 제시 가능성을 예측하는 제시 모델을 구성할 수도 있다.

[0443] **X. 예측 모듈**

[0444] 예측 모듈은 제시 모델을 사용하여 서열 데이터를 수신하고 서열 데이터에서 후보 항원을 선택하는 데 사용될 수 있다. 구체적으로, 서열 데이터는 환자의 감염된 세포로부터 추출된 DNA 서열, RNA 서열 및/또는 단백질 서열일 수 있다. 예측 모듈은 하나 이상의 감염성 질병 유기체 관련 항원을 함유하는 부분을 확인하기 위해 환자의 정상 조직 세포에서 추출한 서열 데이터를 환자의 감염된 세포로부터 추출된 서열 데이터와 비교하여 병원체-유래 펩티드, 바이러스-유래 펩티드, 박테리아-유래 펩티드, 진균-유래 펩티드, 기생충-유래 펩티드인 후보 항원을 확인할 수 있다. 예측 모듈은 부적절하게 발현된 후보 항원을 확인하기 위해(예를 들어, 발현된 폴리뉴클레오티드 및/또는 감염성 질병에 특이적인 폴리펩티드를 확인함으로써) 환자의 정상 조직 세포에서 추출한 서열 데이터를 환자의 감염된 조직 세포에서 추출한 서열 데이터를 비교하여 정상 세포 또는 조직과 비교하여 감염된 세포 또는 감염된 조직에서 발현이 변경된 후보 항원을 확인할 수 있다.

[0445] 제시 모듈은 펩티드 서열의 제시 가능성을 추정하기 위해 처리된 펩티드 서열에 하나 이상의 제시 모델을 적용할 수 있다. 구체적으로, 예측 모듈은 제시 모델을 후보 항원에 적용함으로써 감염된 세포 HLA 분자에 제시될 가능성이 있는 하나 이상의 후보 항원 펩티드 서열을 선택할 수 있다. 한 구현에서, 제시 모듈은 미리 결정된 임계값을 초과하는 추정된 제시 가능성을 갖는 후보 항원 서열을 선택한다. 다른 구현에서, 제시 모델은 가장 높은 추정 제시 가능성을 갖는 N 후보 항원 서열을 선택한다(여기서, N은 일반적으로 백신에서 전달될 수 있는 에피토프의 최대 수이다). 주어진 환자에 대해 선택된 후보 항원을 포함하는 백신을 환자에게 주입하여 면역 반응을 유도할 수 있다.

[0446] **XI.B.카세트 설계 모듈**

[0447] **XI.B.1 개요**

[0448] 카세트 설계 모듈을 사용하여 환자에게 주사하기 위해 선택된 후보 펩티드를 기반으로 하는 백신 카세트 서열을 생성할 수 있다. 예를 들어, 카세트 설계 모듈을 사용하여 연결된 T 세포 에피토프와 같은 연결된 에피토프 서열을 코딩하는 서열을 생성할 수 있다. 다양한 카세트 설계 모듈이 당업자에게 알려져 있으며, 예를 들어 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1 및 US2011/0293637, 및 국제 특허 출원 공개 WO/2018/195357, WO/2018/208856에 카세트 설계 모듈이 상세히 기재되어 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0449] 치료 에피토프의 세트는 미리 결정된 임계값 이상의 제시 가능성과 관련된 예측 모듈에 의해 결정된 선택된 펩티드에 기초하여 생성될 수 있으며, 여기서 제시 가능성은 제시 모델에 의해 결정된다. 그러나 다른 구현양태에서, 치료 에피토프 세트는 다수의 방법(단독 또는 조합으로) 중 임의의 하나 이상에 기초하여, 예를 들어 환자의 HLA 클래스 I 또는 클래스 II 대립유전자에 대한 결합 친화도 또는 예측된 결합 친화도, 환자의 HLA 클래스 I 또는 클래스 II 대립유전자에 대한 결합 안정성 또는 예측된 결합 안정성, 무작위 샘플링 등에 기초하여 생성될 수 있음이 이해된다.

[0450] 치료 에피토프는 선택된 펩티드 자체에 해당할 수 있다. 치료 에피토프는 또한 선택된 펩티드에 더하여 C- 및/또는 N-말단 플랭킹 서열을 포함할 수 있다. N- 및 C-말단 플랭킹 서열은 이의 소스 단백질의 맥락에서 치료 백

신 에피토프의 천연 N- 및 C-말단 플랭킹 서열일 수 있다. 치료 에피토프는 고정-길이 에피토프를 나타낼 수 있다. 치료 에피토프는 가변-길이 에피토프를 나타낼 수 있으며, 여기서 에피토프의 길이는, 예를 들어 C- 또는 N-플랭킹 서열의 길이에 따라 달라질 수 있다. 예를 들어, C-말단 플랭킹 서열 및 N-말단 플랭킹 서열은 각각 2-5개 잔기의 다양한 길이를 가질 수 있으며, 그 결과 에피토프에 대해 16개의 가능한 선택을 초래한다.

[0451] 카세트 설계 모듈은 또한 카세트에서 한 쌍의 치료 에피토프 사이의 접합부에 걸쳐 있는 접합 에피토프의 제시를 고려하여 카세트 서열을 생성할 수 있다. 접합 에피토프는 카세트에서 치료 에피토프 및 링커 서열을 연결하는 과정으로 인해 카세트에서 발생하는 새로운 비-자기이지만 관련이 없는 에피토프 서열이다. 접합 에피토프의 새로운 서열은 카세트 자체의 치료 에피토프와 다르다.

[0452] 카세트 설계 모듈은 접합 에피토프가 환자에게 제시될 가능성을 감소시키는 카세트 서열을 생성할 수 있다. 구체적으로, 카세트가 환자에게 주입될 때, 접합 에피토프는 환자의 HLA 클래스 I 또는 HLA 클래스 II 대립유전자에 의해 제시될 가능성을 가지며 각각 CD8 또는 CD4 T-세포 반응을 자극한다. 접합 에피토프에 반응하는 T 세포는 치료적 이점이 없고 항원 경쟁에 의해 카세트에서 선택된 치료 에피토프에 대한 면역 반응을 감소시킬 수 있기 때문에, 이러한 반응은 종종 바람직하지 않다.[76]

[0453] 카세트 설계 모듈은 하나 이상의 후보 카세트를 통해 반복할 수 있고, 그 카세트 서열과 관련된 접합 에피토프의 제시 점수가 수치 임계값 미만인 카세트 서열을 결정할 수 있다. 접합 에피토프 제시 점수는 카세트 내 접합 에피토프의 제시 가능성과 관련된 양이며, 접합 에피토프 제시 점수의 값이 높을수록 카세트의 접합 에피토프가 HLA 클래스 I 또는 HLA 클래스II 또는 둘 다에 의해 제시될 가능성이 높음을 나타낸다.

[0454] 한 구현양태에서, 카세트 설계 모듈은 후보 카세트 서열 중에서 가장 낮은 접합 에피토프 제시 점수와 관련된 카세트 서열을 결정할 수 있다.

[0455] 카세트 설계 모듈은 하나 이상의 후보 카세트 서열을 통해 반복하고, 후보 카세트에 대한 접합 에피토프 제시 점수를 결정하고, 임계값 미만의 접합 에피토프 제시 점수와 관련된 최적의 카세트 서열을 확인할 수 있다.

[0456] 카세트 설계 모듈은 후보 카세트 서열에서 접합 에피토프 중 임의의 것이 백신이 설계되는 주어진 환자에 대한 자가-에피토프인지 확인하기 위해 하나 이상의 후보 카세트 서열을 추가로 점검할 수 있다. 이를 달성하기 위해 카세트 설계 모듈은 BLAST와 같은 알려진 데이터베이스에 대해 접합 에피토프를 점검한다. 한 구현양태에서, 카세트 설계 모듈은 접합 자가-에피토프를 피하는 카세트를 설계하도록 구성될 수 있다.

[0457] 카세트 설계 모듈은 무차별 대입 접근법을 수행하고 모든 또는 대부분의 가능한 후보 카세트 서열을 반복하여 가장 작은 접합 에피토프 제시 점수를 갖는 서열을 선택할 수 있다. 그러나 이러한 후보 카세트의 수는 백신의 용량이 증가함에 따라 엄청나게 커질 수 있다. 예를 들어, 20개 에피토프의 백신 용량의 경우, 카세트 설계 모듈은 ~10^{18}개의 가능한 후보 카세트를 반복하여 가장 낮은 접합 에피토프 표시 점수를 가진 카세트를 결정해야 한다. 이 결정은 카세트 설계 모듈이 환자를 위한 백신을 생성하기 위해 합리적인 시간 내에 완료하기 위해 계산적으로 부담이 될 수 있고(필요한 계산 처리 자원의 관점에서) 때로는 다루기 어려울 수 있다. 더욱이, 각 후보 카세트에 대해 가능한 접합 에피토프를 설명하는 것은 훨씬 더 부담스러울 수 있다. 따라서, 카세트 설계 모듈은 무차별 대입 접근법을 위한 후보 카세트 서열의 수보다 훨씬 적은 수의 후보 카세트 서열을 반복하는 방법에 기초하여 카세트 서열을 선택할 수 있다.

[0458] 카세트 설계 모듈은 무작위로 또는 적어도 의사-무작위로 생성된 후보 카세트의 서브세트를 생성할 수 있고, 카세트 서열로서 미리 결정된 임계값 미만의 접합 에피토프 제시 점수와 관련된 후보 카세트를 선택할 수 있다. 추가적으로, 카세트 설계 모듈은 카세트 서열로서 가장 낮은 접합 에피토프 제시 점수를 갖는 서브세트로부터 후보 카세트를 선택할 수 있다. 예를 들어, 카세트 설계 모듈은 20개의 선택된 에피토프 세트에 대해 ~1백만 개의 후보 카세트의 서브세트를 생성하고 가장 작은 접합 에피토프 제시 점수를 가진 후보 카세트를 선택할 수 있다. 무작위 카세트 서열의 서브세트를 생성하고 서브세트에서 접합 에피토프 제시 점수가 낮은 카세트 서열을 선택하는 것은 무차별 대입 접근법에 비해 차선책일 수 있지만 훨씬 적은 계산 자원이 필요하므로 기술적으로 구현이 가능하다. 또한, 이와같은 보다 효율적인 기술과 대조적으로 무차별 대입 접근법을 수행하면 접합 에피토프 제시 점수가 미미하거나 무시할 수 있을 정도로 향상될 수 있으므로 자원 할당 관점에서 가치가 없다. 카세트 설계 모듈은 카세트에 대한 에피토프 서열을 비대칭 여행 판매원 문제(TSP)로 공식화하여 개선된 카세트 구성을 결정할 수 있다. 노드 목록과 각 노드 쌍 사이의 거리가 주어지면 TSP는 각 노드를 정확히 한 번 방문하고 원래 노드로 돌아가기 위해 가장 짧은 총 거리와 관련된 노드 순서를 결정한다. 예를 들어, 서로 알려진 거리의 도시 A, B, C가 있는 경우, TSP의 해결책은 가능한 경로 중에서 각 도시를 정확히 한 번 방문하기 위해 이

동한 총 거리가 가장 작은 도시의 닫힌 순서를 생성한다. 비대칭 버전의 TSP는 노드 쌍 사이의 거리가 비대칭일 때 최적의 노드 순서를 결정한다. 예를 들어, 노드 A에서 노드 B로 이동하는 "거리"는 노드 B에서 노드 A로 이동하는 "거리"와 다를 수 있다. 비대칭 TSP를 사용하여 개선된 최적 카세트를 해결함으로써 카세트 설계 모듈은 카세트의 에피토프 사이의 접합부에 걸쳐 감소된 제시 점수를 초래하는 카세트 서열을 찾을 수 있다. 비대칭 TSP의 해결책은 카세트의 접합부에 걸쳐 접합 에피토프 제시 점수를 최소화하기 위해 에피토프가 카세트에서 연결되어야 하는 순서에 해당하는 치료 에피토프의 서열을 나타낸다. 이 접근법을 통해 결정된 카세트 서열은 특히 생성된 후보 카세트 서열의 수가 많을 때 무작위 샘플링 접근법보다 훨씬 적은 계산 자원을 잠재적으로 요구하면서 접합 에피토프의 제시가 상당히 적은 서열을 생성할 수 있다. 카세트 설계를 최적화하기 위한 다양한 계산 접근법 및 비교의 예시적인 예는 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1 및 US2011/0293637, 및 국제 특허 출원 공개 WO/2018/195357, WO/2018/208856에 상세히 기재되어 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0459] 카세트 설계 모듈은 또한 백신에 코딩된 추가 단백질 서열을 고려하여 카세트 서열을 생성할 수 있다. 예를 들어, 연결된 T 세포 에피토프를 코딩하는 서열을 생성하는 데 사용되는 카세트 설계 모듈은 후보 서열 목록에서 추가 단백질 서열에 의해 이미 코딩된 T 세포 에피토프를 제거하는 것과 같이 백신에 존재하는 추가 단백질 서열(예를 들어, 전장 단백질 서열)에 의해 이미 코딩된 T 세포 에피토프를 고려할 수 있다.

[0460] 카세트 설계 모듈은 또한 서열의 크기를 고려하여 카세트 서열을 생성할 수 있다. 이론에 구속되기를 바라지 않고, 일반적으로 증가된 카세트 크기는 백신 생산 및/또는 백신 효능과 같은 백신 측면에 부정적인 영향을 미칠 수 있다. 한 예에서, 카세트 설계 모듈은 중첩 T 세포 에피토프 서열과 같은 중첩 서열을 고려할 수 있다. 일반적으로, 중첩 T 세포 에피토프 서열("프레임"이라고도 함)을 함유하는 단일 서열은 다중 펩티드를 코딩하는 데 필요한 서열 크기를 감소시키기 때문에 개별 T 세포 에피토프 서열을 개별적으로 연결하는 것보다 더 효율적이다. 따라서, 예시적인 예에서, 연결된 T 세포 에피토프를 코딩하는 서열을 생성하는 데 사용되는 카세트 설계 모듈은 추가 T 세포 에피토프를 제시하는 MHC에 대한 추가 집단 적용 범위에서 얻은 이점 대 서열 크기 증가 비용을 결정하는 것과 같은 하나 이상의 추가 T 세포 에피토프를 코딩하도록 후보 T 세포 에피토프를 확장하는 비용/이점을 고려할 수 있다.

[0461] 카세트 설계 모듈은 또한 기능적 단백질, 기능적 단백질 도메인, 기능적 단백질 서브유닛 또는 잠재적으로 현재 안전 위험을 나타내는 기능적 단백질 단편의 코딩을 제한하거나 코딩할 가능성을 제한하는 것과 같은 잠재적인 안전성을 향상시키는 다른 측면을 고려하여 카세트 서열을 생성할 수 있다. 일부 경우에, 카세트 설계 모듈은 번역된 상응하는 전장 단백질의 50% 미만, 49% 미만, 48% 미만, 47% 미만, 46% 미만, 45% 미만, 44% 미만, 43% 미만, 42% 미만, 41% 미만, 40% 미만, 39% 미만, 38% 미만, 37% 미만, 36% 미만, 35% 미만, 34% 미만 또는 33% 미만이 되도록 코딩된 펩티드의 서열 크기를 제한할 수 있다. 일부 경우에, 카세트 설계 모듈은 단일 연속 서열이 번역된 상응하는 전장 단백질의 50% 미만이 되도록 코딩된 펩티드의 서열 크기를 제한할 수 있지만, 하나 이상의 서열이 동일한 번역된 상응하는 전장 단백질에서 유래될 수 있으며 함께 50% 이상을 코딩한다. 예시적인 예에서, 중첩 T 세포 에피토프 서열("프레임")을 함유하는 단일 서열이 번역된 상응하는 전장 단백질의 50%보다 큰 경우, 프레임은 다중 프레임(예를 들어, f1, f2 등)으로 분할될 수 있어서, 각 프레임이 번역된 상응하는 전장 단백질의 50% 미만이 되도록 한다. 카세트 설계 모듈은 또한 단일 연속 서열이 번역된 상응하는 전장 단백질의 49% 미만, 48% 미만, 47% 미만, 46% 미만, 45% 미만, 44% 미만, 43% 미만, 42% 미만, 41% 미만, 40% 미만, 39% 미만, 38% 미만, 37% 미만, 36% 미만, 35% 미만, 34% 미만 또는 33% 미만이 되도록 코딩된 펩티드의 서열 크기를 제한할 수 있다. 동일한 유전자의 다중 프레임이 코딩되는 경우, 다중 프레임은 서로 중첩되는 서열을 가질 수 있고 즉, 각각이 동일한 서열을 분리하여 코딩한다. 동일한 유전자로부터의 다중 프레임이 코딩되는 경우, 제 2 핵산 서열이 상응하는 유전자의 제 1 핵산 서열을 따르거나 바로 따르지 않는 경우, 동일한 유전자로부터 유래된 둘 이상의 핵산 서열은 제 1 핵산 서열이 제 2 핵산 서열에 바로 이어지거나 연결될 수 없도록 정렬될 수 있다. 예를 들어, 동일한 유전자 내에 3개의 프레임이 있는 경우(아미노산 위치의 오름차순으로 f1,f2,f3):

[0462] - 다음 카세트 순서는 허용되지 않는다:

[0463] o f1 바로 다음에 f2

[0464] o f2 바로 다음에 f3

[0465] o f1 바로 다음에 f3

[0466] - 다음 카세트 순서는 허용된다.

[0467] o f3 바로 다음에 f2

[0468] f2 바로 다음에 f1

[0469] **XII. 예시 컴퓨터**

[0470] 컴퓨터는 본원에 설명된 모든 계산 방법에 사용될 수 있다. 당업자는 컴퓨터가 상이한 구조를 가질 수 있음을 인식할 것이다. 컴퓨터의 예는 당업자에게 공지되어 있으며, 예를 들어 US 특허 번호 10,055,540, US 출원 공개 번호 US2020/0010849A1, 및 국제 특허 출원 공개 WO/2018/195357, WO/2018/208856에 상세히 기재되어 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0471] **XIII. 항원 전달 벡터 예**

[0472] 다음은 본 발명을 수행하기 위한 특정 구현양태의 실시예이다. 실시예는 단지 예시의 목적으로 제공되며, 어떤 식으로든 본 발명의 범위를 제한하도록 의도되지 않는다. 사용된 숫자(예를 들어, 양, 온도 등)와 관련하여 정확성을 보장하기 위해 노력했지만 약간의 실험적 오류와 편차는 물론 허용되어야 한다.

[0473] 본 발명의 실행은 달리 명시되지 않는 한, 당업계의 기술 내에서 단백질 화학, 생화학, 재조합 DNA 기술 및 약리학의 통상적인 방법을 사용할 것이다. 이러한 기술은 문헌에 완전히 설명되어 있다(T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, 등, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992)).

[0474] **XIII.A. 항원 카세트 설계**

[0475] 백신접종을 통해 상응하는 세포 면역 반응(들)을 자극하는 다중 클래스 I MHC 제한 항원이 전달될 수 있다. 한 예에서, 백신 카세트는 에피토프가 그들의 천연 주변 펩티드 서열 내에 임베딩되거나 비-천연 링커 서열에 의해 이격되어 있는 단일 유전자 생성물로서 다중 에피토프를 코딩하도록 조작되었다. 항원 처리 및 제시 및 따라서 항원 특이적 CD8 T 세포 반응의 크기와 폭에 잠재적으로 영향을 미칠 수 있는 여러 설계 매개변수가 확인되었다. 본 실시예에서, 다음을 평가하기 위해 여러 모델 카세트를 설계 및 구성하였다: (1) 단일 발현 카세트에 혼입된 다중 에피토프에 대해 강력한 T 세포 반응이 생성될 수 있는지 여부; (2) 발현 카세트 내의 항원 사이에 위치하는 최적의 링커를 만드는 것- 모든 에피토프의 최적 처리 및 제시를 유도하는 것; (3) 카세트 내 에피토프의 상대적 위치가 온전한 T 세포 반응에 영향을 미치는지 여부; (4) 카세트 내의 에피토프 수가 개별 에피토프에 대한 T 세포 반응의 크기 또는 품질에 영향을 미치는지 여부; (5) 세포 표적화 서열의 첨가가 T 세포 반응을 개선하는지 여부.

[0476] 모델 카세트 내의 마커 에피토프에 대해 특이적인 항원 제시 및 T 세포 반응을 평가하기 위해 2개의 판독값이 개발되었다: (1) 특별히 조작된 리포터 T 세포의 활성화에 의해 측정된 항원 제시의 평가를 허용하는 시험관내 세포-기반 스크린 (Aarnoudse 등, 2002; Nagai 등, 2012); 및 (2) HLA-A2 형질전환 마우스(Vitiello 등, 1991)를 사용하여 상응하는 에피토프-특이적 T 세포 반응에 의해 인간 기원의 카세트-유래 에피토프의 백신접종 후 면역원성을 평가하기 위해사용된 생체내 검정(Cornet 등, 2006; Depla 등, 2008; Ishioka 등, 1999).

[0477] **XIV. ChAd 항원 카세트 전달 벡터**

[0478] **XIV.A. ChAd 항원 카세트 전달 벡터 구성**

[0479] 한 예에서, 침팬지 아데노바이러스(ChAd)는 항원 카세트에 대한 전달 벡터가 되도록 조작되었다. 추가 예에서, E1(nt 457 내지 3014) 및 E3(nt 27,816-31,332) 서열이 결실된 AC_000011.1(특허 US 6083716의 서열 2)에 기초하여 전장 ChAdV68 벡터를 합성하였다. CMV 프로모터/인핸서의 제어 하에 있는 리포터 유전자가 결실된 E1 서열 대신에 삽입되었다. 이 클론의 HEK293 세포로의 형질주입은 감염성 바이러스를 생성하지 않았다. 야생형 C68 바이러스의 서열을 확인하기 위해 ATCC로부터 VR-594를 분리하여 계대배양한 후 독립적으로 서열분석하였다(서열번호: 10). AC_000011.1 서열을 야생형 ChAdV68 바이러스의 ATCC VR-594 서열(서열번호: 10)과 비교할 때, 6개의 뉴클레오티드 차이가 확인되었다. 한 예에서, 5개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드를

갖는 변형된 ChAdV68 벡터가 AC_000011.1을 기반으로 생성되었다(ChAdV68.5WTnt 서열번호: 1).

[0480] 또 다른 예에서, 변형된 ChAdV68 벡터는 E1(nt 577 내지 3403) 및 E3(nt 27,816-31,332) 서열이 결실되고 상응하는 ATCC VR-594 뉴클레오티드가 4개 위치에서 치환된 AC_000011.1에 기초하여 생성되었다. CMV 프로모터/인핸서의 제어 하에 GFP 리포터(ChAdV68.4WTnt.GFP; 서열번호: 11) 또는 모델 신생항원 카세트(ChAdV68.4WTnt.MAG25mer; 서열번호: 12)가 결실된 E1 서열 대신에 삽입되었다.

[0481] 다른 예에서, 변형된 ChAdV68 벡터는 E1(nt 577 내지 3403) 및 E3(nt 27,125-31,825) 서열이 결실되고 상응하는 ATCC VR-594 뉴클레오티드가 5개 위치에서 치환된 AC_000011.1에 기초하여 생성되었다. CMV 프로모터/인핸서의 제어 하에 GFP 리포터(ChAdV68.5WTnt.GFP; 서열번호: 13) 또는 모델 신생항원 카세트(ChAdV68.5WTnt.MAG25mer; 서열번호: 2)가 결실된 E1 서열 대신에 삽입되었다.

[0482] 관련 벡터는 아래에 설명되어 있다:

[0483] - 전장 ChAdVC68 서열 "ChAdV68.5WTnt"(서열번호: 1); 5개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드가 있는 AC_000011.1 서열.

[0484] - ATCC VR-594 C68(서열번호: 10); 독립적으로 시퀀싱됨; 전장 C68

[0485] - ChAdV68.4WTnt.GFP(서열번호: 11); E1(nt 577-3403) 및 E3(nt 27,816-31,332) 서열이 결실된 AC_000011.1; 4개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드; 결실된 E1 대신 삽입된 CMV 프로모터/인핸서의 제어 하에 GFP 리포터

[0486] - ChAdV68.4WTnt.MAG25mer(서열번호: 12); E1(nt 577-3403) 및 E3(nt 27,816-31,332) 서열이 결실된 AC_000011.1; 4개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드; 결실된 E1 대신 삽입된 CMV 프로모터/인핸서의 제어 하에 있는 모델 신생항원 카세트

[0487] - ChAdV68.5WTnt.GFP(서열번호: 13); E1(nt 577-3403) 및 E3(nt 27,125-31,825) 서열이 결실된 AC_000011.1; 5개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드; 결실된 E1 대신 삽입된 CMV 프로모터/인핸서의 제어 하에 GFP 리포터

[0488] **XV. 알파바이러스 항원 카세트 전달 벡터**

[0489] **XV.B. 알파바이러스 벡터**

[0490] 본 발명의 한 구현에서, 항원 발현 시스템을 위한 RNA 알파바이러스 백본은 베네수엘라 말 뇌염(VEE)(Kinney, 1986, Virology 152:400-413) 기반 자가-복제 RNA(srRNA) 벡터로부터 생성되었다. 한 예에서, 26S 서브-게놈 프로모터의 3'에 위치한 VEE의 구조적 단백질을 코딩하는 서열이 결실되었고(VEE 서열 7544-11,175가 결실; Kinney 등 1986에 기초한 넘버링; 서열번호: 6) 항원 서열(서열번호: 14 및 서열번호: 4) 또는 루시퍼라제 리포터(예를 들어, VEE-루시퍼라제, 서열번호: 15)로 대체되었다.

[0491] **XVI. 비-인간 영장류 연구**

[0492] ChAdV68 및 자가-복제 RNA(srRNA)를 사용하는 다양한 투여 프로토콜을 비-인간 영장류(NHP)에서 평가하였다.

[0493] **재료 및 방법**

[0494] 연구(백신 프라임)를 개시하기 위해 프라이밍 백신을 각 NHP에 근육내(IM) 주사하였다. 하나 이상의 부스팅 백신(백신 부스트)도 각 NHP에 근육내 주사되었다. 용량 당 양측 주사는 표에 요약되어 있고 아래에 요약된 군에 따라 투여되었다.

[0495] **예방접종**

[0496] Mamu-A*01 인도 붉은털 원숭이를 LNP-1 또는 LNP-2로 제형화된 ChAdV68.5WTnt.MAG25mer의 $1x10^{12}$ 바이러스 입자(주사당 $5x10^{11}$ 바이러스 입자), 30 ug의 VEE-MAG25MER srRNA, 100 ug의 VEE-MAG25MER srRNA 또는 300 ug VEE-MAG25mer srRNA로 면역화시켰다. 30 ug, 100 ug 또는 300 ug VEE-MAG25mer srRNA의 백신 부스트를 프라임 백신 접종 후 표시된 시간에 근육내 투여했다.

[0497] **면역 모니터링**

[0498] 림프구 분리 배지(LSM, MP Biomedicals) 및 류코셉(LeucoSep) 분리 튜브(Greiner Bio-One)를 사용하여 프라임

백신접종 후 지정된 시간에 PBMC를 분리하고 10% FBS 및 페니실린/스트렙토마이신을 함유하는 RPMI에 재현탁했다. 세포는 죽은 세포와 세포자연사 세포를 배제하기 위해 프로피듐 요오다이드 염색을 사용하여 어튠(Attune) NxT 유세포 분석기(Thermo Fisher)에서 계수되었다. 그런 다음 세포를 후속 분석을 위해 살아있는 세포의 적절한 농도로 조정했다. 연구의 각 원숭이에 대해, T 세포 반응은 ELISpot 또는 유세포 분석법을 사용하여 측정되었다. 백신에서 코딩된 6개의 다른 붉은털원숭이 원숭이 Mamu-A*01 클래스 I 에피토프에 대한 T 세포 반응은 생체외 효소-결합 면역스팟(ELISpot) 분석을 사용하여 IFN-감마와 같은 사이토카인의 유도를 측정함으로써 PBMC에서 모니터링되었다. 원숭이 IFNg ELISpotPLUS 키트(MABTECH)를 사용하여 ELISpot 조화 지침(DOI: 10.1038/nprot.2015.068)에 따라 ELISpot 분석을 수행했다. 200,000개의 PBMC를 96-웰 IFNg 항체 코팅된 플레이트에서 16시간 동안 10 uM의 표시된 펩티드와 함께 인큐베이션하였다. 알칼리 포스파타제를 사용하여 반점을 발전시켰다. 반응 시간을 10분으로 측정하고 수돗물에서 플레이트를 작동시켜 종결시켰다. AID vSpot 판독 스펙트럼을 사용하여 반점을 계수했다. ELISpot 분석의 경우, 포화 >50%인 웰은 "너무 많아 셀 수 없음"으로 기록되었다. 복제 웰의 편차가 > 10%인 샘플은 분석에서 제외되었다. 다음 공식을 사용하여 웰 합류에 대해 반점 계수를 수정했다: 반점 수 + 2 x (반점 수 x % 합류 /[100% - % 합류]). 음성 배경은 항원 자극 웰에서 음성 펩티드 자극 웰의 반점 계수를 빼서 수정했다. 마지막으로, 셀 수 없을 정도로 많다고 표시된 웰은 가장 가까운 100단위로 반올림하여 관찰된 가장 높은 수정 값으로 설정되었다.

[0499] 백신에서 코딩된 6개의 상이한 붉은털 원숭이 Mamu-A*01 클래스 I 에피토프에 대한 특이적 CD4 및 CD8 T 세포 반응은 유세포 분석을 사용하여 IFN-감마와 같은 세포내 사이토카인의 유도를 측정함으로써 PBMC에서 모니터링되었다. 두 방법의 결과는 사이토카인이 항원-특이적 방식으로 에피토프에 유도되었음을 나타낸다.

[0500] **붉은털 원숭이의 면역원성**

[0501] 이 연구는 (a) ChAdV68.5WTnt.MAG25mer와 조합된 동종 프라임/부스트 또는 이종 프라임/부스트로서 VEE-MAG25mer srRNA 30 μg 및 100 μg 용량의 면역원성 및 예비 안전성을 평가하고; (b) LNP1 대 LNP2를 사용하여 지질 나노입자에서 VEE-MAG25mer srRNA의 면역 반응을 비교하고; (c) VEE-MAG25mer srRNA 및 ChAdV68.5WTnt.MAG25mer 면역화에 대한 T-세포 반응의 동역학을 평가하기 위해 설계되었다.

[0502] 연구 부문은 면역원성을 입증하기 위해 Mamu-A*01 인도 붉은털 원숭이에서 수행되었다. 이 연구에 사용된 선택 항원은 붉은털 원숭이, 특히 Mamu-A*01 MHC 클래스 I 일배체형을 가진 원숭이에서만 인식된다. Mamu-A*01 인도 붉은털 원숭이를 다른 연구 부문으로 무작위 배정하고(군 당 6마리의 원숭이) 다중 Mamu-A*01 제한된 에피토프를 포함하는 모델 항원을 코딩하는 ChAdV68.5WTnt.MAG25mer 또는 VEE-MAG25mer srRNA 벡터로 IM 주사를 양측에 투여했다. 연구 암은 하기에 기술된 바와 같다.

[0503] **표 A**: 인도 붉은털 원숭이에서의 비-GLP 면역원성 연구

군	프라임	부스트 1	부스트 2
1	VEE-MAG25mer srRNA -LNP1(30 μg)	VEE-MAG25mer srRNA -LNP1 (30 μg)	VEE-MAG25mer srRNA -LNP1 (30 μg)
2	VEE-MAG25mer srRNA -LNP1 (100 μg)	VEE-MAG25mer srRNA -LNP1 (100 μg)	VEE-MAG25mer srRNA -LNP1 (100 μg)
3	VEE-MAG25mer srRNA -LNP2 (100 μg)	VEE-MAG25mer srRNA -LNP2 (100 μg)	VEE-MAG25mer srRNA -LNP2 (100 μg)
4	ChAdV68.5WTnt. MAG25mer	VEE-MAG25mer srRNA -LNP1 (100 μg)	VEE-MAG25mer srRNA -LNP1 (100 μg)

[0504]

[0505] 면역 모니터링을 위해 예방접종 전 및 초기 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9, 및 10주에 PBMC를

수집하였다.

[0506] **결과**

[0507] 말초 혈액 단핵 세포(PBMC)에서 항원-특이적 세포 면역 반응을 6개의 상이한 Mamu-A*01 제한된 에피토프에 대해 예방접종 전에 및 초기 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9 및 10주에 측정하였다. 동물은 4주 및 8주에 30㎍ 또는 100㎍ 용량으로 VEE-MAG25mer srRNA로 부스터 예방접종을 받았고 표 A에 설명된 대로 LNP1 또는 LNP2로 제형화되었다. 모든 6개의 에피토프에 대한 조합된 면역 반응을 각 면역 모니터링 시점에 대해 플롯팅했다(도 1a-d 및 표 B-E).

[0508] 각각 초기 VEE-MAG25mer srRNA-LNP1(30 ㎍) 프라임 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9 또는 10주에 10^6개의 PBMC(조합된 6개의 에피토프) 당 170, 14, 15, 11, 7, 8, 14, 17, 12 SFC로 모든 측정에서 조합된 항원-특이적 면역 반응이 관찰되었다(도 1a). 조합된 항원-특이적 면역 반응은 각각 초기 VEE-MAG25mer srRNA-LNP1(100 ㎍) 프라임 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9 또는 10주에 10^6개 PBMC(조합된 6개의 에피토프) 당 108, -3, 14, 1, 37, 4, 105, 17, 25 SFC로 모든 측정에서 관찰되었다(도 1b). 조합된 항원-특이적 면역 반응은 각각 초기 VEE-MAG25mer srRNA-LNP2(100 ㎍) 프라임 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9 또는 10주에 10^6PBMC(조합된 6개의 에피토프) 당 -17, 38, 14, -2, 87, 21, 104, 129, 89 SFC로 모든 측정에서 관찰되었다(도 1c). 음수 값은 각 에피토프/동물에 대한 사전-출혈 값으로 정규화한 결과이다.

[0509] 조합된 항원 특이적 면역 반응은 각각 초기 ChAdV68.5WTnt.MAG25mer 프라임 예방접종 후 1, 2, 3, 4, 5, 6, 8, 9 또는 10주에 10^6 PBMC(조합된 6개의 에피토프) 당 1218, 1784, 1866, 973, 1813, 747, 797, 1249, 및 547 SFC로 모든 측정에서 관찰되었다(도 1d). 면역 반응은 프라임 예방접종 후 ~2-3주 후에 측정된 피크 면역 반응과 4주 후에 면역 반응의 수축으로 예상되는 프로파일을 보여주었다. 10^6개 PBMC(조합된 6개의 에피토프) 당 1813개 SFC의 조합된 항원-특이적 세포 면역 반응은 ChAdV68.5WTnt.MAG25mer로 초기 예방접종 5주 후(즉, VEE-MAG25mer srRNA를 사용한 첫 번째 부스트 후 1주) 측정되었다. VEE-MAG25mer srRNA를 사용한 첫 번째 부스트(5주) 후 1주에 측정된 면역 반응은 ChAdV68.5WTnt.MAG25mer 프라임 예방접종(3주)에 대해 측정된 피크 면역 반응과 유사했다(도 20d). 10^6개 PBMC(조합된 6개의 에피토프) 당 1249개 SFC의 조합된 항원-특이적 세포 면역 반응은 각각 ChAdV68.5WTnt.MAG25mer로 초기 예방접종 후 9주 후에 측정되었다(즉, VEE-MAG25mer srRNA로 두 번째 부스트 후 1주). VEE-MAG25mer srRNA를 사용한 2차 부스트(9주) 후 1주에 측정된 면역 반응은 부스트 예방접종 직전에 측정된 것보다 ~2배 더 높았다(도 1d).

[0510] **표 B**: VEE-MAG25mer srRNA-LNP1(30 ㎍)(군 1)에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	39.7±22.7	35.4±25.1	3.2±3.6	33±28.1	30.9±20.3	28.3±17.5
3	2±2.4	0.2±1.8	1.8±2.4	3.7±1.9	1.7±2.8	4.9±2.3
4	1±1.8	0.3±1.2	5.5±3.6	2.3±2.2	5.7±2.7	0.8±0.8
5	0.5±0.9	1.4±3.8	3.1±1.6	2.3±2.7	1.9±2	1.4±1.2
6	1.9±1.8	-0.3±3	1.7±1.2	1.4±1.4	0.8±1.1	1.1±1
8	-0.4±0.8	-0.9±2.9	0.5±1.3	3±1.1	2.2±2.1	3.7±2
9	1±1.7	1.2±4.2	7.2±3.9	0.5±0.7	1.6±3	3±1
10	3.8±1.8	11±5	-1.1±1.1	1.9±0.9	1.3±1.6	0.2±0.5

[0511]

[0512] **표 C**: VEE-MAG25mer srRNA-LNP1(100 ㎍)(군 2)에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성

세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	7.9±17.2	23.2±17.4	11.4±4.9	41.7±16.5	15±13.5	8.9±6.2
3	-3.1±4.6	-7.2±6.5	2.3±2.3	-0.3±2.7	2.7±5.1	2.2±1.4
4	1.9±3.8	-6.2±7.6	10.5±4.1	1.2±2.9	5.6±4.9	1.1±0.8
5	-2.6±7	-8±5.9	1.5±1.7	6.4±2.3	0.7±4.3	3.3±1.3
6	6.3±6.3	4.4±8.3	6.6±4.4	5.2±5.2	3.9±5	10.8±6.9
8	-3.6±7.2	-6.8±7.3	-0.8±1.2	3.4±4.2	6.4±7.5	5.7±2.7
9	8.1±2.4	20.6±23.4	18.9±5.7	8.1±8.9	9±11.2	40±17.6
10	3.1±8	-3.9±8.5	3.3±1.8	0.6±2.9	7.4±6.4	6.1±2.5

[0513]

[0514] **표 D**: VEE-MAG25mer srRNA-LNP2(100 μg)(군 3)에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	-5.9±3.8	-0.3±0.5	-0.5±1.5	-5.7±6.1	-1±1.3	-3.2±5.5
3	0.7±5.2	3.4±2.4	4.2±4.6	18.3±15.5	11.9±5.1	-0.4±8.2
4	-3.8±5.5	2.3±1.8	11.3±6.1	-3.1±5.6	8.5±4	-1.5±6.1
5	-3.7±5.7	-0.1±0.7	-0.2±1.6	3.4±8.5	3±3.1	-4.6±5
6	12.3±15	7.8±4.9	24.7±19.8	23.2±22.5	18.7±15.8	0.5±6.2
8	5.9±12.3	-0.1±0.7	-0.5±1.3	8.8±14.4	8.7±8	-1.3±4
9	16.1±13.4	16.5±4	22.9±4.2	13±13.2	16.4±7.8	19.6±9.2
10	29.9±21.8	22±19.5	0.5±2.6	22.2±22.6	35.3±15.8	19.4±17.3

[0515]

[0516] **표 E**: ChAdV68.5WTnt.MAG25mer 프라임에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	178±68.7	206.5±94.8	221.2±120	15.4±16.7	33.3±25.9	563.5±174.4
2	311.2±165.5	278.8±100.9	344.6±110.8	46.3±13.5	181.6±76.8	621.4±220.9
3	277.3±101.1	359.6±90.5	468.2±106.6	41.7±11.1	169.8±57.8	549.4±115.7
4	140±46.5	169.6±46.8	239.4±37	26.5±11.4	75±31.6	322.2±50.7
5	155.6±62.1	406.7±96.4	542.7±143.3	35.1±16.6	134.2±53.7	538.5±91.9
6	78.9±42.5	95.5±29.4	220.9±75.3	-1.4±5.3	43.4±19.6	308.1±42.6
8	88.4±30.4	162.1±30.3	253.4±78.6	21.4±11.2	53.7±22.3	217.8±45.2
9	158.5±69	322.3±87.2	338.2±137.1	5.6±12.4	109.2±17.9	314.8±43.4
10	97.3±32.5	133.2±27	154.9±59.2	10±6	26±16.7	125.5±27.7

[0517]

[0518] **인도 붉은털 원숭이에서 비-GLP RNA 용량 범위 연구(고용량)**

[0519] 이 연구는 (a) ChAdV68.5WTnt.MAG25mer와 조합된 상동 프라임/부스트 또는 이종 프라임/부스트로서 300 μg의 용량에서 VEE-MAG25mer srRNA의 면역원성을 평가하고; (b) 300 μg 용량에서 LNP1 대 LNP2를 사용한 지질 나노입자에서 VEE-MAG25mer srRNA의 면역 반응을 비교하고; 및 (c) VEE-MAG25mer srRNA 및 ChAdV68.5WTnt.MAG25mer 예방접종에 대한 T 세포 반응의 동역학을 평가하기 위해 설계되었다.

[0520] 연구 부문은 면역원성을 입증하기 위해 Mamu-A*01 인도 붉은털 원숭이에서 수행되었다. 붉은털 원숭이와 같은 비인간 영장류 종에서의 백신 면역원성은 인간에서 백신 효능의 가장 좋은 예측 인자이다. 또한, 이 연구에 사용된 선택 항원은 붉은털 원숭이, 특히 Mamu-A*01 MHC 클래스 I 일배체형을 가진 원숭이에서만 인식된다. Mamu-A*01 인도 붉은털 원숭이를 다른 연구 부문으로 무작위 배정하고(군 당 6마리의 원숭이) 다중 Mamu-A*01 제한된 항원을 포함하는 모델 항원을 코딩하는 ChAdV68.5-WTnt.MAG25mer 또는 VEE-MAG25mer srRNA로 IM 주사를 양측에 투여했다. 연구 부문은 하기에 기술된 바와 같다.

[0521] PBMC는 군 1(이종 프라임/부스트)에 대한 면역 모니터링을 위해 예방접종 전 및 초기 예방접종 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 또는 24주에 수집되었다. PBMC는 군 2 및 3(동종 프라임/부스트)에 대한 면역 모니터링을 위해 예방접종 전 및 초기 예방접종 후 4, 5, 7, 8, 10, 11, 12, 13, 14 또는 15주에 수집되었다.

[0522] **표 F**: 인도 붉은털 원숭이에서의 비-GLP 면역원성 연구

군	프라임	부스트 1	부스트 2	부스트 3
1	ChAdV68.5WTnt .MAG25mer	VEE-MAG25mer srRNA -LNP2 (300 μg)	VEE-MAG25mer srRNA -LNP2 (300 μg)	VEE-MAG25mer srRNA -LNP2 (300 μg)
2	VEE-MAG25mer srRNA -LNP2 (300 μg)	VEE-MAG25mer srRNA -LNP2 (300 μg)	VEE-MAG25mer srRNA -LNP2 (300 μg)	
3	VEE-MAG25mer srRNA -LNP1 (300 μg)	VEE-MAG25mer srRNA -LNP1 (300 μg)	VEE-MAG25mer srRNA -LNP1 (300 μg)	

[0523]

[0524] **결과**

[0525] Mamu-A*01 인도 붉은털 원숭이를 ChAdV68.5-WTnt.MAG25mer로 면역화시켰다. 말초 혈액 단핵세포(PBMC)의 항원-특이적 세포 면역 반응은 예방접종 전 및 초기 예방접종 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 또는 24주에 6개의 다른 Mamu-A*01 제한된 에피토프에 대해 측정했다(도 2 및 표 G). 동물은 4주, 12주 및 20주에 LNP2 제형을 사용하여 VEE-MAG25mer srRNA로 부스트 예방접종을 받았다. 106 PBMC(조합된 6개의 에피토프) 당 1750, 4225, 1100, 2529, 3218, 1915, 1708, 1561, 5077, 4543, 4920, 5820, 3395, 2728, 1996, 1465, 4730, 2984, 2828, 또는 3043 SFC의 조합된 항원-특이적 면역 반응은

ChAdV68.5WTnt.MAG25mer로 초기 예방접종 후 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 또는 24주에 측정되었다(도 2). VEE-MAG25mer srRNA를 사용한 두 번째 부스트 예방접종(13주차) 후 1주에 측정된 면역 반응은 부스트 예방접종 직전(12주차) 전에 측정된 것보다 ~3배 더 높았다. VEE-MAG25mer srRNA를 사용한 세 번째 부스트 예방접종(21주차) 후 1주에 측정된 면역 반응은 두 번째 부스트에서 관찰된 반응과 유사한, 부스트 예방접종 직전(20주차)보다 ~3배 더 높았다.

[0526] Mamu-A*01 인도 붉은털 원숭이를 또한 2개의 상이한 LNP 제형(LNP1 및 LNP2)을 사용하여 VEE-MAG25mer srRNA로 면역화시켰다. 말초 혈액 단핵세포(PBMC)에서 항원-특이적 세포 면역 반응은 예방접종 전 및 초기 예방접종 후 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 또는 15주에 6개의 다른 Mamu-A*01 제한된 에피토프에 대해 측정했다(도 3 및 4, 표 H 및 I). 동물은 4주 및 12주에 각각의 LNP1 또는 LNP2 제형을 사용하여 VEE-MAG25mer srRNA로 부스트 예방접종을 받았다. 10^6 PBMC(조합된 6개의 에피토프) 당 168, 204, 103, 126, 140, 145, 330, 203 및 162 SFC의 조합된 항원-특이적 면역 반응을 VEE-MAG25mer srRNA-LNP2로 예방접종한 후 4, 5, 7, 8, 10, 11, 13, 14, 15주 후에 측정하였다(도 3). 10^6개의 PBMC(조합된 6개의 에피토프) 당 189, 185, 349, 437, 492, 570, 233, 886, 369 및 381 SFC의 조합된 항원-특이적 면역 반응은 VEE-MAG25mer srRNA-LNP1로 예방접종한 후 4, 5, 7, 8, 11, 10, 12, 13, 14, 15주 후에 측정하였다(도 4).

[0527] **표 G**: ChAdV68.5WTnt.MAG25mer(군 1)로 프라이밍 백신접종에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	173±41.6	373.5±87.3	461.4±74.2	38.4±26.1	94.5±26	609.2±121.9
5	412.7±138.4	987.8±283.3	1064.4±266.9	85.6±31.2	367.2±135.2	1306.8±332.8
6	116.2±41.2	231.1±46.3	268.3±90.7	86.1±42	174.3±61	223.9±38.1
7	287.4±148.7	588.9±173.9	693.2±224.8	92.1±33.5	172.9±55.6	694.6±194.8
8	325.4±126.6	735.8±212	948.9±274.5	211.3±62.7	179.1±50	817.3±185.2
10	312±129.7	543.2±188.4	618.6±221.7	-5.7±4.1	136.5±51.3	309.9±85.6
11	248.5±81.1	348.7±129.8	581.1±205.5	-3.1±4.4	119±51.2	413.7±144.8
12	261.9±68.2	329.9±83	486.5±118.6	-1.2±5.1	132.8±31.8	350.9±69.3
13	389.3±167.7	1615.8±418.3	1244.3±403.6	1.3±8.1	522.5±155	1303.3±385.6
14	406.3±121.6	1616±491.7	1142.3±247.2	6.6±11.1	322.7±94.1	1048.6±215.6
15	446.8±138.7	1700.8±469.1	1306.3±294.4	43±24.5	421.2±87.9	1001.5±236.4
16	686.8±268.8	1979.5±541.7	1616.8±411.8	2.4±7.8	381.9±116.4	1152.8±352.7
17	375.8±109.3	1378.6±561.2	773.1±210.3	-1.4±4.3	177.6±93.7	691.7±245
18	255.9±99.7	1538.4±498.1	498.7±152.3	-5.3±3.3	26.2±13.4	413.9±164.8
19	133±62.6	955.9±456.8	491.1±121.8	-5.7±4.1	50.3±25.4	371.2±123.7
20	163.7±55.8	641.7±313.5	357.9±91.1	2.6±7.5	41.4±24.2	257.8±68.9
21	319.9±160.5	2017.1±419.9	1204.8±335.2	-3.7±5.1	268.1±109.6	924.1±301
22	244.7±105.6	1370.9±563.5	780.3±390	-3.6±5.1	118.2±68.1	473.3±249.3
23	176.7±81.8	1263.7±527.3	838.6±367.9	-5.7±4.1	73.6±49	480.9±163.9
24	236.5±92	1324.7±589.3	879.7±321	-0.4±5.7	104±53.1	498±135.8

[0528]

[0529] **표 H**: VEE-MAG25mer srRNA-LNP2 (300 μg)(군 2)로 프라이밍 백신접종에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	46±27.1	18.4±6.8	58.3±45.8	29.9±20.8	4.9±2.3	10.7±4
5	85.4±54	5.2±5.8	52.4±51.2	34.5±35	11.8±12.2	14.4±7.9
7	18.6±32.5	1.9±1.7	59.4±55.7	9.3±10.7	3.3±3	10.7±6.1
8	36.6±39.4	6.3±3.9	48.7±39.9	13.5±8.8	3.8±3.6	17.2±9.7
10	69.1±59.1	4.4±1.9	39.3±38	14.7±10.8	4.4±5.3	8.5±5.3
11	43±38.8	22.6±21.1	30.2±26.2	3.3±2.2	5.8±3.5	40.3±25.5
13	120.4±78.3	68.2±43.9	54.2±36.8	21.8±7.4	17.7±6.1	47.4±27.3
14	76±44.8	28±19.5	65.9±64.3	-0.3±1.3	2.5±2	31.1±26.5
15	58.9±41.4	19.5±15.1	55.4±51	2.5±2	5.5±3.6	20.1±15.7

[0530]

[0531] **표 I**: VEE-MAG25mer srRNA-LNP1 (300 μg)(군 3)로 프라이밍 백신접종에 대한 각 에피토프 ± SEM에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	19.5±8.7	13.3±13.1	16.5±15.3	10.5±7.3	35.9±24.8	92.9±91.6
5	87.9±43.9	12.7±11.7	37.2±31.9	21.1±23.8	13.2±13.7	12.6±13.7
7	21.1±13.3	48.8±48.4	51.7±39.5	9.1±10.5	58.6±55.8	159.4±159
8	47.7±21.7	66.4±52.2	59.8±57.4	49.4±28	79.4±63	133.8±132.1
10	49±30.2	42.2±41.1	139.3±139.3	51.6±51.2	78.2±75.8	131.7±131.6
11	42±26.8	20.9±21.4	177.1±162	-6.3±4.3	104.3±104.1	231.5±230.1
12	40.2±19	20.3±11.9	42.2±46.7	3.7±6.7	57±44.7	70±69.2
13	81.2±48.9	38.2±37.6	259.4±222.2	-4±4.1	164.1±159.3	347.3±343.5
14	34.5±31.8	5.3±11.6	138.6±137.3	-4.7±5.2	52.3±52.9	142.6±142.6
15	49±24	6.7±9.8	167.1±163.8	-6.4±4.2	47.8±42.3	116.6±114.5

[0532]

[0533] **srRNA 용량 범위 연구**

[0534] 본 발명의 한 구현에서, srRNA 용량 범위 연구는 NHP 면역원성 연구로 진행하기 위한 srRNA 용량을 확인하기 위해 mamu A01 인도 붉은털 원숭이에서 수행될 수 있다. 한 예에서, Mamu A01 인도 붉은털 원숭이는 IM 주사에 의해 다중 mamu A01 제한된 에피토프를 포함하는 모델 항원을 코딩하는 srRNA 벡터로 투여받을 수 있다. 다른 예에서, 항-CTLA-4 모노클로날 항체는 동물의 한 군에서 백신 배수 림프절을 표적화하기 위해 IM 백신 주사 부위에 근접하게 SC 투여될 수 있다. PBMC는 면역 모니터링을 위해 초기 백신 접종 후 2주마다 수집할 수 있다. 연구 부문은 아래에 설명되어 있다(표 J).

[0535] **표 J**: 인도 붉은털 원숭이의 비-GLP RNA 용량 범위 연구

군	프라임	부스트 1	부스트 2
1	srRNA-LNP (저용량)	srRNA-LNP (저용량)	srRNA-LNP (저용량)
2	srRNA-LNP (중간 용량)	srRNA-LNP (중간 용량)	srRNA-LNP (중간 용량)
3	srRNA-LNP (고용량)	srRNA-LNP (고용량)	srRNA-LNP (고용량)
4	srRNA-LNP (고용량) + anti-CTLA-4	srRNA-LNP (고용량) + anti-CTLA-4	srRNA-LNP (고용량) + anti-CTLA-4

[0536]

[0537] * srRNA의 용량 범위는 고용량 ≤300 μg에서 결정된다.

[0538] **인도 붉은털 원숭이의 면역원성 연구**

[0539] 항원 벡터를 사용하여 면역원성을 입증하기 위해 mamu A01 인도 붉은털 원숭이(NHP)에서 백신 연구를 수행하였다. 도 5는 백신접종 전략을 예시한다. NHP의 3개 군은 ChAdV68.5-WTnt.MAG25mer 및 관문 억제제 항-CTLA-4 항체 이필리무맙(군 5 및 6)으로 또는 관문 억제제 없이(군 4) 면역화되었다. 항체는 정맥내(군 5) 또는 피하(군 6)에 투여되었다. 삼각형은 0주 및 32주에 chAd68 백신접종(1e12 vp/동물)을 나타낸다. 원은 0, 4, 12, 20, 28 및 32주에 알파바이러스 백신접종을 나타낸다.

[0540] 면역화된 NHP에서 CD8+ 항-에피토프 반응의 시간 경과는 chAd-MAG 예방접종 단독(도 6 및 표 K), IV 전달된 관문 억제제를 사용한 chAd-MAG 예방접종(도 7 및 표 L), 및 SC 전달된 관문 억제제를 사용한 chAd-MAG 예방접종(도 8 및 표 M)에 대해 제시된다. 결과는 영장류에서 chAd68 벡터가 CD8+ 반응을 효율적으로 프라이밍하고, 알파바이러스 벡터가 chAD68 백신 프라이밍 반응을 효율적으로 부스트했으며, IV 또는 SC 전달 여부에 관계없이 관문 억제제가 프라이밍 및 부스팅 반응을 모두 증폭시켰고, chAd 벡터가 백신접종 후 재투여되어 면역 반응을 효과적으로 부스팅했음을 입증한다.

[0541] **표 K**: chAd-MAG가 투여된 붉은털 원숭이(군 4)에서 CD8+ 항-에피토프 반응. 평균 SFC/1e6 비장 세포 +/- 표준 오류가 표시된다.

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	173±41.6	373.5±87.3	461.4±74.2	38.4±26.1	94.5±26	609.2±121.9
5	412.7±138.4	987.8±283.3	1064.4±266.9	85.6±31.2	367.2±135.2	1306.8±332.8
6	116.2±41.2	231.1±46.3	268.3±90.7	86.1±42	174.3±61	223.9±38.1
7	287.4±148.7	588.9±173.9	693.2±224.8	92.1±33.5	172.9±55.6	694.6±194.8
8	325.4±126.6	735.8±212	948.9±274.5	211.3±62.7	179.1±50	817.3±185.2
10	312±129.7	543.2±188.4	618.6±221.7	-5.7±4.1	136.5±51.3	309.9±85.6
11	248.5±81.1	348.7±129.8	581.1±205.5	-3.1±4.4	119±51.2	413.7±144.8
12	261.9±68.2	329.9±83	486.5±118.6	-1.2±5.1	132.8±31.8	350.9±69.3
13	389.3±167.7	1615.8±418.3	1244.3±403.6	1.3±8.1	522.5±155	1303.3±385.6
14	406.3±121.6	1616±491.7	1142.3±247.2	6.6±11.1	322.7±94.1	1048.6±215.6
15	446.8±138.7	1700.8±469.1	1306.3±294.4	43±24.5	421.2±87.9	1001.5±236.4
16	686.8±268.8	1979.5±541.7	1616.8±411.8	2.4±7.8	381.9±116.4	1152.8±352.7
17	375.8±109.3	1378.6±561.2	773.1±210.3	-1.4±4.3	177.6±93.7	691.7±245
18	255.9±99.7	1538.4±498.1	498.7±152.3	-5.3±3.3	26.2±13.4	413.9±164.8
19	133±62.6	955.9±456.8	491.1±121.8	-5.7±4.1	50.3±25.4	371.2±123.7
20	163.7±55.8	641.7±313.5	357.9±91.1	2.6±7.5	41.4±24.2	257.8±68.9
21	319.9±160.5	2017.1±419.9	1204.8±335.2	-3.7±5.1	268.1±109.6	924.1±301
22	244.7±105.6	1370.9±563.5	780.3±390	-3.6±5.1	118.2±68.1	473.3±249.3
23	176.7±81.8	1263.7±527.3	838.6±367.9	-5.7±4.1	73.6±49	480.9±163.9
24	236.5±92	1324.7±589.3	879.7±321	-0.4±5.7	104±53.1	498±135.8
25	136.4±52.6	1207.1±501.6	924±358.5	6.2±10.5	74.1±34.4	484.6±116.7
26	278.2±114.4	1645±661.7	1170.2±469.9	-2.9±5.7	80.6±55.8	784.4±214.1
27	159±56.8	961.7±547.1	783.6±366.4	-5±4.3	63.6±27.5	402.9±123.4
28	189.6±75.7	1073.1±508.8	668.3±312.5	-5.7±4.1	80.3±38.3	386.4±122
29	155.3±69.1	1102.9±606.1	632.9±235	34.5±24.2	80±35.5	422.5±122.9
30	160.2±59.9	859±440.9	455±209.1	-3±5.3	60.5±28.4	302.7±123.2
31	122.2±49.7	771.1±392.7	582.2±233.5	-5.7±4.1	55.1±27.3	295.2±68.3
32	119.3±28.3	619.4±189.7	566±222.1	-3.7±5.1	21.9±4.5	320.5±76.4
33	380.5±122	1636.1±391.4	1056.2±205.7	-5.7±4.1	154.5±38.5	988.4±287.7
34	1410.8±505.4	972.4±301.5	319.6±89.6	-4.8±4.2	141.1±49.8	1375.5±296.7
37	130.8±29.2	500±156.9	424.9±148.9	-3.5±4.7	77.7±24.6	207.1±42.4
38	167.7±54.8	1390.8±504.7	830.4±329.1	-5.5±4.1	111.8±43.2	516±121.7

[0542]

[0543] **표 L**: IV 전달된 chAd-MAG + 항-CTLA4 항체(이필리무맙)가 투여된 붉은털 원숭이(군 5)에서 CD8+ 항-에피토프 반응. 평균 SFC/1e6 비장 세포 +/- 표준 오류가 표시된다.

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	1848.1±432.2	1295.7±479.7	1709.8±416.9	513.7±219.8	838.5±221.1	2514.6±246.5
5	1844.1±410.2	2367.5±334.4	1983.1±370.7	732.1±249.4	1429.7±275.3	2517.7±286.5
6	822.4±216.7	1131.2±194.7	796.8±185.8	226.8±70	802.2±101.4	913.5±222.7
7	1147.2±332.9	1066±311.2	1149.8±467.3	267.4±162.6	621.5±283.2	1552.2±395.1
8	1192.7±188.8	1461.5±237.7	1566.9±310.5	522.5±118.6	662.3±142.4	1706±216.7
10	1249±220.3	1170.6±227.7	1297.3±264.7	-0.3±4.4	551.8±90.5	1425.3±142.6
11	934.2±221.7	808±191.3	1003.1±293.4	1.9±4.3	364.2±76.6	1270.8±191.6
12	1106.2±216.6	896.7±190.7	1020.1±243.3	1.3±3.9	436.6±90	1222±155.4
13	2023.8±556.3	3696.7±1.7	2248.5±436.4	-4.5±3.5	2614±406.1	3700±0
14	1278.7±240	2639.5±387	1654.6±381.1	-6±2.1	988.8±197.9	2288.3±298.7
15	1458.9±281.8	2932.5±488.7	1893.4±499	74.6±15.6	1657.8±508.9	2709.1±428.7
16	1556.8±243	2143.8±295.2	2082.8±234.2	-5.8±2.5	1014.6±161.4	2063.7±86.7
17	1527±495.1	2213±677.1	1767.7±391.8	15.1±5.9	633.8±133.9	2890.8±433.9
18	1068.2±279.9	1940.9±204.1	1114.1±216.1	-5.8±2.5	396.6±77.6	1659.4±171.7
19	760.7±362.2	1099.5±438.4	802.7±192.5	-2.4±3.3	262.2±62.2	1118.6±224.2
20	696.3±138.2	954.9±198	765.1±248.4	-1.4±4.4	279.6±89.3	1139±204.5
21	1201.4±327.9	3096±1.9	1901±412.1	-5.8±2.5	1676.3±311.5	2809.3±195.8
22	1442.5±508.3	2944.7±438.6	1528.4±349.6	2.8±5.1	940.7±160.5	2306.3±218.6
23	1400.4±502.2	2757.1±452.9	1604.2±450.1	-5.1±2.3	708.1±162.6	2100.4±362.9
24	1351±585.1	2264.5±496	1080.6±253.8	0.3±6.5	444.2±126.4	1823.7±306.5
25	1211.5±505.2	2160.4±581.8	970.8±235.9	2.5±3.8	450.4±126.9	1626.2±261.3
26	1443±492.5	2485±588	1252.5±326.4	-0.2±6	360.2±92.3	2081.9±331.1
27	896.2±413.3	1686±559.5	751±192.1	-3.7±2.5	247.4±82.8	1364.1±232
28	1147.8±456.9	1912.1±417.1	930.3±211.4	-5.8±2.5	423.9±79.6	1649.3±315
29	1038.5±431.9	1915.2±626.1	786.8±205.9	23.5±8.3	462.8±64	1441.5±249.7
30	730.5±259.3	1078.6±211.5	699.1±156.2	-4.4±2.7	234.4±43.9	1160.6±112.6
31	750.4±328.3	1431±549.9	650.6±141.1	-5.2±3	243.4±56.4	868.9±142.8
32	581.4±227.7	1326.6±505.2	573.3±138	-3.2±4.2	160.8±49.2	936.4±110.4
33	2198.4±403.8	3093.4±123.3	2391.8±378.4	7.1±8.5	1598.1±343.1	2827.5±289.5
34	2654.3±337	2709.9±204.3	1297.5±291.4	0.4±4.2	1091.8±242.9	1924±245.7
37	846.8±301.7	1706.9±196	973.6±149.3	50.5±45.2	777.3±140.2	1478.8±94.3

[0544]

[0545] **표 M**: SC 전달된 chAd-MAG + 항-CTLA4 항체(이필리무맙)가 투여된 붉은털 원숭이(군 6)에서 CD8+ 항-에피토프 반응. 평균 SFC/1e6 비장 세포 +/- 표준 오류가 표시된다.

주	항원					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	598.3±157.4	923.7±306.8	1075.6±171.8	180.5±74.1	752.3±245.8	1955.3±444.4
5	842.2±188.5	1703.7±514.2	1595.8±348.4	352.7±92.3	1598.9±416.8	2163.7±522.1
6	396.4±45.3	728.3±232.7	503.8±151.9	282±69	463.1±135.7	555.2±191.5
7	584.2±177	838.3±254.9	1013.9±349.4	173.6±64.3	507.4±165.7	1222.8±368
8	642.9±134	1128.6±240.6	1259.1±163.8	366.1±72.8	726.7±220.9	1695.6±359.4
10	660.4±211.4	746.9±222.7	944.8±210.2	-1.2±1.9	523.4±230.7	787.3±308.3
11	571.2±162	609.4±194.3	937.9±186.5	-8.9±2.3	511.6±229.6	1033.3±315.7
12	485.3±123.7	489.4±142.7	919.3±214.1	-8.9±2.3	341.6±139.4	1394.7±432.1
13	986.9±154.5	2811.9±411.3	1687.7±344.3	-4.1±5.1	1368.5±294.2	2751±501.9
14	945.9±251.4	2027.7±492.8	1386.7±326.7	-5.7±2.8	708.9±277.1	1588.2±440.1
15	1075.2±322.4	2386±580.7	1606.3±368.1	-5.4±3.2	763.3±248.8	1896.5±507.8
16	1171.8±341.6	2255.1±439.6	1672.2±342.3	-7.8±2.4	1031.6±228.8	1896.4±419.9
17	1118.2±415.4	2156.3±476	1345.3±377.7	-1.1±6.7	573.7±118.8	1614.4±382.3
18	861.3±313.8	2668.2±366.8	1157.2±259.6	-8.9±2.3	481.2±164	1725.8±511.4
19	719.2±294.2	1447.2±285	968±294.5	-2.2±4.6	395.6±106.1	1199.6±289.2
20	651.6±184	1189.8±242.8	947.4±249.8	-8.9±2.3	355±106.3	1234.7±361.7
21	810.3±301.9	2576.2±283.7	1334±363.1	-8.9±2.3	892.2±305	1904.4±448.1
22	775±196.4	2949±409.7	1421.8±309.7	38±27.8	577±144.2	2330.6±572.3
23	584.9±240.2	1977.9±361.4	1209.8±405.1	-7.3±3.2	273.7±93.3	1430.6±363.9
24	485.4±194.4	1819.8±325.5	837.2±261.4	-3.4±4.1	234.4±71.1	943.9±243.3
25	452.3±175	2072±405.7	957.1±293.1	-8.9±2.3	163±43.2	1341.2±394.7
26	517.9±179.1	2616±567.5	1126.6±289	-8.3±2.3	199.9±89.2	1615.7±385.6
27	592.8±171.7	1838.3±372.4	749.3±170.4	-7.3±2.5	325.5±98.7	1110.7±308.8
28	793±228.5	1795.4±332.3	1068.7±210.3	2.5±4.1	553.1±144.3	1480.8±357.1
29	661.8±199.9	2140.6±599.3	1202.7±292.2	-8.7±2.8	558.9±279.2	1424.2±408.6
30	529.1±163.3	1528.2±249.8	840.5±218.3	-8.9±2.3	357.7±149.4	1029.3±335
31	464.8±152.9	1332.2±322.7	726.3±194.3	-8.9±2.3	354.3±158.6	884.4±282
32	612.9±175.3	1584.2±390.2	1058.3±219.8	-8.7±2.8	364.6±149.8	1388.8±467.3
33	1600.2±416.7	2597.4±367.9	2086.4±414.8	-6.3±3.3	893.8±266	2490.6±416.4
34	2814.6±376.2	2713.6±380.8	1888.8±499.4	-7.5±3.1	1288.9±438.9	2428.1±458.9
37	1245.9±471.7	1877.7±291.2	1606.6±441.9	14.2±13	1227.5±348.1	1260.7±342

[0546]

[0547] **인도 붉은털 원숭이의 기억 표현형**

[0548] 붉은털 원숭이 원숭이를 항-CTLA4와 함께 또는 없이 ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA 이종 프라임/부스트 요법으로 면역화하고, ChAdV68.5WTnt.MAG25mer로 다시 부스트시켰다. 군은 최종 ChAdV68 투여 후 11개월(연구 18개월) 후 평가되었다. ELISpot에 의해 설명된 대로 수행되었다. 도 9 및 표 N은 예방접종 전(왼쪽 패널)과 18개월 후(오른쪽 패널) 모두 ELISpot에 의해 측정된 6개의 다른 Mamu-A*01 제한 에피토프에 대한 세포 반응을 보여준다. 제한된 에피토프에 대한 반응의 검출은 항원-특이적 기억 반응이 ChAdV68/samRNA 백신 프로토콜에 의해 생성되었음을 입증한다.

[0549] 기억력을 평가하기 위해, 백신에서 코딩된 4개 다른 붉은털 원숭이 Mamu-A*01 클래스 I 에피토프를 인식하는 CD8+ T-세포를 이중 색상 Mamu-A*01 사량체 표지를 사용하여 모니터링했으며, 각 항원은 고유한 이중 양성 조합을 통해 표시되며단일 샘플 내에서 4개 항원-특이적 집단을 모두 확인할 수 있었다. 기억 세포 표현형은 세포 표면 마커 CD45RA 및 CCR7과 함께 공동-염색하여 수행되었다. 도 10 및 표 O는 4개의 상이한 Mamu-A*01 제한된 에피토프를 인식하는 기억 T-세포에 대한 조합 사량체 염색 및 CD45RA/CCR7 공동-염색의 결과를 나타낸다. T 세포 표현형은 또한 유세포 분석에 의해 평가되었다. 도 11은 연구 18개월에 4개의 Mamu-A*01 사량체+ CD8+ T-세포 집단의 합 내에서 기억 세포 유형의 분포를 보여준다. 기억 세포는 다음과 같이 특성화되었다: CD45RA+CCR7+=나이브, CD45RA+CCR7-= 효과기(Teff), CD45RA-CCR7+=중추 기억(Tcm), CD45RA-CCR7-=효과기 기억(Tem). 종합적으로, 결과는 기억 반응이 효과기, 중추 기억 및 효과기 기억 집단을 포함하여 오래 지속되는 면역을 나타내는 마지막 부스트 후 적어도 1년 후에 감지되었음을 보여준다.

[0550] **표 N** 프라임 전 및 기억 평가 시점(18개월)에서 각 동물에 대한 10^6 PBMC 당 평균 반점 형성 세포(SFC)

동물	프라임 전 기준선						18 개월					
	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9
1	1.7	0.0	0.0	5.0	0.0	13.7	683.0	499.2	1100.3	217.5	47.7	205.3
2	0.0	0.0	0.0	0.2	0.1	0.0	773.4	ND	1500.0	509.3	134.5	242.5
3	0.0	0.0	6.7	6.8	10.2	3.3	746.3	167.5	494.1	402.8	140.6	376.0
4	0.0	0.0	0.0	0.0	0.0	0.0	47.6	1023.9	85.1	44.2	44.2	47.6
5	15.3	6.7	18.6	15.6	5.2	12.1	842.4	467.7	1500.0	805.9	527.8	201.8
6	3.1	0.0	0.0	15.5	6.9	5.3	224.3	720.3	849.0	296.9	32.4	121.9

[0551]

[0552] ND=기술적 배제로 인해 결정되지 않음

[0553] **표 O** 살아있는 CD8+ 세포 중 Mamu-A*01 사량체 양성 퍼센트

동물	Tat TL8	Gag CM9	Env TL9	Env CL9
1	0.42	0.11	0.19	0.013
2	0.36	0.048	0.033	0.00834
3	0.97	0.051	0.35	0.048
4	0.46	0.083	0.17	0.028
5	0.77	0.45	0.14	0.2
6	0.71	0.16	0.17	0.04

[0554]

[0555] **XVII. 인플루엔자 연구**

[0556] ChAdV68 및 자가-복제 RNA(srRNA)를 사용하는 다양한 투여 프로토콜을 인플루엔자 백신에 대해 평가하였다.

[0557] **재료 및 방법**

[0558] **인플루엔자 ChAdV68 및 SAM 벡터**

[0559] 하기 인플루엔자 A 균주로부터 인플루엔자 혈구응집소(HA)를 발현하는 ChAdV68 및 SAM 인플루엔자 벡터를 제조하였다: H1N1 캘리포니아7/09 및 H7N9 Anhui1/13. HA 뉴클레오티드 서열은 HA 단백질의 최적화된 포유동물 발현을 위해 최적화된 코돈이었다(서열 참조). 구성물은 깁슨(Gibson) 어셈블리에 의해 ChAdV68 및 SAM 플라스미드로 클로닝되었다. 사용된 ChAdV68 벡터 백본은 다음과 같다: E1(nt 577-3403) 및 E3(nt 27,125-31,825) 서열이 결실된 AC_000011.1; 5개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드; 결실된 E1 대신 삽입된 CMV 프로모터/인핸서의 제어 하에 있는 항원. ChAdV68 플라스미드를 PacI 분해하고 293F 세포로 형질주입시켜 ChAd-HA 바이러스 벡터를 생성했다. ChAd 생산은 293F 현탁 세포에서 400 mL 규모로 수행되었다. 바이러스를 불연속 CsCl 밀도 구배의 2회 라운드로 정제하고 바이러스를 정제하고 ARM 완충액(10 mM Tris pH8.0, 25 mM NaCl, 2.5% 글리세롤)에 투석했다. 벡터를 동물에 투여하기 위한 표적 용량으로 ARM 완충액에 희석하였다. 사용된 SAM 벡터 백본은 다음과 같다: 26S 하위-게놈 프로모터의 3'에 위치한 VEE의 구조적 단백질을 코딩하는 서열을 갖는 베네수엘라 말 뇌염(VEE)(Kinney, 1986, Virology 152: 400-413) 기반 자가-증폭 RNA(SAM) 벡터는 결실되고(VEE 서열 7544에서 11,175가 결실되었고; Kinney 등 1986을 기반으로 하는 넘버링; 서열번호: 6) 및 항원 서열로 대체되었다. H1N1캘리포니아7/09 또는 H7N9 Anhui1/13 균주의 HA를 코딩하는 고품질 전장 SAM을 시험관내에서 합성하고 및 공동-전사 캡핑하고 실리카 컬럼을 통해 정제했다. 그런 다음 SAM을 지질 나노입자(LNP)에 캡슐화하고 저장 완충액에 보관했다.

[0560] **Balb/C 마우스에서 생체내 평가**

[0561] 체액 평가를 위해, ChAd-CA709-HA 및 ChAd-Anhui1/13-HA 벡터를 5e10 VP 또는 1e9 VP/동물로 투여하였다. 혈청을 주사 전 및 프라임 후 28일째 및 SAM 부스트 후 4주째인 56일째에 수확하였다. 혈청의 한계 희석을 적혈구의 혈구응집을 방지하는 능력에 대해 분석하고 혈구응집 억제(HAI) 역가를 결정했다.

[0562] T 세포 반응 평가를 위해 Balb/c 마우스를 5e10 VP ChAd-CA709-HA 또는 10 ug SAM-CA709-HA로 면역화하고 항원 특이적 T 세포 면역 반응을 평가했다. 프라임 후 2주에, 비장세포를 분리하고 인플루엔자 A CA/7/09의 HA 단백질에 걸쳐 있는 중첩 펩티드의 6개 미니풀(15 AA 길이, 11 AA 중첩)로 자극한 후 밤새 IFNɣ ELISpot에 의해 T 세포 반응을 평가했다.

[0563] 마우스는 또한 ChAdV68-프라임 및 SAM-부스트 요법으로 면역화되었고 여기서 마우스는 ChAd-HA(5e10 vp)으로 면역화된 후 8주에 SAM-HA로 부스트되었고 T 세포 반응은 부스트 후 2주(프라임 후 10주) IFNɣ ELISpot에 의해

평가되었다. T 세포 다작용성은 또한 chAd & SAM-HA를 사용한 프라임/부스트 후 IFNγ, TNFα 또는 IL-2를 발현하는 T 세포를 확인하기 위해 세포내 사이토카인 염색(ICS)에 의해 부스트 후 2주째에 비장세포에서 평가되었다.

[0564] **결과**

[0565] 인플루엔자 백신을 생체내에서 평가하였다.

[0566] 예방접종 전 및 후의 인플루엔자 A H1N1 균주 CA/7/09에 대한 HAI 역가는 도 12에 도시되고 표 P에서 정량화되었다. 동물 당 고용량의 5e10 vp로 투여된 마우스(도 12a)는 동물의 100% 혈청전환을 나타내었고 높은 역가 >300 HAI를 가졌고 모든 동물은 보호적일 것으로 예측되는 HAI 역가 >40를 가졌다. 저용량에서(도 12b), 5/6 동물이 HAI 역가 > 40을 가졌고 이러한 역가는 생체내에서 보호될 가능성이 있는 성공적인 혈청전환을 나타내는 4배 초과의 사전-샘플이었다. 예방접종 전 및 후의 인플루엔자 A H7N9 Anhui1/13 균주에 대한 HAI 역가가 도 13에 도시되고 표 Q에 정량화되어 있다(도 13a 고용량; 도 13b 저용량). H7N9 균주에 대해 유사하지만 일반적으로 더 낮은 반응이 관찰되었으며, 모든 동물은 5e10 vp의 고용량에서 혈청전환되었다. 결과는 ChAd-HA 프라이밍이 H1N1 및 H7N9 인플루엔자 균주에 대한 강한 체액성 면역 반응을 유도함을 입증한다.

[0567] ChAdV68-프라임(도 14a), SAM-프라임(도 14b) 또는 ChAdV68-프라임 및 SAM-부스트(도 14c)을 갖는 인플루엔자 A H1N1 균주 CA/7/09에 대한 항-HA T 세포 반응이 도 14에 도시되고 표 R에서 정량화되어 있다. ChAdV68-프라임은 인터페론 감마 ELISpot에 의해 측정된 바와 같이 펩티드 풀에 걸쳐 더 강력한 T 세포 반응을 생성했다. SAM-프라임은 인터페론 감마 ELISpot에 의해 측정된 바와 같이 펩티드 풀에 걸쳐 보다 가변적이고 일반적으로 감소된 T 세포 반응을 생성했다. ChAdV68-프라임 및 SAM-부스트는 두 벡터 중 하나를 사용한 프라임 후 반응과 비교하여 상당히 증가된 T 세포 반응을 보여주었다. CD8 T 세포(도 15a) 및 CD4 T 세포(도 15b)에 대한 ChAdV68-프라임 및 SAM-부스트를 갖는 인플루엔자 A H1N1 균주 CA/7/09에 대한 다작용성 T 세포 반응이 도 15 에 도시되고 표 S에서 정량화되어 있다. T 세포 반응은 현저한 CD8 T 세포 편향을 보여주었다. 결과는 특히 ChAd-HA 프라이밍 뿐만 아니라 SAM-기반 백신으로의 후속 부스팅을 통해 백신 요법이 H1N1 인플루엔자 균주에 대해 강력한 다기능 T 세포 반응을 유도했음을 보여준다.

[0568] **표 P** - 예방접종 전 및 후의 H1N1 HAI 역가("1-X" 5e10 VP, "2-X" 1e9 VP)

마우스 #	GMT 백신접종 전	GMT 백신접종 후
1-1	5	320
1-2	5	320
1-3	5	640
1-4	5	640
1-5	5	320
1-6	5	320
2-1	5	5
2-2	5	160
2-3	5	80
2-4	5	40
2-5	5	160
2-6	5	40

[0569]

[0570] **표 Q** - 예방접종 전 및 후의 H7N9 HAI 역가("3-X" 5e10 VP, "4-X" 1e9 VP)

마우스 #	GMT 백신접종 전	GMT 백신접종 후
3-1	5	56.57
3-2	5	80
3-3	5	56.57
3-4	5	226.27
3-5	5	320
3-6	5	160
4-1	5	22.28
4-2	5	56.57
4-3	5	10
4-4	5	20
4-5	5	40
4-6	5	80

[0571]

[0572] **표 R** - H1N1 T 세포 반응(인터페론 감마 ELISpot)

		미니풀								
치료	동물 ID	1	2	3	4	5	6	합	평균	SE
ChAd-HA 5e10 vp	7	34	966	638	7	4116	2304	8065	10630	1732
	8	307	5526	1607	169	4790	5913	18312		
	9	81	1538	745	38	4926	2600	9927		
	10	59	1832	457	3	2463	1873	6687		
	11	59	3855	1307	22	5630	1556	12430		
	12	145	6100[a]	2504	98	6100[a]	5613	8360		
SAM-HA 10 ug	21	487	4240	1373	374	5366	4515	16355	6313	2467
	22	354	3649	1451	231	2213	3101	11000		
	23	49	1707	299	0	1332	476	3863		
	24	138	1519	504	30	176	1339	3706		
	25	104	648	217	0	87	666	1721		
	26	112	400	213	30	215	261	1232		
ChAd-HA + SAM-HA 프라임/부스트	1	696	11991	2356	525	25000[a]	4441	20009	45546	5690
	2	247	15232	3041	454	21429	14969	55371		
	3	276	16216	2078	465	23563	10576	53174		
	4	489	19196	2849	347	20170	9534	52584		
	5	292	12085	2109	572	20931	2514	38503		
	6	162	21544	2986	233	17148	11561	53635		

[0573]

[0574] **표 S** - H1N1 다작용성 T 세포 반응(사이토카인 ICS)

	% CD8+			% CD4+		
	IFNg	**TNFa**	**IL2**	**IFNg**	**TNFa**	**IL2**
	32.25	23.25	0.59	0.94	1.02	0.47
	25.67	16.70	0.47	0.90	1.01	0.44
	25.98	17.48	0.77	0.68	1.04	0.42
	31.97	22.10	0.55	1.38	1.43	0.61
	26.27	18.08	0.88	0.77	0.91	0.40
	40.89	26.66	1.02	0.99	0.81	0.35
평균	**30.51**	**20.71**	**0.71**	**0.94**	**1.04**	**0.45**
SE	**2.41**	**1.60**	**0.09**	**0.10**	**0.09**	**0.04**

[0575]

[0576] 사이토카인 반응은 개별 동물에서 나온 것이며 2개의 중첩 펩티드 풀에 대한 반응의 합이다.

[0577] **특정 서열**

[0578] 본원에 언급된 벡터, 카세트 및 항체는 하기 기재되어 있고 서열번호로 언급된다.

트레멜리무맙 VL (서열번호:16)
트레멜리무맙 VH (서열번호:17)
트레멜리무맙 VH CDR1 (서열번호:18)
트레멜리무맙 VH CDR2 (서열번호:19)
트레멜리무맙 VH CDR3 (서열번호:20)
트레멜리무맙 VL CDR1 (서열번호:21)
트레멜리무맙 VL CDR2 (서열번호:22)
트레멜리무맙 VL CDR3 (서열번호:23)
더발루맙 (MEDI4736) VL (서열번호:24)
MEDI4736 VH (서열번호:25)
MEDI4736 VH CDR1 (서열번호:26)
MEDI4736 VH CDR2 (서열번호:27)
MEDI4736 VH CDR3 (서열번호:28)
MEDI4736 VL CDR1 (서열번호:29)
MEDI4736 VL CDR2 (서열번호:30)
MEDI4736 VL CDR3 (서열번호:31)
UbA76-25merPDTT 뉴클레오티드 (서열번호:32)
UbA76-25merPDTT 폴리펩티드 (서열번호:33)
MAG-25merPDTT 뉴클레오티드 (서열번호:34)
MAG-25merPDTT 폴리펩티드 (서열번호:35)
Ub7625merPDTT_NoSFL 뉴클레오티드 (서열번호:36)
Ub7625merPDTT_NoSFL 폴리펩티드 (서열번호:37)
ChAdV68.5WTnt.MAG25mer (서열번호:2); E1 (nt 577 내지 3403) 및 E3 (nt 27,125-31,825) 서열이 결실된 AC_000011.1; 5개의 위치에서 치환된 상응하는 ATCC VR-594 뉴클레오티드; 결실된 E1 대신 삽입된 CMV 프로모터/인핸서의 제어 하에 있는 모델 신생항원 카세트; 카세트의 SV40 폴리A 3′
베네수엘라 말 뇌염 바이러스 [VEE] (서열번호:3) GenBank: L01442.2
VEE-MAG25mer (서열번호:4); MAG-25merPDTT 뉴클레오티드 함유(염기 30-1755)
베네수엘라 말 뇌염 바이러스 균주 TC-83 [TC-83](서열번호:5) GenBank: L01443.1
VEE 전달 벡터 (서열번호:6); 뉴클레오티드 7544-11175가 결실된 VEE 게놈 [알파바이러스 구조 단백질 제거]
TC-83 전달 벡터 (서열번호:7); 뉴클레오티드 7544-11175가 결실된 TC-83 게놈 [알파바이러스 구조 단백질 제거]
VEE 생산 벡터 (서열번호:8); 뉴클레오티드 7544-11175가 결실되고, + 5′ T7-프로모터, + 3′ 제한 부위를 갖는 VEE 게놈
TC-83 생산 벡터 (서열번호:9); 뉴클레오티드 7544-11175가 결실되고, + 5′ T7-프로모터, + 3′ 제한 부위를 갖는 TC-83 게놈
VEE-UbAAY (서열번호:14);MHC 클래스 I 마우스 종양 에피토프 SIINFEKL 및 AH1-A5가 삽입된 VEE 전달 벡터
VEE-루시퍼라제 (서열번호:15); 7545에 루시퍼라제 유전자가 삽입된 VEE 전달 벡터

[0579]

유비퀴틴 (서열번호:38)>UbG76 0-228
유비퀴틴 A76 (서열번호:39)>UbA76 0-228
HLA-A2 (MHC 클래스 I) 신호 팹티드 (서열번호:40)>MHC 신호Pep 0-78
HLA-A2 (MHC 클래스 I) 막횡단 도메인 (서열번호:41)>HLA A2 TM 도메인 0-201
IgK 리더서열 (서열번호:42)>IgK 리더서열 0-60
인간 DC-Lamp (서열번호:43)>인간DCLAMP 0-3178
마우스 LAMP1 (서열번호:44)>마우스Lamp1 0-1858
인간 Lamp1 cDNA (서열번호:45)>인간 Lamp1 0-2339
파상풍 톡소이드 핵산 서열 (서열번호:46)
파상풍 톡소이드 아미노산 서열 (서열번호:47)
PADRE 뉴클레오티드 서열 (서열번호:48)
PADRE 아미노산 서열 (서열번호:49)
WPRE (서열번호:50)>WPRE 0-593
IRES (서열번호:51)>eGFP_IRES_SEAP_삽입 1746-2335
GFP (서열번호:52)
SEAP (서열번호:53)
반딧불이 루시퍼라제 (서열번호:54)
FMDV 2A (서열번호:55)
이필리무맙 중쇄 (서열번호: 29365)
QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYTMHWVRQAPGKGLEWVTFISYDGNNKYYADSVKGRFTISRDNS KNTLYLQMNSLRAEDTAIYYCARTGWLGPFDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDY FPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTH TCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNST YRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYP SDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
이필리무맙 경쇄 (서열번호: 29366)
EIVLTQSPGTLSLSPGERATLSCRASQSVGSSYLAWYQQKPGQAPRLLIYGAFSRATGIPDRFSGSGSGTDFTLT ISRLEPEDFAVYYCQQYGSSPWTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQ주 VDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
니볼루맙 중쇄 (서열번호: 29367)
QVQLVESGGGVVQPGRSLRLDCKASGITFSNSGMHWVRQAPGKGLEWVAVIWYDGSKRYYADSVKGRFTISRDNS KNTLFLQMNSLRAEDTAVYYCATNDDYWGQGTLVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPV TVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPPCPPCPAP EFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLT VLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWE SNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK
니볼루맙 경쇄 (서열번호: 29368)
EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTI SSLEPEDFAVYYCQQSSNWPRTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQ주V DNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
인플루엔자 A H1N1 CA/7/09 HA
ATGAAAGCTATCCTTGTAGTCCTTCTCTATACTTTTGCCACAGCTAACGCGGATACGCTGTGCATCGGGTATCAC GCAAATAACTCCACCGATACGGTGGACACGGTGCTTGAGAAGAATGTAACTGTAACTCATTCCGTGAACTTGCTG GAGGACAAACACAACGGGAAGTTGTGCAAGCTCAGGGGGGTCGCGCCGTTGCATTTGGGAAAATGTAATATCGCT GGATGGATCTTGGGTAATCCCGAGTGCGAAAGCCTCAGCACCGCCAGCAGCTGGAGCTACATTGTGGAAACTCCC TCAAGCGATAACGGGACCTGTTACCCAGGGGATTTCATCGACTATGAGGAATTGCGGGAACAGTTGAGTTCTGTG TCTTCATTTGAACGATTtGAAATTTTCCCCAAGACCAGTTCTTGGCCTAATCACGACTCTAACAAGGGTGTTACG GCAGCATGCCCGCATGCCGGAGCAAAGAGTTTCTACAAAAATTTGATCTGGTTGGTGAAGAAGGGCAACTCATAC CCTAAGCTCAGTAAGTCTTATATCAATGACAAAGGAAAAGAAGTACTGGTTTTGTGGGGAATCCACCATCCGTCC ACATCTGCAGACCAACAGTCACTCTACCAGAACGCGGATGCCTACGTTTTTGTGGGAAGCTCAAGATATTCCAAA

[0580]

AAATTCAAGCCTGAGATTGCTATTCGCCCAAAGGTCCGCGACCAAGAAGGCAGGATGAATTACTACTGGACCTTG
GTCGAGCCTGGTGATAAGATTACATTTGAAGCCACCGGTAACCTTGTTGTCCCGAGGTACGCCTTCGCGATGGAG
CGGAATGCAGGGTCAGGGATTATTATATCAGACACCCCAGTACACGACTGCAACACAACTTGTCAGACCCCTAAG
GGTGCCATCAATACATCCCTGCCGTTTCAGAATATCCATCCGATCACTATAGGCAAGTGTCCAAAATATGTGAAG
AGCACGAAGCTTAGGCTGGCGACCGGATTGCGGAACATACCTTCTATCCAGAGTCGCGGGCTCTTCGGAGCTATC
GCGGGCTTTATAGAGGGAGGATGGACTGGAATGGTAGATGGATGGTATGGTTACCACCATCAGAACGAACAGGGA
TCCGGGTACGCAGCAGATTTGAAATCAACACAGAACGCCATCGACGAGATCACCAATAAGGTGAACTCTGTAATT
GAAAAAATGAATACGCAATTCACTGCAGTGGGGAAGGAATTCAACCATCTGGAGAAACGAATTGAAAACCTTAAC
AAGAAGGTAGATGACGGTTTCCTCGATATCTGGACATATAATGCAGAACTTTTGGTATTGCTGGAAAATGAACGG
ACCCTGGATTATCACGATTCAAACGTTAAAAATCTCTATGAGAAGGTTCGATCTCAACTGAAGAACAACGCCAAG
GAAATAGGAAACGGATGTTTCGAGTTCTATCATAAATGCGATAACACATGCATGGAGAGCGTCAAGAACGGTACC
TACGACTATCCTAAGTATAGTGAGGAAGCCAAACTCAATAGGGAAGAGATCGACGGAGTCAAATTGGAATCAACG
CGAATATATCAGATTCTTGCAATTTACAGCACTGTCGCGAGTAGCCTTGTGTTGGTTGTGAGCCTCGGTGCTATT
TCCTTTTGGATGTGCTCAAACGGCTCTCTCCAGTGTAGAATCTGCATTTGA

인플루엔자 A H7N9 Anhui/1/13 HA

ATGAATACCCAAATATTGGTATTCGCTCTGATTGCAATTATTCCGACTAACGCAGATAAGATTTGCCTTGGCCAC
CATGCTGTGAGTAATGGAACTAAAGTTAACACACTTACCGAACGGGGCGTTGAAGTCGTGAACGCCACAGAGACA
GTCGAGAGAACAAACATTCCACGAATATGCAGCAAAGGCAAACGAACTGTAGACCTCGGGCAATGCGGCCTCCTC
GGTACCATTACCGGTCCTCCGCAGTGTGACCAGTTTCTGGAGTTTTCAGCCGACCTCATCATTGAGCGACGGGAG
GGAAGCGACGTCTGCTACCCTGGGAAGTTCGTGAACGAAGAAGCGCTGCGGCAGATTTTGAGAGAAAGTGGTGGT
ATAGATAAAGAGGCGATGGGCTTCACGTATTCCGGTATAAGGACAAATGGTGCCACGTCTGCATGCAGGCGGAGC
GGCAGCTCTTTTTACGCAGAGATGAAATGGCTGTTGTCAAACACCGATGATGCGGCTTTTCCTCAAATGACCAAA
AGCTATAAGAACACCAGGAAATCCCCCGCACTCATAGTCTGGGGTATACATCACTCAGTGTCCACAGCAGAACAA
ACGAAACTTTATGGGTCTGGCAACAAACTGGTGACAGTGGGGTCCTCTAACTATCAACAAAGCTTTGTGCCATCA
CCAGGGGCTCGACCACAAGTAAACGGACTCAGTGGGCGGATCGACTTCCACTGGTTGATGCTCAATCCTAACGAT
ACCGTTACCTTCTCTTTCAACGGAGCCTTCATAGCGCCTGACAGGGCCAGCTTTCTCAGGGGTAAATCCATGGGG
ATACAGTCAGGGGTCCAGGTGGATGCCAATTGCGAAGGCGATTGCTATCACTCTGGCGGAACAATAATCTCCAAT
CTCCCGTTCCAGAATATTGACTCACGGGCAGTAGGGAAATGTCCCCGCTATGTCAAACAGAGGAGCTTGCTGCTT
GCAACCGGCATGAAGAACGTGCCTGAAATACCCAAAGGTAGGGGCCTTTTCGGGGCTATCGCGGGATTTATCGAA
AACGGGTGGGAGGGACTCATCGACGGCTGGTACGGCTTTAGGCATCAAAACGCGCAAGGTGAAGGCACGGCAGCT
GACTACAAGAGCACGCAGTCTGCCATCGATCAGATAACCGGGAAACTTAATCGCCTGATCGAAAAGACAAATCAG
CAATTTGAACTCATCGACAATGAATTTAACGAAGTGGAGAAGCAAATTGGCAATGTCATCAACTGGACAAGAGAC
TCAATTACGGAAGTTTGGAGCTACAATGCTGAATTGCTTGTAGCAATGGAAAACCAGCATACGATAGACCTGGCT
GATTCTGAGATGGACAAGCTCTATGAGCGGGTAAAAAGGCAGCTCCGAGAAAACGCCGAGGAGGACGGTACGGGA
TGCTTCGAGATTTTCCATAAGTGTGACGACGACTGTATGGCAAGTATCCGAAATAACACTTACGATCATTCAAAA
TACCGGGAGGAAGCTATGCAAAACAGAATACAGATTGACCCGGTTAAATTGAGCAGCGGCTATAAAGATGTGATC
CTGTGGTTTAGCTTCGGAGCTTCCTGTTTCATTCTTCTGGCAATAGTTATGGGTCTTGTATTTATTTGTGTAAAA
AACGGGAATATGCGATGCACGATCTGTATCTGA

[0581]

[0582] **참고문헌**

[0583] 1. Desrichard, A., Snyder, A. & Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-3175

[0584] 2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348,** 69-74 (2015).

[0585] 3. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* **125,** 3413-3421 (2015).

[0586] 4. Rizvi, N. A. 등 Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348,** 124-128 (2015).

[0587] 5. Snyder, A. 등 Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371,** 2189-2199 (2014).

[0588] 6. Carreno, B. M. 등 Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348,** 803-808 (2015).

[0589] 7. Tran, E. 등 Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344,** 641-645 (2014).

[0590] 8. Hacohen, N. & Wu, C. J.-Y. United States Patent Application: 20110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). at <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PGNR.>

[0591] 9. Lundegaard, C., Hoof, I., Lund, O. & Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res.* **6 Suppl 2,** S3 (2010).

[0592] 10. Yadav, M. 등 Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515,** 572-576 (2014).

[0593] 11. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* **14**, 658-673 (2015).

[0594] 12. Van Allen, E. M. 등 Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211 (2015).

[0595] 13. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5**, 445-459 (2014).

[0596] 14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550 (2014).

[0597] 15. Rajasagi, M. 등 Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453-462 (2014).

[0598] 16. Downing, S. R. 등 United States Patent Application: 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). at <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PGNR.>

[0599] 17. Target Capture for NextGen Sequencing - IDT. at <http://www.idtdna.com/pages/products/nextgen/target-capture>

[0600] 18. Shukla, S. A. 등 Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152-1158 (2015).

[0601] 19. Cieslik, M. 등 The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372-1381 (2015).

[0602] 20. Bodini, M. 등 The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* **125**, 600-605 (2015).

[0603] 21. Saunders, C. T. 등 Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28**, 1811-1817 (2012).

[0604] 22. Cibulskis, K. 등 Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213-219 (2013).

[0605] 23. Wilkerson, M. D. 등 Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* **42**, e107 (2014).

[0606] 24. Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. & Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* **30**, 2813-2815 (2014).

[0607] 25. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* **25**, 2865-2871 (2009).

[0608] 26. Lam, H. Y. K. 등 Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28**, 47-55 (2010).

[0609] 27. Frampton, G. M. 등 Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023-1031 (2013).

[0610] 28. Boegel, S. 등 HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).

[0611] 29. Liu, C. 등 ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* **41**, e142 (2013).

[0612] 30. Mayor, N. P. 등 HLA Typing for the Next Generation. *PloS One* **10**, e0127153 (2015).

[0613] 31. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. & Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* **4,** (2015).

[0614] 32. Song, L. & Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* **14 Suppl 5,** S14 (2013).

[0615] 33. Maretty, L., Sibbesen, J. A. & Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* **15,** 501 (2014).

[0616] 34. Pertea, M. 등 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33,** 290-295 (2015).

[0617] 35. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355

[0618] 36. Vitting-Seerup, K., Porse, B. T., Sandelin, A. & Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15,** 81 (2014).

[0619] 37. Rivas, M. A. 등 Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348,** 666-669 (2015).

[0620] 38. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* **21,** 1728-1737 (2011).

[0621] 39. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31,** 166-169 (2015).

[0622] 40. Furney, S. J. 등 SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330

[0623] 41. Zhou, Q. 등 A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930

[0624] 42. Maguire, S. L. 등 SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235,** 571-580 (2015).

[0625] 43. Carithers, L. J. 등 A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* **13,** 311-319 (2015).

[0626] 44. Xu, G. 등 RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* **9,** e89445 (2014).

[0627] 45. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639

[0628] 46. Jorgensen, K. W., Rasmussen, M., Buus, S. & Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141,** 18-26 (2014).

[0629] 47. Larsen, M. V. 등 An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **35,** 2295-2303 (2005).

[0630] 48. Nielsen, M., Lundegaard, C., Lund, O. & Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57,** 33-41 (2005).

[0631] 49. Boisvert, F.-M. 등 A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells.

Mol. Cell. Proteomics **11,** M111.011429-M111.011429 (2012).

[0632] 50. Duan, F. 등 Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211,** 2231-2248 (2014).

[0633] 51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books @ Amazon.com. at <http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>

[0634] 52. Calis, J. J. A. 등 Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9,** e1003266 (2013).

[0635] 53. Zhang, J. 등 Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346,** 256-259 (2014)

[0636] 54. Walter, M. J. 등 Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090-1098 (2012).

[0637] 55. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 1992. 255: 1261-1263.

[0638] 56. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, Engelhard VH, Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy.__Proc Natl Acad Sci U S A. 2006 Oct 3;103(40):14889-94.

[0639] 57. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics. 2015 Mar;14(3):658-73. doi: 10.1074/mcp.M114.042812.

[0640] 58. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. Nat Protoc. 2015 Sep;10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 2015 Aug 6

[0641] 59. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. Cell. 1978 May;14(1):9-20.

[0642] 60. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. Br J Haematol. 1982 Nov;52(3):411-20.

[0643] 61. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013 Jan;13(1):22-4. doi: 10.1002/pmic.201200439. Epub 2012 Dec 4.

[0644] 62. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet--implementation and features. J Am Soc Mass Spectrom. 2015 Nov;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 2015 Jun 27.

[0645] 63. Lukas Kall, Jesse Canterbury, Jason Weston, William Stafford Noble and Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature Methods 4:923 - 925, November 2007

[0646] 64. Lukas Kall, John D. Storey, Michael J. MacCoss and William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. Journal of Proteome Research, 7(1):29-34, January 2008

[0647] 65. Lukas Kall, John D. Storey and William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. Bioinformatics, 24(16):i42-i48, August 2008

[0648] 66. Kinney RM, BJ Johnson , VL Brown , DW Trent. Nucleotide Sequence of the 26 S mRNA of the Virulent Trinidad Donkey Strain of Venezuelan Equine Encephalitis Virus and Deduced Sequence of the Encoded Structural Proteins. Virology 152 (2), 400-413. 1986 Jul 30.

[0649] 67. Jill E Slansky, Frederique M Rattis, Lisa F Boyd, Tarek Fahmy, Elizabeth M Jaffee, Jonathan P Schneck, David H Margulies, Drew M Pardoll. Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex. Immunity, Volume 13, Issue 4, 1 October 2000, Pages 529-538.

[0650] 68. A Y Huang, P H Gulden, A S Woods, M C Thomas, C D Tong, W Wang, V H Engelhard, G Pasternack, R Cotter, D Hunt, D M Pardoll, and E M Jaffee. The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. Proc Natl Acad Sci U S A.; 93(18): 9730-9735, 1996 Sep 3.

[0651] 69. JOHNSON, BARBARA J. B., RICHARD M. KINNEY, CRYSTLE L. KOST AND DENNIS W. TRENT. Molecular Determinants of Alphavirus Neurovirulence: Nucleotide and Deduced Protein Sequence Changes during Attenuation of Venezuelan Equine Encephalitis Virus. J Gen Virol 67:1951-1960, 1986.

[0652] 70. Aarnoudse, C.A., Kruse, M., Konopitzky, R., Brouwenstijn, N., and Schrier, P.I. (2002). TCR reconstitution in Jurkat reporter cells facilitates the identification of novel tumor antigens by cDNA expression cloning. Int J Cancer 99, 7-13.

[0653] 71. Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H.M., Kubo, R.T., and Sette, A. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. Immunity 1, 751-761.

[0654] 72. Banu, N., Chia, A., Ho, Z.Z., Garcia, A.T., Paravasivam, K., Grotenbreg, G.M., Bertoletti, A., and Gehring, A.J. (2014). Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections. Scientific Reports 4, 4166.

[0655] 73. Cornet, S., Miconnet, I., Menez, J., Lemonnier, F., and Kosmatopoulos, K. (2006). Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic tumor peptides with enhanced immunogenicity. Vaccine 24, 2102-2109.

[0656] 74. Depla, E., van der Aa, A., Livingston, B.D., Crimi, C., Allosery, K., de Brabandere, V., Krakover, J., Murthy, S., Huang, M., Power, S., 등 (2008). Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. Journal of Virology 82, 435-450.

[0657] 75. Ishioka, G.Y., Fikes, J., Hermanson, G., Livingston, B., Crimi, C., Qin, M., del Guercio, M.F., Oseroff, C., Dahlberg, C., Alexander, J., 등 (1999). Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes. J Immunol 162, 3915-3925.

[0658] 76. Janetzki, S., Price, L., Schroeder, H., Britten, C.M., Welters, M.J.P., and Hoos, A. (2015). Guidelines for the automated evaluation of Elispot assays. Nat Protoc 10, 1098-1115.

[0659] 77. Lyons, G.E., Moore, T., Brasic, N., Li, M., Roszkowski, J.J., and Nishimura, M.I. (2006). Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells. Cancer Res 66, 11455-11461.

[0660] 78. Nagai, K., Ochi, T., Fujiwara, H., An, J., Shirakata, T., Mineno, J., Kuzushima, K., Shiku, H., Melenhorst, J.J., Gostick, E., 등 (2012). Aurora kinase A-specific T-cell receptor gene transfer redirects T lymphocytes to display effective antileukemia reactivity. Blood 119, 368-376.

[0661] 79. Panina-Bordignon, P., Tan, A., Termijtelen, A., Demotz, S., Corradin, G., and Lanzavecchia, A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. Eur J Immunol 19, 2237-2242.

[0662] 80. Vitiello, A., Marchesini, D., Furze, J., Sherman, L.A., and Chesnut, R.W. (1991). Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. J Exp Med 173, 1007-1015.

[0663] 81. Yachi, P.P., Ampudia, J., Zal, T., and Gascoigne, N.R.J. (2006). Altered peptide ligands induce delayed CD8-T cell receptor interaction--a role for CD8 in distinguishing antigen quality. Immunity 25, 203-211.

[0664] 82. Pushko P, Parker M, Ludwig GV, Davis NL, Johnston RE, Smith JF. Replicon-helper systems from attenuated Venezuelan equine encephalitis virus: expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo. Virology. 1997 Dec 22;239(2):389-401.

[0665] 83. Strauss, JH and E G Strauss. The alphaviruses: gene expression, replication, and evolution. Microbiol Rev. 1994 Sep; 58(3): 491-562.

[0666] 84. Rheme C, Ehrengruber MU, Grandgirard D. Alphaviral cytotoxicity and its implication in vector development. Exp Physiol. 2005 Jan;90(1):45-52. Epub 2004 Nov 12.

[0667] 85. Riley, Michael K. II, and Wilfred Vermerris. Recent Advances in Nanomaterials for Gene Delivery—A Review. Nanomaterials 2017, 7(5), 94.

[0668] 86. Frolov I, Hardy R, Rice CM. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. RNA. 2001 Nov;7(11):1638-51.

[0669] 87. Jose J, Snyder JE, Kuhn RJ. A structural and functional perspective of alphavirus replication and assembly. Future Microbiol. 2009 Sep;4(7):837-56.

[0670] 88. Bo Li and C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a referenfe genome. BMC Bioinformatics, 12:323, August 2011

[0671] 89. Hillary Pearson, Tariq Daouda, Diana Paola Granados, Chantal Durette, Eric Bonneil, Mathieu Courcelles, Anja Rodenbrock, Jean-Philippe Laverdure, Caroline Côté, Sylvie Mader, Sébastien Lemieux, Pierre Thibault, and Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. The Journal of Clinical Investigation, 2016,

[0672] 90. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Kloetzel, Michael P. H. Stumpf, Albert J. R. Heck, Michele Mishto. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science, 21, October 2016.

[0673] 91. Mommen GP., Marino, F., Meiring HD., Poelen, MC., van Gaans-van den Brink, JA., Mohammed S., Heck AJ., and van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. Mol Cell Proteomics 15(4): 1412-1423, April 2016.

[0674] 92. Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Lower, Jan Diekmann, Sebastian Boegel, Barbara Schrors, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Ozlem Tureci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to caner. Nature 520, 692-696, April 2015.

[0675] 93. Tran E., Turcotte S., Gros A., Robbins P.F., Lu Y.C., Dudley M.E., Wunderlich J.R., Somerville R.P., Hogan K., Hinrichs C.S., Parkhurst M.R., Yang J.C., Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science 344(6184) 641-645, May 2014.

[0676] 94. Andreatta M., Karosiene E., Rasmussen M., Stryhn A., Buus S., Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics 67(11-12) 641-650, November 2015.

[0677] 95. Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 10:296, September 2009.

[0678] 96. Nielsen, M., Lundegaard, C., Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics 8:238, July 2007.

[0679] 97. Zhang, J., 등 PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Molecular & Cellular Proteomics. 11(4):1-8. 1/2/2012.

[0680] 98. Jensen, Kamilla Kjaergaard, 등 "Improved Methods for Prediting Peptide Binding Affinity to MHC Class II Molecules." Immunology, 2018, doi:10.1111/imm.12889.

[0681] 99. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., 등 (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413-421

[0682] 100. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., 등 (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell 171, 1259-1271.e11.

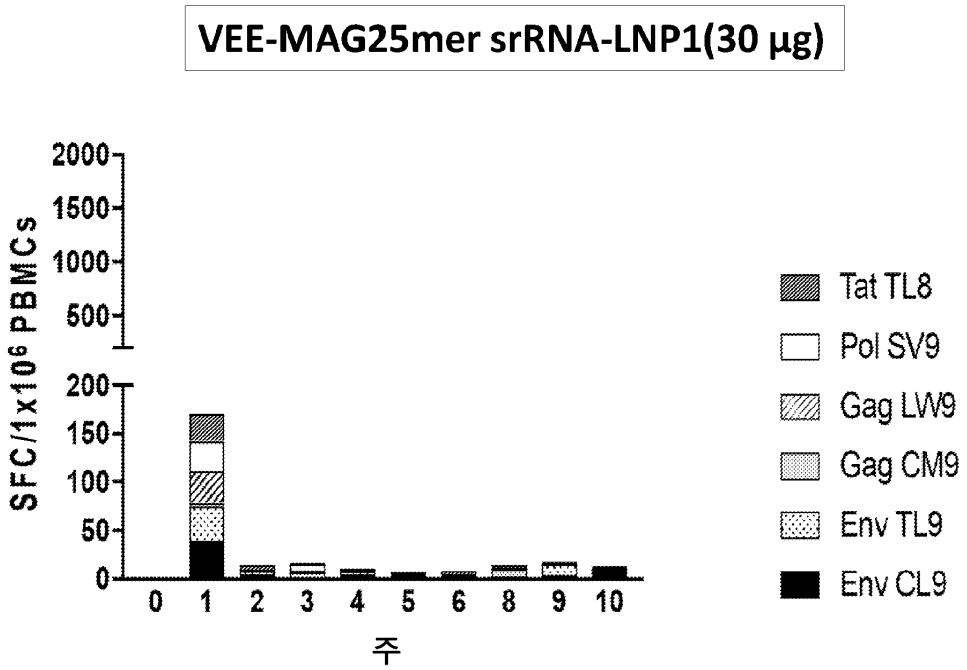
[0683] 101. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., 등 (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 33, 1152-1158.

[0684] 102. Van Loo, P., Nordgard, S.H., Lingjarde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., 등 (2010). Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U. S. A. 107, 16910-16915.

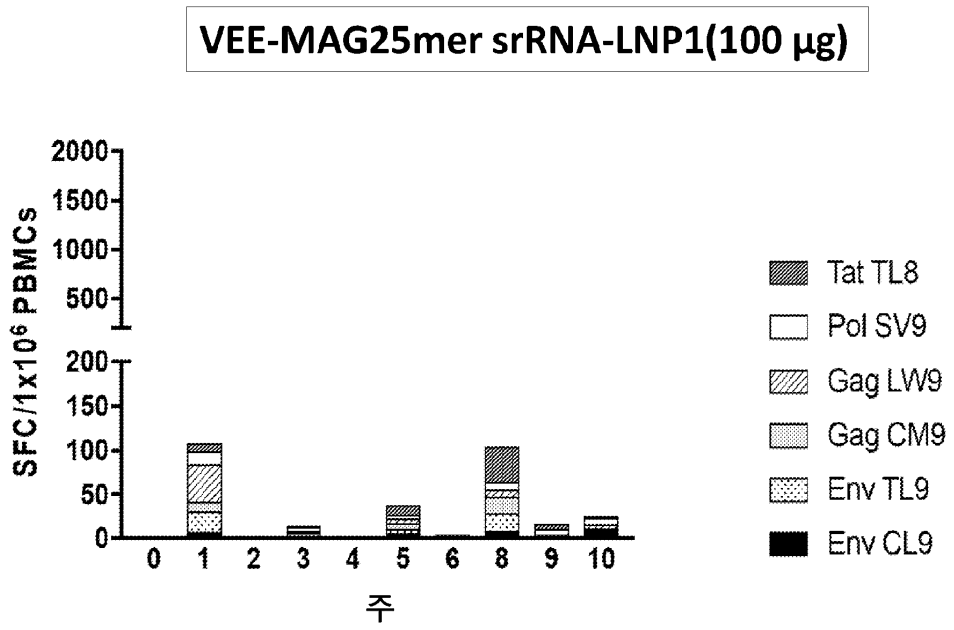
[0685] 103. Van Loo, P., Nordgard, S.H., Lingjarde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., 등 (2010). Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U. S. A. 107, 16910-16915.

도면

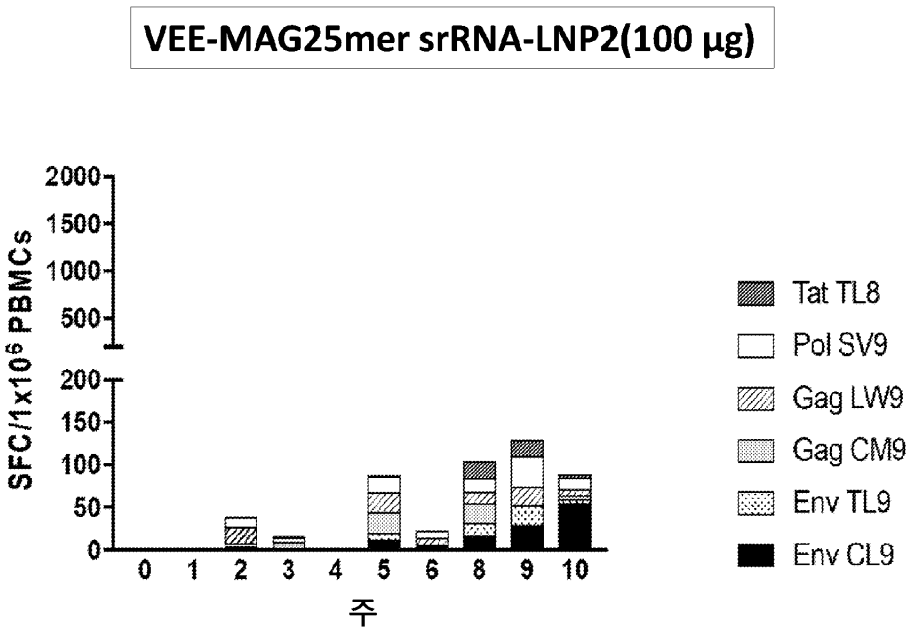
도면1a



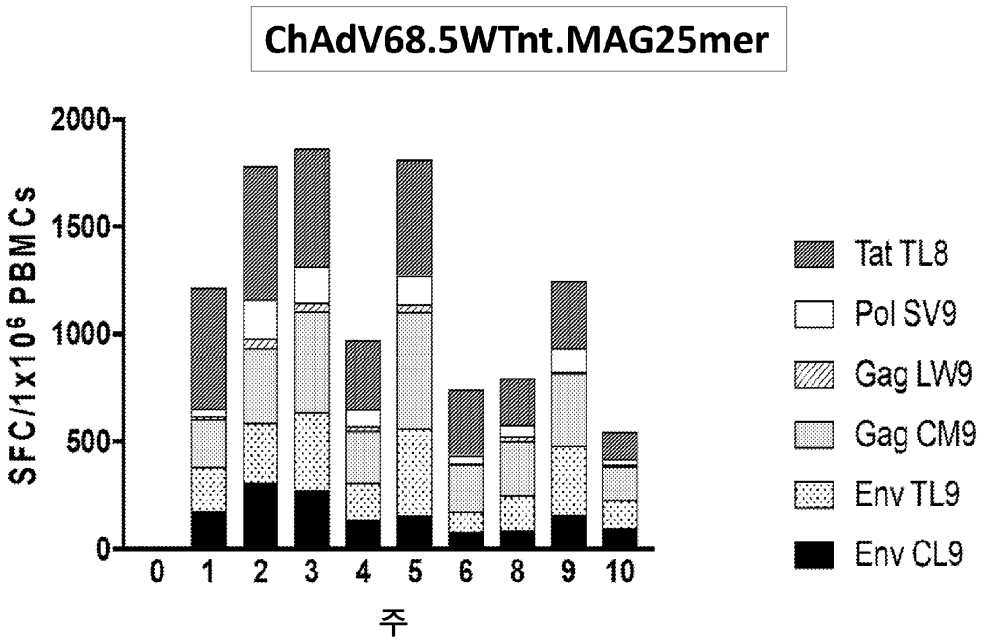
도면1b


VEE-MAG25mer srRNA-LNP1(100 μg)
SFC/1x10⁶ PBMCs
2000
1500
1000
500
200
150
100
50
0
0 1 2 3 4 5 6 8 9 10
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

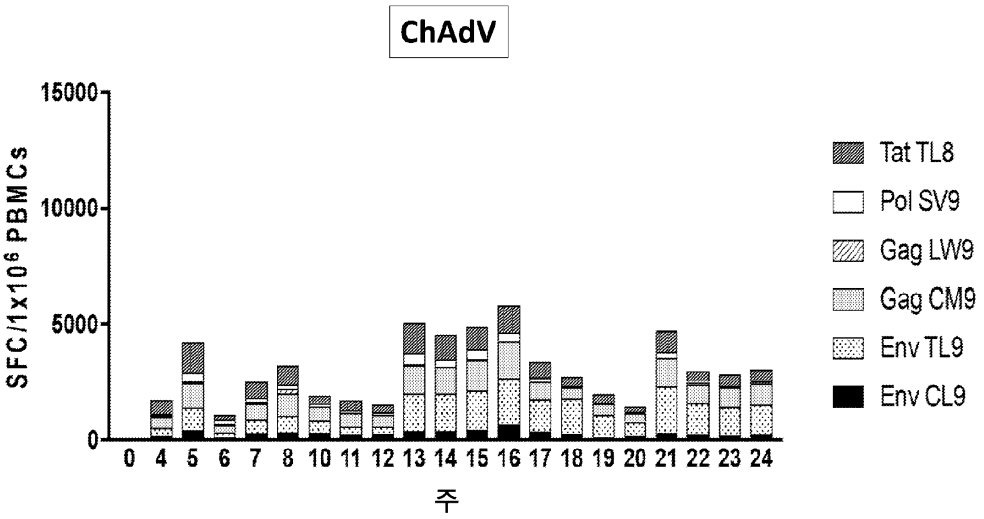
도면1c


VEE-MAG25mer srRNA-LNP2(100 μg)
SFC/1x10⁶ PBMCs
2000
1500
1000
500
200
150
100
50
0
0 1 2 3 4 5 6 8 9 10
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

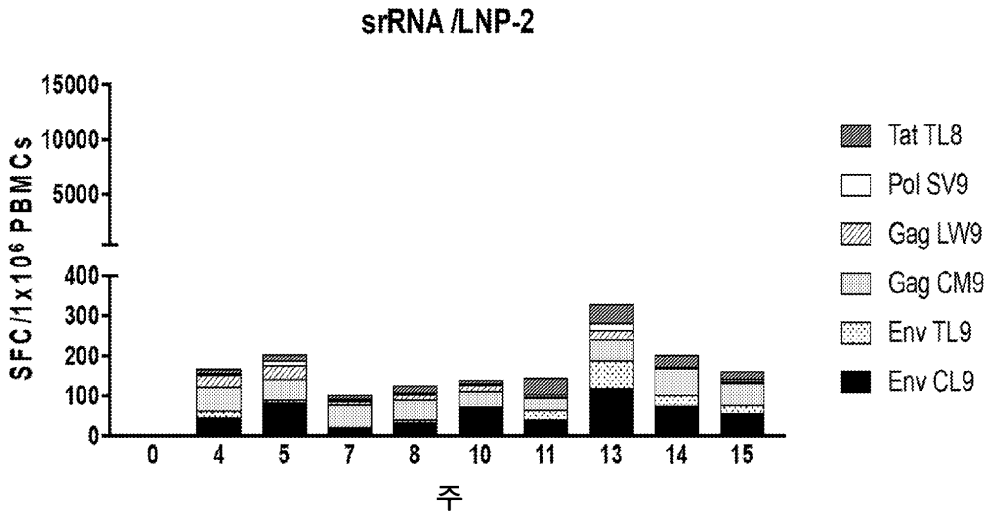
도면1d


ChAdV68.5WTnt.MAG25mer
SFC/1x10⁶ PBMCs
2000
1500
1000
500
0
0 1 2 3 4 5 6 8 9 10
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

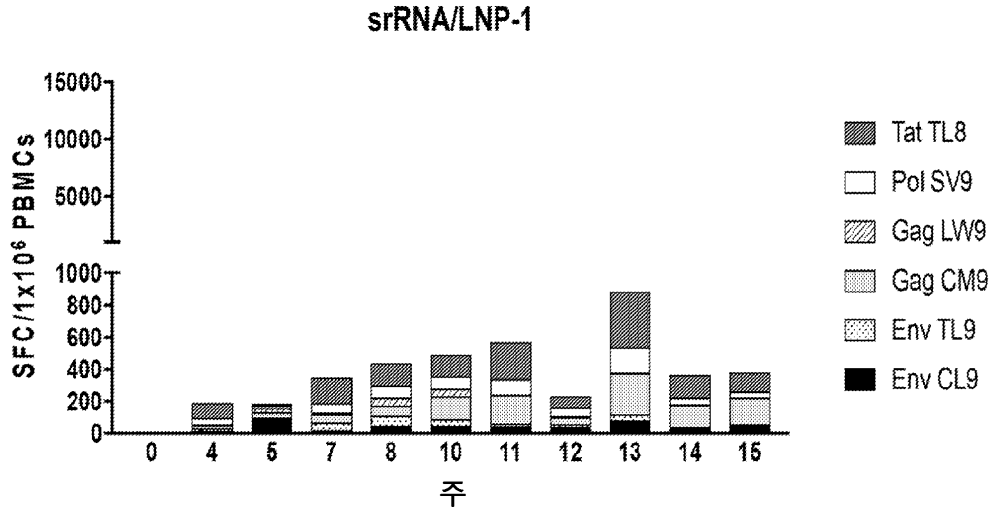
도면2


ChAdV
SFC/1x10⁶ PBMCs
15000
10000
5000
0
0 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

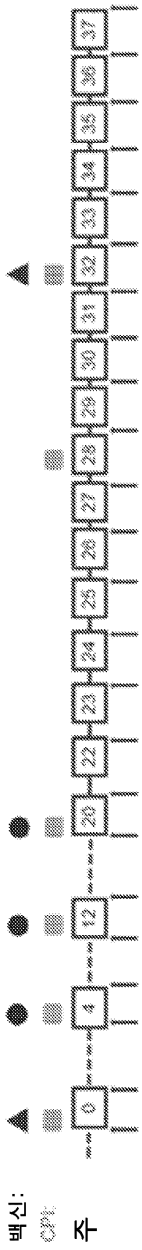
도면3


srRNA /LNP-2
SFC/1x10⁶ PBMCs
15000
10000
5000
400
300
200
100
0
0 4 5 7 8 10 11 13 14 15
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

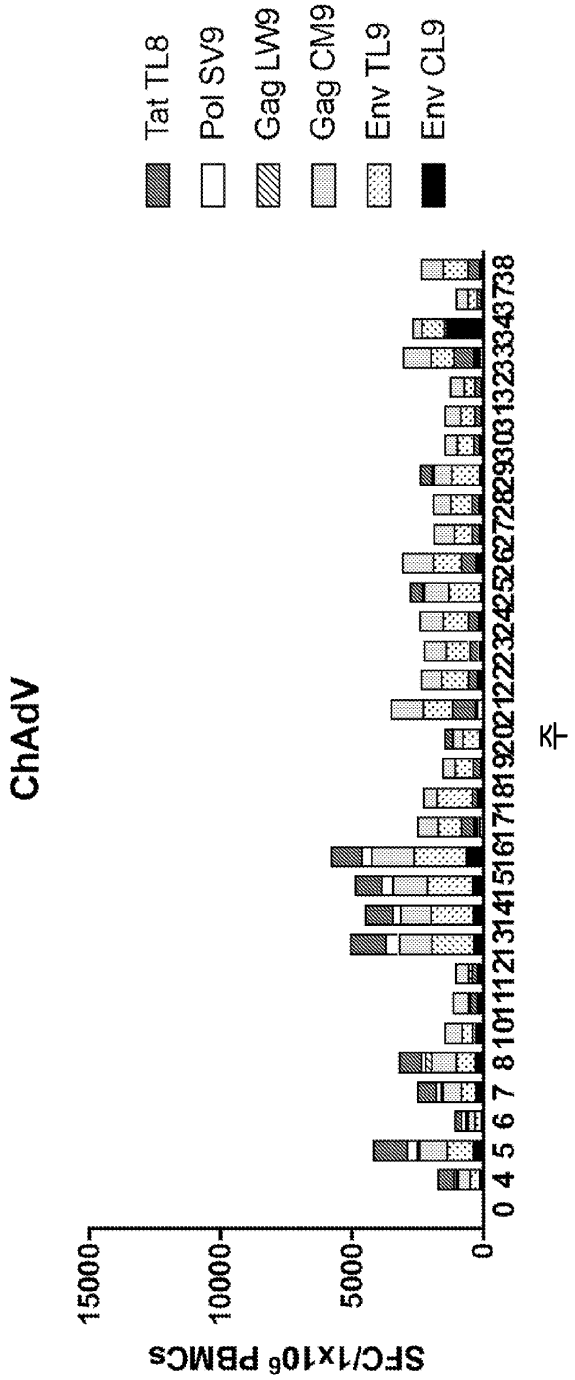
도면4


srRNA/LNP-1
SFC/1x10⁶ PBMCs
15000
10000
5000
1000
800
600
400
200
0
0 4 5 7 8 10 11 12 13 14 15
주
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9

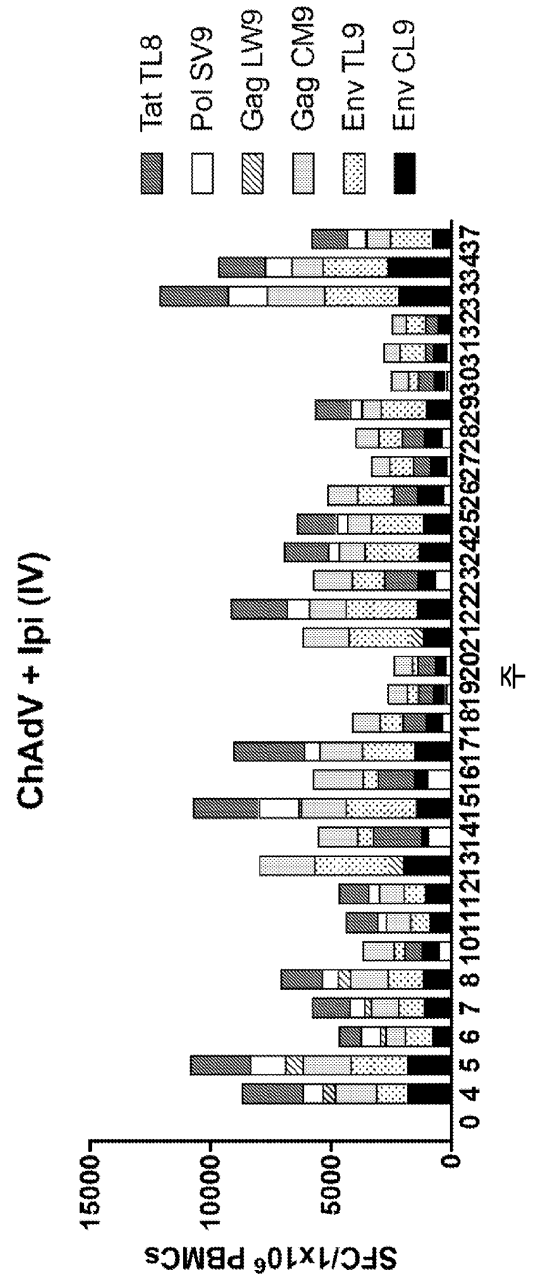
도면5



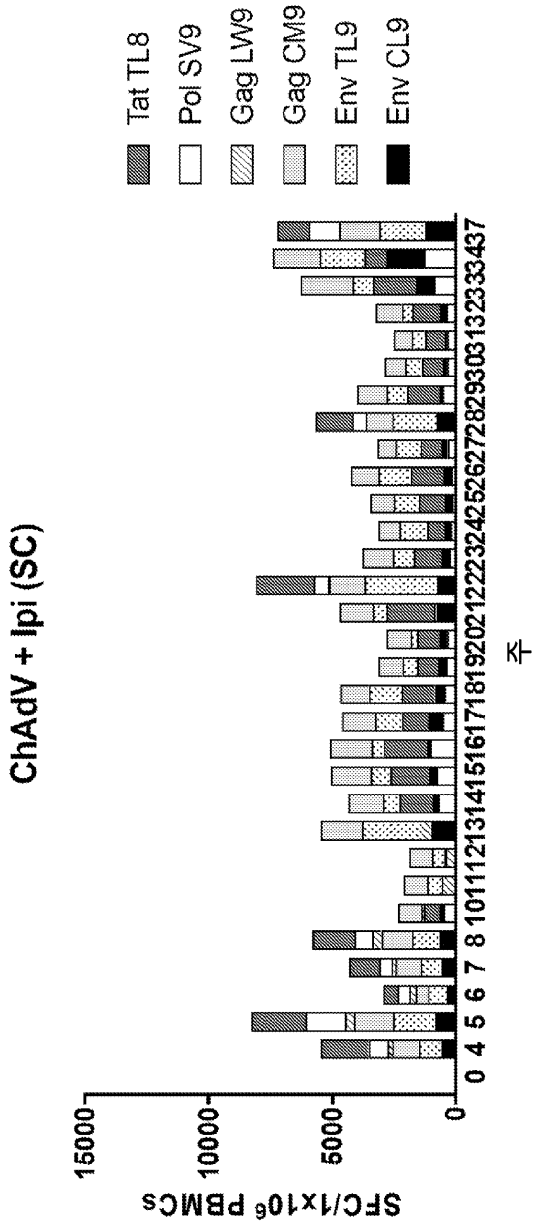
도면6


ChAdV
Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9
SFC/1x10⁶ PBMCs
15000
10000
5000
0
0 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 37 38
부

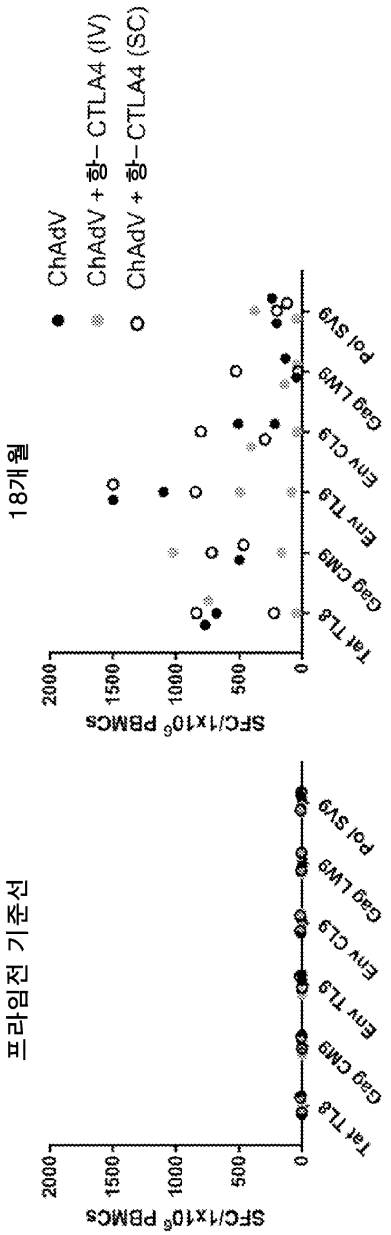
도면7


Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9
ChAdV + Ipi (IV)
SFC/1x10⁶ PBMCs
15000
10000
5000
0
0 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 37
주

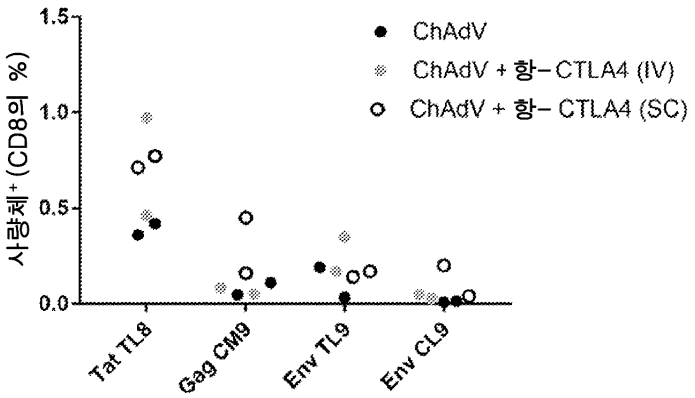
도면8


Tat TL8
Pol SV9
Gag LW9
Gag CM9
Env TL9
Env CL9
ChAdV + Ipi (SC)
SFC/1x10⁶ PBMCs
15000
10000
5000
0
0 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 37
주

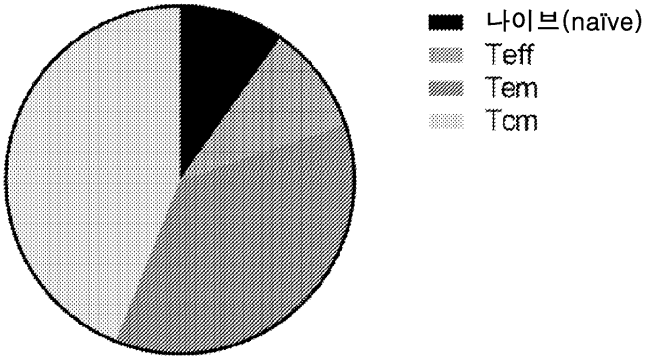
도면9


ChAdV
ChAdV + 항- CTLA4 (IV)
ChAdV + 항- CTLA4 (SC)
18개월
프라임전 기준선
SFC/1x10^6 PBMCs
2000
1500
1000
500
0
Tat TL8
Gag CM9
Env TL9
Env CL9
Gag LW9
Pol SV9

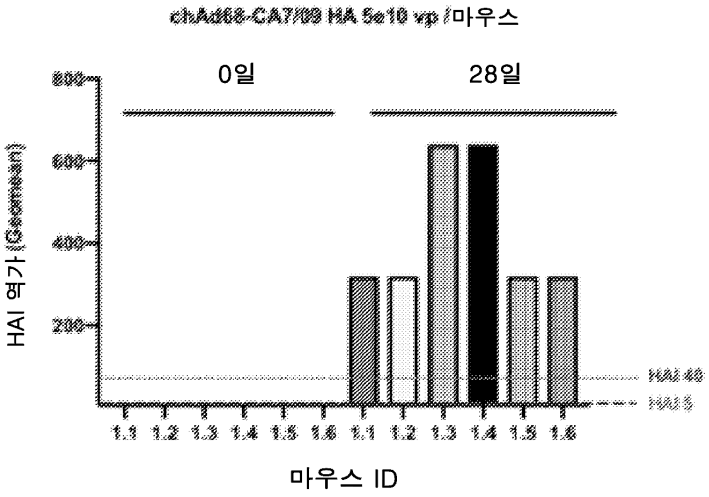
도면10


사량체+ (CD8의 %)
1.5
1.0
0.5
0.0
ChAdV
ChAdV + 항- CTLA4 (IV)
ChAdV + 항- CTLA4 (SC)
Tat TL8
Gag CM9
Env TL9
Env CL9

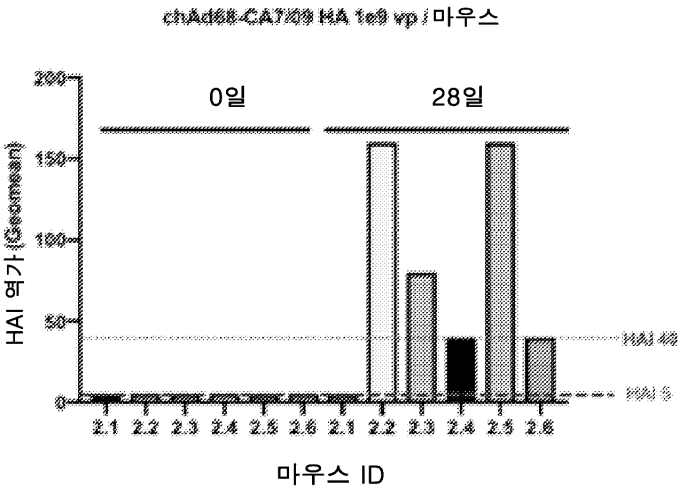
도면11



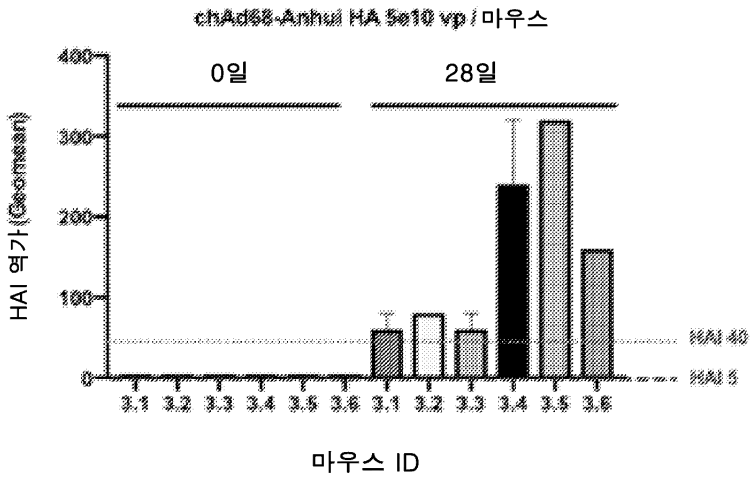
도면12a



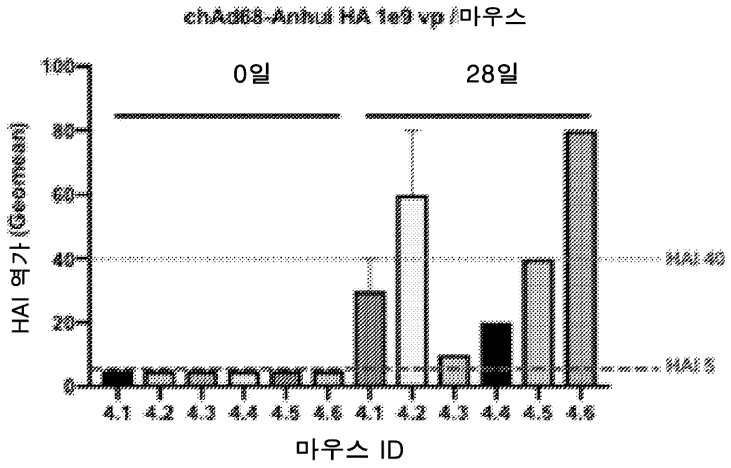
도면12b



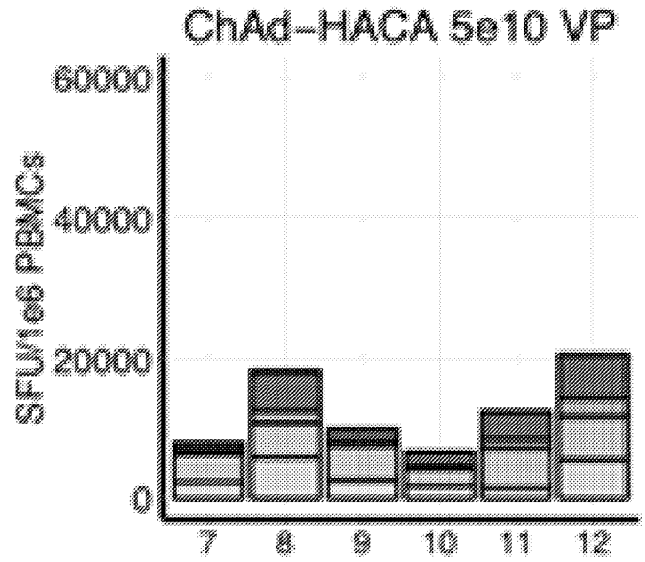
도면13a



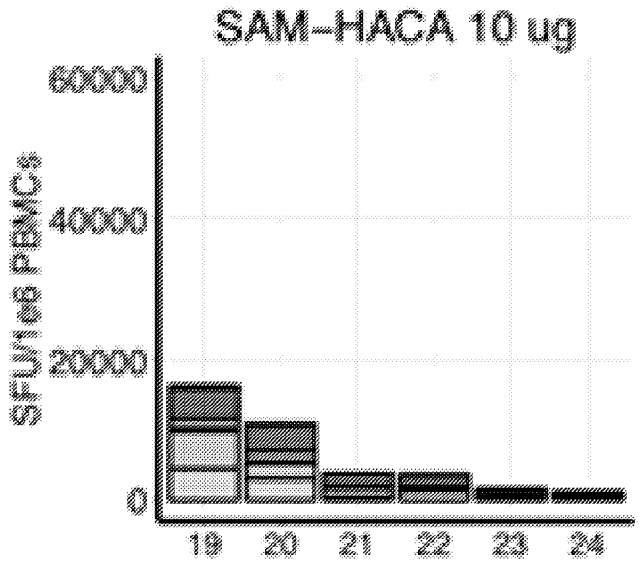
도면13b



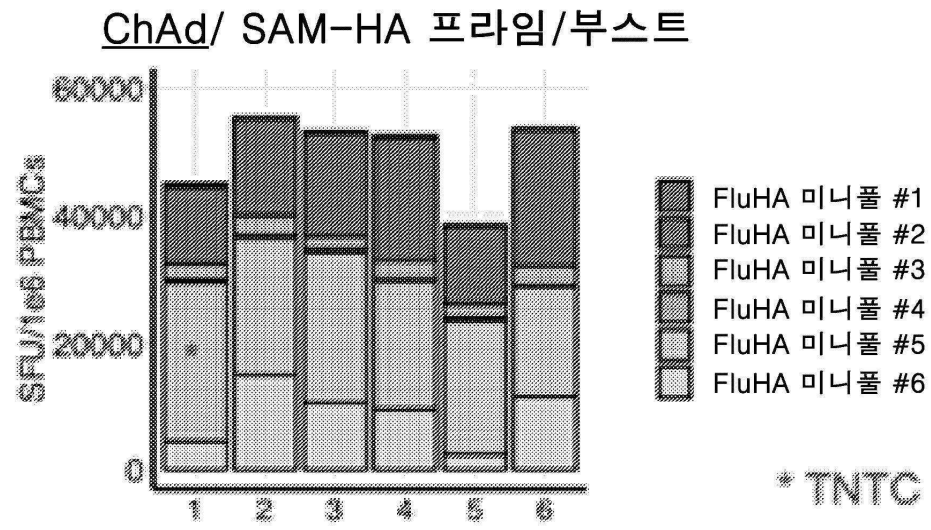
도면14a



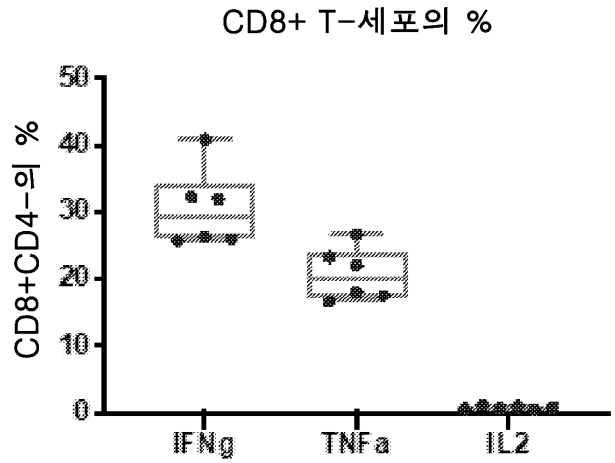
도면14b


SAM-HACA 10 ug
SFU/10^6 PBMCs
60000
40000
20000
0
19
20
21
22
23
24

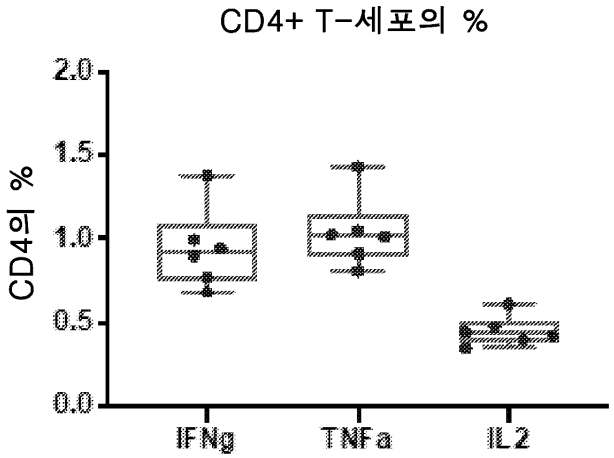
도면14c


ChAd/ SAM-HA 프라임/부스트
SFU/10^6 PBMCs
60000
40000
20000
0
1
2
3
4
5
6
FluHA 미니풀 #1
FluHA 미니풀 #2
FluHA 미니풀 #3
FluHA 미니풀 #4
FluHA 미니풀 #5
FluHA 미니풀 #6
* TNTC

도면15a



도면15b



서 열 목 록

SEQUENCE LISTING

<110> GRITSTONE ONCOLOGY, INC.

<120> INFECTIOUS DISEASE ANTIGENS AND VACCINES

<130> GSO-090WO

<140>

<141>

<150> 63/005,160

<151> 2020-04-03

<160> 56

<170> PatentIn version 3.5

<210> 1

<211> 36519

<212> DNA

<213> Artificial Sequence

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

<400> 1

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc 600

gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactggtggt aaatgccatg 660

atgggcgacg accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg 720

tatgatctgg aggtggatgt gcccgaggac gatcccaatg aggaggcggt aaatgatttt 780

tttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac 840

tcttcactgc ataccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg 900

gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag 960

caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg 1020

gactgcccgc ctctgcccgg acacggctgt aagtcttgtg aatttcatcg catgaatact 1080

ggagataaag ctgtgttgtg tgcactttgc tatatgagag cttacaacca ttgtgtttac 1140

agtaagtgtg attaagttga actttagagg gaggcagaga gcagggtgac tgggcgatga 1200

ctggtttatt tatgtatata tgttctttat ataggtcccg tctctgacgc agatgatgag 1260

acccccacta caaagtccac ttcgtcaccc ccagaaattg gcacatctcc acctgagaat 1320

attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat 1380

gacttgctac agggtggggt tgaacctttg gacttgtgta cccggaaacg ccccaggcac 1440

taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc 1500

aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt 1560

atataagcag gtgcagacct gtgtggttag ctcagagcgg catggagatt tggacggtct 1620

tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc 1680

tgtggagatt ctgcttcggt ggcgacctag ctaggctagt ctacagggcc aaacaggatt 1740

atagtgaaca atttgaggtt attttgagag agtgttctgg tctttttgac gctcttaact 1800

tgggccatca gtctcacttt aaccagagga tttcgagagc ccttgatttt actactcctg 1860

gcagaaccac tgcagcagta gccttttttg cttttattct tgacaaatgg agtcaagaaa 1920

cccatttcag cagggattac cagctggatt tcttagcagt agctttgtgg agaacatgga 1980

agtgccagcg cctgaatgca atctccggct acttgccggt acagccgcta gacactctga 2040

ggatcctgaa tctccaggag agtcccaggg cacgccaacg tcgccagcag cagcagcagg 2100

aggaggatca agaagagaac ccgagagccg gcctggaccc tccggcggag gaggaggagt 2160

agctgacctg tttcctgaac tgcgccgggt gctgactagg tcttcgagtg gtcgggagag 2220

ggggattaag cgggagaggc atgatgagac taatcacaga actgaactga ctgtgggtct 2280

gatgagtcgc aagcgcccag aaacagtgtg gtggcatgag gtgcagtcga ctggcacaga 2340

tgaggtgtcg gtgatgcatg agaggttttc tctagaacaa gtcaagactt gttggttaga 2400

gcctgaggat gattgggagg tagccatcag gaattatgcc aagctggctc tgaggccaga 2460

caagaagtac aagattacta agctgataaa tatcagaaat gcctgctaca tctcagggaa 2520

tggggctgaa gtggagatct gtctccagga aagggtggct ttcagatgct gcatgatgaa 2580

tatgtacccg ggagtggtgg gcatggatgg ggttaccttt atgaacatga ggttcagggg 2640

agatgggtat aatggcacgg tctttatggc caataccaag ctgacagtcc atggctgctc 2700

cttctttggg tttaataaca cctgcatcga ggcctggggt caggtcggtg tgaggggctg 2760

cagtttttca gccaactgga tgggggtcgt gggcaggacc aagagtatgc tgtccgtgaa 2820

gaaatgcttg tttgagaggt gccacctggg ggtgatgagc gagggcgaag ccagaatccg 2880

ccactgcgcc tctaccgaga cgggctgctt tgtgctgtgc aagggcaatg ctaagatcaa 2940

gcataatatg atctgtggag cctcggacga gcgcggctac cagatgctga cctgcgccgg 3000

cgggaacagc catatgctgg ccaccgtaca tgtggcttcc catgctcgca agccctggcc 3060

cgagttcgag cacaatgtca tgaccaggtg caatatgcat ctggggtccc gccgaggcat 3120

gttcatgccc taccagtgca acctgaatta tgtgaaggtg ctgctggagc ccgatgccat 3180

gtccagagtg agcctgacgg gggtgtttga catgaatgtg gaggtgtgga agattctgag 3240

atatgatgaa tccaagacca ggtgccgagc ctgcgagtgc ggagggaagc atgccaggtt 3300

ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg 3360

caccgggacg gagttcggtt ccagcgggga agaatctgac tagagtgagt agtgttctgg 3420

ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg 3480

cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg 3540

gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg tggacggccg 3600

gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcgtt 3660

ggacgcagct gccgccgcag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat 3720

gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atcccgccag 3780

cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct 3840

gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac 3900

ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttgtt gattttaaca 3960

cagagtctga atctttattt gatttttcgc gcgcggtagg ccctggacca ccggtctcga 4020

tcattgagca cccggtggat cttttccagg acccggtaga ggtgggcttg gatgttgagg 4080

tacatgggca tgagcccgtc ccgggggtgg aggtagctcc attgcagggc ctcgtgctcg 4140

ggggtggtgt tgtaaatcac ccagtcatag caggggcgca gggcatggtg ttgcacaata 4200

tctttgagga ggagactgat ggccacgggc agccctttgg tgtaggtgtt tacaaatctg 4260

ttgagctggg agggatgcat gcgggggggag atgaggtgca tcttggcctg gatcttgaga 4320

ttggcgatgt taccgcccag atcccgcctg gggttcatgt tgtgcaggac caccagcacg 4380

gtgtatccgg tgcacttggg gaatttatca tgcaacttgg aagggaaggc gtgaaagaat 4440

ttggcgacgc ctttgtgccc gcccaggttt tccatgcact catccatgat gatggcgatg 4500

ggcccgtggg cggcggcctg ggcaaagacg tttcgggggt cggacacatc atagttgtgg 4560

tcctgggtga ggtcatcata ggccatttta atgaatttgg ggcggagggt gccggactgg 4620

gggacaaagg taccctcgat cccgggggcg tagttcccct cacagatctg catctcccag 4680

gctttgagct cggaggggggg gatcatgtcc acctgcgggg cgataaagaa cacggtttcc 4740

ggggcggggg agatgagctg ggccgaaagc aagttccgga gcagctggga cttgccgcag 4800

ccggtggggc cgtagatgac cccgatgacc ggctgcaggt ggtagttgag ggagagacag 4860

ctgccgtcct cccggaggag gggggccacc tcgttcatca tctcgcgcac gtgcatgttc 4920

tcgcgcacca gttccgccag gaggcgctct ccccccaggg ataggagctc ctggagcgag 4980

gcgaagtttt tcagcggctt gagtccgtcg gccatgggca ttttggagag ggtttgttgc 5040

aagagttcca ggcggtccca gagctcggtg atgtgctcta cggcatctcg atccagcaga 5100

cctcctcgtt tcgcgggttg ggacggctgc gggagtaggg caccagacga tgggcgtcca 5160

gcgcagccag ggtccggtcc ttccagggtc gcagcgtccg cgtcagggtg gtctccgtca 5220
cggtgaaggg gtgcgcgccg ggctgggcgc ttgcgagggt gcgcttcagg ctcatccggc 5280
tggtcgaaaa ccgctcccga tcggcgccct gcgcgtcggc caggtagcaa ttgaccatga 5340
gttcgtagtt gagcgcctcg gccgcgtggc ctttggcgcg gagcttacct ttggaagtct 5400
gcccgcaggc gggacagagg agggacttga gggcgtagag cttgggggcg aggaagacgg 5460
actcgggggc gtaggcgtcc gcgccgcagt gggcgcagac ggtctcgcac tccacgagcc 5520
aggtgaggtc gggctggtcg gggtcaaaaa ccagtttccc gccgttcttt ttgatgcgtt 5580

tcttaccttt ggtctccatg agctcgtgtc cccgctgggt gacaaagagg ctgtccgtgt 5640
ccccgtagac cgactttatg ggccggtcct cgagcggtgt gccgcggtcc tcctcgtaga 5700
ggaaccccgc ccactccgag acgaaagccc gggtccaggc cagcacgaag gaggccacgt 5760
gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggta tgcaaacaca 5820
tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg 5880
gggtcccggc cgggggggta taaaagggtg cgggtccctg ctcgtcctca ctgtcttccg 5940
gatcgctgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga 6000

cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg 6060
cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt 6120
cgagcttggt ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca 6180
tggtctggtt tttttccttg tcggcgcgct ccttggcggc gatgttgagc tgcacgtact 6240
cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctcgtcgggc acgattctga 6300
cctgccagcc ccgattatgc agggtgatga ggtccacact ggtggccacc tcgccgcgca 6360
ggggctcatt agtccagcag aggcgtccgc ccttgcgcga gcagaagggg ggcaggggggt 6420

ccagcatgac ctcgtcgggg gggtcggcat cgatggtgaa gatgccgggc aggaggtcgg 6480
ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg 6540
ccagcgcgcg ctcgtaggga ctgaggggcg tgccccaggg catgggatgg gtaagcgcgg 6600
aggcgtacat gccgcagatg tcgtagacgt agaggggctc ctcgaggatg ccgatgtagg 6660
tggggtagca gcgccccccg cggatgctgg cgcgcacgta gtcatacagc tcgtgcgagg 6720
gggcgaggag ccccgggccc aggttggtgc gactgggctt ttcggcgcgg tagacgatct 6780
ggcggaaaat ggcatgcgag ttggaggaga tggtgggcct ttggaagatg ttgaagtggg 6840

cgtggggcag tccgaccgag tcgcggatga agtgggcgta ggagtcttgc agcttggcga 6900

cgagctcggc ggtgactagg acgtccagag cgcagtagtc gagggtctcc tggatgatgt 6960

catacttgag ctgtcccttt tgtttccaca gctcgcggtt gagaaggaac tcttcgcggt 7020

ccttccagta ctcttcgagg gggaacccgt cctgatctgc acggtaagag cctagcatgt 7080

agaactggtt gacggccttg taggcgcagc agcccttctc cacggggagg gcgtaggcct 7140

gggcggcctt gcgcagggag gtgtgcgtga gggcgaaagt gtccctgacc atgaccttga 7200

ggaactggtg cttgaagtcg atatcgtcgc agccccccctg ctcccagagc tggaagtccg 7260

tgcgcttctt gtaggcgggg ttgggcaaag cgaaagtaac atcgttgaag aggatcttgc 7320

ccgcgcgggg cataaagttg cgagtgatgc ggaaaggttg gggcacctcg gcccggttgt 7380

tgatgacctg ggcggcgagc acgatctcgt cgaagccgtt gatgttgtgg cccacgatgt 7440

agagttccac gaatcgcgga cggcccttga cgtggggcag tttcttgagc tcctcgtagg 7500

tgagctcgtc ggggtcgctg agcccgtgct gctcgagcgc ccagtcggcg agatgggggt 7560

tggcgcggag gaaggaagtc cagagatcca cggccagggc ggtttgcaga cggtcccggt 7620

actgacggaa ctgctgcccg acggccattt tttcgggggt gacgcagtag aaggtgcggg 7680

ggtccccgtg ccagcgatcc catttgagct ggagggcgag atcgagggcg agctcgacga 7740

gccggtcgtc cccggagagt ttcatgacca gcatgaaggg gacgagctgc ttgccgaagg 7800

accccatcca ggtgtaggtt tccacatcgt aggtgaggaa gagcctttcg gtgcgaggat 7860

gcgagccgat ggggaagaac tggatctcct gccaccaatt ggaggaatgg ctgttgatgt 7920

gatggaagta gaaatgccga cggcgcgccg aacactcgtg cttgtgttta tacaagcggc 7980

cacagtgctc gcaacgctgc acgggatgca cgtgctgcac gagctgtacc tgagttcctt 8040

tgacgaggaa tttcagtggg aagtggagtc gtggcgcctg catctcgtgc tgtactacgt 8100

cgtggtggtc ggcctggccc tcttctgcct cgatggtggt catgctgacg agcccgcgcg 8160

ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc 8220

cggagctgtc cagggtcctg agacgctgcg gagtcaggtc agtgggcagc ggcggcgcgc 8280

ggttgacttg caggagtttt tccagggcgc gcgggaggtc cagatggtac ttgatctcca 8340

ccgcgccatt ggtggcgacg tcgatggctt gcagggtccc gtgcccctgg ggtgtgacca 8400

ccgtcccccg tttcttcttg ggcggctggg gcgacggggg cggtgcctct tccatggtta 8460

gaagcggcgg cgaggacgcg cgccgggcgg caggggcggc tcggggcccg gaggcagggg 8520

cggcaggggc acgtcggcgc cgcgcgcggg taggttctgg tactgcgccc ggagaagact 8580

ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgcctctggg tgaaggccac 8640

gggacccgtg agtttgaacc tgaaagagag ttcgacagaa tcaatctcgg tatcgttgac 8700

ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcggt 8760

catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc 8820

cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt 8880

ccagacgcgg ctgtagacca cgacgccctc gggatcgcgg gcgcgcatga ccacctgggc 8940

gaggttgagc tccacgtggc gcgtgaagac cgcgtagttg cagaggcgct ggtagaggta 9000

gttgagcgtg gtggcgatgt gctcggtgac gaagaaatac atgatccagc ggcggagcgg 9060

catctcgctg acgtcgccca gcgcctccaa acgttccatg gcctcgtaaa agtccacggc 9120

gaagttgaaa aactgggagt tgcgcgccga gacggtcaac tcctcctcca gaagacggat 9180

gagctcggcg atggtggcgc gcacctcgcg ctcgaaggcc cccgggagtt cctccacttc 9240

ctcttcttcc tcctccacta acatctcttc tacttcctcc tcaggcggca gtggtggcgg 9300

gggaggggggc ctgcgtcgcc ggcggcgcac gggcagacgg tcgatgaagc gctcgatggt 9360

ctcgccgcgc cggcgtcgca tggtctcggt gacggcgcgc ccgtcctcgc ggggccgcag 9420

cgtgaagacg ccgccgcgca tctccaggtg gccgggggggg tccccgttgg gcagggagag 9480

ggcgctgacg atgcatctta tcaattgccc cgtagggact ccgcgcaagg acctgagcgt 9540

ctcgagatcc acgggatctg aaaaccgctg aacgaaggct tcgagccagt cgcagtcgca 9600

aggtaggctg agcacggttt cttctggcgg gtcatgttgg ttgggagcgg ggcgggcgat 9660

gctgctggtg atgaagttga ataggcggt tctgagacgg cggatggtgg cgaggagcac 9720

caggtctttg ggcccggctt gctggatgcg cagacggtcg gccatgcccc aggcgtggtc 9780

ctgacacctg gccaggtcct tgtagtagtc ctgcatgagc cgctccacgg gcacctcctc 9840

ctcgcccgcg cggccgtgca tgcgcgtgag cccgaagccg cgctggggct ggacgagcgc 9900

caggtcggcg acgacgcgct cggcgaggat ggcttgctgg atctgggtga gggtggtctg 9960

gaagtcatca aagtcgacga agcggtggta ggctccggtg ttgatggtgt aggagcagtt 10020

ggccatgacg gaccagttga cggtctggtg gcccggacgc acgagctcgt ggtacttgag 10080

gcgcgagtag gcgcgcgtgt cgaagatgta gtcgttgcag gtgcgcacca ggtactggta 10140

gccgatgagg aagtgcggcg gcggctggcg gtagagcggc catcgctcgg tggcgggggc 10200

gccgggcgcg aggtcctcga gcatggtgcg gtggtagccg tagatgtacc tggacatcca 10260

ggtgatgccg gcggcggtgg tggaggcgcg cgggaactcg cggacgcggt tccagatgtt 10320

gcgcagcggc aggaagtagt tcatggtggg cacggtctgg cccgtgaggc gcgcgcagtc 10380

gtggatgctc tatacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtattg gcactcccgt ctcgacccaa gcctgcacca accctccagg 10560

atacggaggc gggtcgtttt gcaacttttt tttggaggcc ggatgagact agtaagcgcg 10620

gaaagcggcc gaccgcgatg gctcgctgcc gtagtctgga gaagaatcgc cagggttgcg 10680
ttgcggtgtg ccccggttcg aggccggccg gattccgcgg ctaacgaggg cgtggctgcc 10740
ccgtcgtttc caagacccca tagccagccg acttctccag ttacggagcg agcccctctt 10800
ttgttttgtt tgtttttgcc agatgcatcc cgtactgcgg cagatgcgcc cccaccaccc 10860
tccaccgcaa caacagcccc ctccacagcc ggcgcttctg cccccgcccc agcagcaact 10920
tccagccacg accgccgcgg ccgccgtgag cggggctgga cagagttatg atcaccagct 10980
ggccttggaa gagggcgagg ggctggcgcg cctgggggcg tcgtcgccgg agcggcaccc 11040

gcgcgtgcag atgaaaaggg acgctcgcga ggcctacgtg cccaagcaga acctgttcag 11100
agacaggagc ggcgaggagc ccgaggagat gcgcgcggcc cggttccacg cggggcggga 11160
gctgcggcgc ggcctggacc gaaagagggt gctgagggac gaggatttcg aggcggacga 11220
gctgacgggg atcagccccg cgcgcgcgca cgtggccgcg gccaacctgg tcacggcgta 11280
cgagcagacc gtgaaggagg agagcaactt ccaaaaatcc ttcaacaacc acgtgcgcac 11340
cctgatcgcg cgcgaggagg tgaccctggg cctgatgcac ctgtgggacc tgctggaggc 11400
catcgtgcag aaccccacca gcaagccgct gacggcgcag ctgttcctgg tggtgcagca 11460

tagtcgggac aacgaagcgt tcagggaggc gctgctgaat atcaccgagc ccgagggccg 11520
ctggctcctg gacctggtga acattctgca gagcatcgtg gtgcaggagc gcgggctgcc 11580
gctgtccgag aagctggcgg ccatcaactt ctcggtgctg agtttgggca agtactacgc 11640
taggaagatc tacaagaccc cgtacgtgcc catagacaag gaggtgaaga tcgacgggtt 11700
ttacatgcgc atgaccctga aagtgctgac cctgagcgac gatctggggg tgtaccgcaa 11760
cgacaggatg caccgtgcgg tgagcgccag caggcggcgc gagctgagcg accaggagct 11820
gatgcatagt ctgcagcggg ccctgaccgg ggccgggacc gaggggggaga gctactttga 11880

catgggcgcg gacctgcact ggcagcccag ccgccgggcc ttggaggcgg cggcaggacc 11940
ctacgtagaa gaggtggacg atgaggtgga cgaggagggc gagtacctgg aagactgatg 12000
gcgcgaccgt atttttgcta gatgcaacaa caacagccac ctcctgatcc cgcgatgcgg 12060
gcggcgctgc agagccagcc gtccggcatt aactcctcgg acgattggac ccaggccatg 12120
caacgcatca tggcgctgac gacccgcaac cccgaagcct ttagacagca gccccaggcc 12180
aaccggctct cggccatcct ggaggccgtg gtgccctcgc gctccaaccc cacgcacgag 12240
aaggtcctgg ccatcgtgaa cgcgctggtg gagaacaagg ccatccgcgg cgacgaggcc 12300

ggcctggtgt acaacgcgct gctggagcgc gtggcccgct acaacagcac caacgtgcag 12360

accaacctgg accgcatggt gaccgacgtg cgcgaggccg tggcccagcg cgagcggttc 12420

caccgcgagt ccaacctggg atccatggtg gcgctgaacg ccttcctcag cacccagccc 12480

gccaacgtgc cccggggcca ggaggactac accaacttca tcagcgccct gcgcctgatg 12540

gtgaccgagg tgccccagag cgaggtgtac cagtccgggc cggactactt cttccagacc 12600

agtcgccagg gcttgcagac cgtgaacctg agccaggctt tcaagaactt gcagggcctg 12660

tggggcgtgc aggccccggt cggggaccgc gcgacggtgt cgagcctgct gacgccgaac 12720

tcgcgcctgc tgctgctgct ggtggcccccc ttcacggaca gcggcagcat caaccgcaac 12780

tcgtacctgg gctacctgat taacctgtac cgcgaggcca tcggccaggc gcacgtggac 12840

gagcagacct accaggagat cacccacgtg agccgcgccc tgggccagga cgacccgggc 12900

aacctggaag ccaccctgaa ctttttgctg accaaccggt cgcagaagat cccgccccag 12960

tacgcgctca gcaccgagga ggagcgcatc ctgcgttacg tgcagcagag cgtgggcctg 13020

ttcctgatgc aggagggggc cacccccagc gccgcgctcg acatgaccgc gcgcaacatg 13080

gagcccagca tgtacgccag caaccgcccg ttcatcaata aactgatgga ctacttgcat 13140

cgggcggccg ccatgaactc tgactatttc accaacgcca tcctgaatcc ccactggctc 13200

ccgccgccgg ggttctacac gggcgagtac gacatgcccg accccaatga cgggttcctg 13260

tgggacgatg tggacagcag cgtgttctcc ccccgaccgg gtgctaacga gcgccccttg 13320

tggaagaagg aaggcagcga ccgacgcccg tcctcggcgc tgtccggccg cgagggtgct 13380

gccgcggcgg tgcccgaggc cgccagtcct ttcccgagct tgcccttctc gctgaacagt 13440

atccgcagca gcgagctggg caggatcacg cgcccgcgct tgctgggcga agaggagtac 13500

ttgaatgact cgctgttgag acccgagcgg gagaagaact tccccaataa cgggatagaa 13560

agcctggtgg acaagatgag ccgctggaag acgtatgcgc aggagcacag ggacgatccc 13620

cgggcgtcgc agggggccac gagccggggc agcgccgccc gtaaacgccg gtggcacgac 13680

aggcagcggg gacagatgtg ggacgatgag gactccgccg acgacagcag cgtgttggac 13740

ttgggtggga gtggtaaccc gttcgctcac ctgcgccccc gtatcgggcg catgatgtaa 13800

gagaaaccga aaataaatga tactcaccaa ggccatggcg accagcgtgc gttcgtttct 13860

tctctgttgt tgttgtatct agtatgatga ggcgtgcgta cccggagggt cctcctccct 13920

cgtacgagag cgtgatgcag caggcgatgg cggcggcggc gatgcagccc ccgctggagg 13980

ctccttacgt gcccccgcgg tacctggcgc ctacggaggg gcggaacagc attcgttact 14040

cggagctggc accttgtac gataccaccc ggttgtacct ggtggacaac aagtcggcgg 14100

acatcgcctc gctgaactac cagaacgacc acagcaactt cctgaccacc gtggtgcaga 14160

acaatgactt cacccccacg gaggccagca cccagaccat caactttgac gagcgctcgc 14220

ggtggggcgg ccagctgaaa accatcatgc acaccaacat gcccaacgtg aacgagttca 14280

tgtacagcaa caagttcaag gcgcgggtga tggtctcccg caagaccccc aatggggtga 14340

cagtgacaga ggattatgat ggtagtcagg atgagctgaa gtatgaatgg gtggaatttg 14400

agctgcccga aggcaacttc tcggtgacca tgaccatcga cctgatgaac aacgccatca 14460

tcgacaatta cttggcggtg gggcggcaga acggggtgct ggagagcgac atcggcgtga 14520

agttcgacac taggaacttc aggctgggct gggaccccgt gaccgagctg gtcatgcccg 14580

gggtgtacac caacgaggct ttccatcccg atattgtctt gctgcccggc tgcggggtgg 14640

acttcaccga gagccgcctc agcaacctgc tgggcattcg caagaggcag cccttccagg 14700

aaggcttcca gatcatgtac gaggatctgg agggggggcaa catccccgcg ctcctggatg 14760

tcgacgccta tgagaaaagc aaggaggatg cagcagctga agcaactgca gccgtagcta 14820

ccgcctctac cgaggtcagg ggcgataatt ttgcaagcgc cgcagcagtg gcagcggccg 14880

aggcggctga aaccgaaagt aagatagtca ttcagccggt ggagaaggat agcaagaaca 14940

ggagctacaa cgtactaccg gacaagataa acaccgccta ccgcagctgg tacctagcct 15000

acaactatgg cgaccccgag aagggcgtgc gctcctggac gctgctcacc acctcggacg 15060

tcacctgcgg cgtggagcaa gtctactggt cgctgcccga catgatgcaa gacccggtca 15120

ccttccgctc cacgcgtcaa gttagcaact acccggtggt gggcgccgag ctcctgcccg 15180

tctactccaa gagcttcttc aacgagcagg ccgtctactc gcagcagctg cgcgccttca 15240

cctcgcttac gcacgtcttc aaccgcttcc ccgagaacca gatcctcgtc cgcccgcccg 15300

cgcccaccat taccaccgtc agtgaaaacg ttcctgctct cacagatcac gggaccctgc 15360

cgctgcgcag cagtatccgg ggagtccagc gcgtgaccgt tactgacgcc agacgccgca 15420

cctgccccta cgtctacaag gccctgggca tagtcgcgcc gcgcgtcctc tcgagccgca 15480

ccttctaaat gtccattctc atctcgccca gtaataacac cggttggggc ctgcgcgcgc 15540

ccagcaagat gtacggaggc gctcgccaac gctccacgca acaccccgtg cgcgtgcgcg 15600

ggcacttccg cgctccctgg ggcgccctca agggccgcgt gcggtcgcgc accaccgtcg 15660

acgacgtgat cgaccaggtg gtggccgacg cgcgcaacta cacccccgcc gccgcgcccg 15720

tctccaccgt ggacgccgtc atcgacagcg tggtggccga cgcgcgccgg tacgcccgcg 15780

ccaagagccg gcggcggcgc atcgcccggc ggcaccggag cacccccgcc atgcgcgcgg 15840

cgcgagcctt gctgcgcagg gccaggcgca cgggacgcag ggccatgctc agggcggcca 15900

gacgcgcggc ttcaggcgcc agcgccggca ggacccggag acgcgcggcc acggcggcgg 15960

cagcggccat cgccagcatg tcccgcccgc ggcgagggaa cgtgtactgg gtgcgcgacg 16020

ccgccaccgg tgtgcgcgtg cccgtgcgca cccgcccccc tcgcacttga agatgttcac 16080

ttcgcgatgt tgatgtgtcc cagcggcgag gaggatgtcc aagcgcaaat tcaaggaaga 16140
gatgctccag gtcatcgcgc ctgagatcta cggccctgcg gtggtgaagg aggaaagaaa 16200
gccccgcaaa atcaagcggg tcaaaaagga caaaaaggaa gaagaaagtg atgtggacgg 16260
attggtggag tttgtgcgcg agttcgcccc ccggcggcgc gtgcagtggc gcgggcggaa 16320
ggtgcaaccg gtgctgagac ccggcaccac cgtggtcttc acgcccggcg agcgctccgg 16380
caccgcttcc aagcgctcct acgacgaggt gtacggggat gatgatattc tggagcaggc 16440
ggccgagcgc ctgggcgagt ttgcttacgg caagcgcagc cgttccgcac cgaaggaaga 16500

ggcggtgtcc atcccgctgg accacggcaa ccccacgccg agcctcaagc ccgtgacctt 16560
gcagcaggtg ctgccgaccg cggcgccgcg ccgggggttc aagcgcgagg gcgaggatct 16620
gtaccccacc atgcagctga tggtgcccaa gcgccagaag ctggaagacg tgctggagac 16680
catgaaggtg gacccggacg tgcagcccga ggtcaaggtg cggcccatca agcaggtggc 16740
cccgggcctg ggcgtgcaga ccgtggacat caagattccc acggagccca tggaaacgca 16800
gaccgagccc atgatcaagc ccagcaccag caccatggag gtgcagacgg atccctggat 16860
gccatcggct cctagtcgaa gaccccggcg caagtacggc gcggccagcc tgctgatgcc 16920

caactacgcg ctgcatcctt ccatcatccc cacgccgggc taccgcggca cgcgcttcta 16980
ccgcggtcat accagcagcc gccgccgcaa gaccaccact cgccgccgcc gtcgccgcac 17040
cgccgctgca accacccctg ccgccctggt gcggagagtg taccgccgcg gccgcgcacc 17100
tctgaccctg ccgcgcgcgc gctaccaccc gagcatcgcc atttaaactt tcgcctgctt 17160
tgcagatcaa tggccctcac atgccgcctt cgcgttccca ttacgggcta ccgaggaaga 17220
aaaccgcgcc gtagaaggct ggcggggaac gggatgcgtc gccaccacca ccggcggcgg 17280
cgcgccatca gcaagcggtt gggggggaggc ttcctgcccg cgctgatccc catcatcgcc 17340

gcggcgatcg gggcgatccc cggcattgct tccgtggcgg tgcaggcctc tcagcgccac 17400
tgagacacac ttggaaacat cttgtaataa accaatggac tctgacgctc ctggtcctgt 17460
gatgtgtttt cgtagacaga tggaagacat caatttttcg tccctggctc cgcgacacgg 17520
cacgcggccg ttcatgggca cctggagcga catcggcacc agccaactga acgggggcgc 17580
cttcaattgg agcagtctct ggagcgggct taagaatttc gggtccacgc ttaaaaccta 17640
tggcagcaag gcgtggaaca gcaccacagg gcaggcgctg agggataagc tgaaagagca 17700
gaacttccag cagaaggtgg tcgatgggct cgcctcgggc atcaacgggg tggtggacct 17760

ggccaaccag gccgtgcagc ggcagatcaa cagccgcctg gacccggtgc cgcccgccgg 17820

ctccgtggag atgccgcagg tggaggagga gctgcctccc ctggacaagc ggggcgagaa 17880

gcgaccccgc cccgatgcgg aggagacgct gctgacgcac acggacgagc cgcccccgta 17940

cgaggaggcg gtgaaactgg gtctgcccac cacgcggccc atcgcgcccc tggccaccgg 18000

ggtgctgaaa cccgaaaagc ccgcgaccct ggacttgcct cctccccagc cttcccgccc 18060

ctctacagtg gctaagcccc tgccgccggt ggccgtggcc cgcgcgcgac ccgggggcac 18120

cgcccgccct catgcgaact ggcagagcac tctgaacagc atcgtgggtc tgggagtgca 18180

gagtgtgaag cgccgccgct gctattaaac ctaccgtagc gcttaacttg cttgtctgtg 18240

tgtgtatgta ttatgtcgcc gccgccgctg tccaccagaa ggaggagtga agaggcgcgt 18300

cgccgagttg caagatggcc accccatcga tgctgcccca gtgggcgtac atgcacatcg 18360

ccggacagga cgcttcggag tacctgagtc cgggtctggt gcagtttgcc cgcgccacag 18420

acacctactt cagtctgggg aacaagttta ggaaccccac ggtggcgccc acgcacgatg 18480

tgaccaccga ccgcagccag cggctgacgc tgcgcttcgt gcccgtggac cgcgaggaca 18540

acacctactc gtacaaagtg cgctacacgc tggccgtggg cgacaaccgc gtgctggaca 18600

tggccagcac ctactttgac atccgcggcg tgctggatcg gggccctagc ttcaaaccct 18660

actccggcac cgcctacaac agtctggccc ccaagggagc acccaacact tgtcagtgga 18720

catataaagc cgatggtgaa actgccacag aaaaaaccta tacatatgga aatgcacccg 18780

tgcagggcat taacatcaca aaagatggta ttcaacttgg aactgacacc gatgatcagc 18840

caatctacgc agataaaacc tatcagcctg aacctcaagt gggtgatgct gaatggcatg 18900

acatcactgg tactgatgaa aagtatggag gcagagctct taagcctgat accaaaatga 18960

agccttgtta tggttctttt gccaagccta ctaataaaga aggaggtcag gcaaatgtga 19020

aaacaggaac aggcactact aaagaatatg acatagacat ggctttcttt gacaacagaa 19080

gtgcggctgc tgctggccta gctccagaaa ttgttttgta tactgaaaat gtggatttgg 19140

aaactccaga tacccatatt gtatacaaag caggcacaga tgacagcagc tcttctatta 19200

atttgggtca gcaagccatg cccaacagac ctaactacat tggtttcaga gacaacttta 19260

tcgggctcat gtactacaac agcactggca atatgggggt gctggccggt caggcttctc 19320

agctgaatgc tgtggttgac ttgcaagaca gaaacaccga gctgtcctac cagctcttgc 19380

ttgactctct gggtgacaga acccggtatt tcagtatgtg gaatcaggcg gtggacagct 19440

atgatcctga tgtgcgcatt attgaaaatc atggtgtgga ggatgaactt cccaactatt 19500

gtttccctct ggatgctgtt ggcagaacag atacttatca gggaattaag gctaatggaa 19560

ctgatcaaac cacatggacc aaagatgaca gtgtcaatga tgctaatgag ataggcaagg 19620

gtaatccatt cgccatggaa atcaacatcc aagccaacct gtggaggaac ttcctctacg 19680

ccaacgtggc cctgtacctg cccgactctt acaagtacac gccggccaat gttaccctgc 19740

ccaccaacac caacacctac gattacatga acggccgggt ggtggcgccc tcgctggtgg 19800

actcctacat caacatcggg gcgcgctggt cgctggatcc catggacaac gtgaacccct 19860

tcaaccacca ccgcaatgcg gggctgcgct accgctccat gctcctgggc aacgggcgct 19920

acgtgccctt ccacatccag gtgccccaga aatttttcgc catcaagagc ctcctgctcc 19980

tgcccgggtc ctacacctac gagtggaact tccgcaagga cgtcaacatg atcctgcaga 20040

gctccctcgg caacgacctg cgcacggacg gggcctccat ctccttcacc agcatcaacc 20100

tctacgccac cttcttcccc atggcgcaca acacggcctc cacgctcgag gccatgctgc 20160

gcaacgacac caacgaccag tccttcaacg actacctctc ggcggccaac atgctctacc 20220

ccatcccggc caacgccacc aacgtgccca tctccatccc ctcgcgcaac tgggccgcct 20280

tccgcggctg gtccttcacg cgtctcaaga ccaaggagac gccctcgctg ggctccgggt 20340

tcgaccccta cttcgtctac tcgggctcca tcccctacct cgacggcacc ttctacctca 20400

accacacctt caagaaggtc tccatcacct tcgactcctc cgtcagctgg cccggcaacg 20460

accggctcct gacgcccaac gagttcgaaa tcaagcgcac cgtcgacggc gagggctaca 20520

acgtggccca gtgcaacatg accaaggact ggttcctggt ccagatgctg gcccactaca 20580

acatcggcta ccagggcttc tacgtgcccg agggctacaa ggaccgcatg tactccttct 20640

tccgcaactt ccagcccatg agccgccagg tggtggacga ggtcaactac aaggactacc 20700

aggccgtcac cctggcctac cagcacaaca actcgggctt cgtcggctac ctcgcgccca 20760

ccatgcgcca gggccagccc taccccgcca actaccccta cccgctcatc ggcaagagcg 20820

ccgtcaccag cgtcacccag aaaaagttcc tctgcgacag ggtcatgtgg cgcatcccct 20880

tctccagcaa cttcatgtcc atgggcgcgc tcaccgacct cggccagaac atgctctatg 20940

ccaactccgc ccacgcgcta gacatgaatt tcgaagtcga ccccatggat gagtccaccc 21000

ttctctatgt tgtcttcgaa gtcttcgacg tcgtccgagt gcaccagccc caccgcggcg 21060

tcatcgaggc cgtctacctg cgcaccccct tctcggccgg taacgccacc acctaagctc 21120

ttgcttcttg caagccatgg ccgcgggctc cggcgagcag gagctcaggg ccatcatccg 21180

cgacctgggc tgcgggccct acttcctggg caccttcgat aagcgcttcc cgggattcat 21240

ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccgggggcga 21300

gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgacccctt 21360

cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg 21420

ccgcagcgcc ctggccaccg aggaccgctg cgtcaccctg gaaaagtcca cccagaccgt 21480

gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt 21540

gcactggccc gaccgcccca tggacaagaa ccccaccatg aacttgctga cgggggtgcc 21600
caacggcatg ctccagtcgc cccaggtgga acccaccctg cgccgcaacc aggaggcgct 21660
ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa 21720
ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaatgtc 21780
tttaataaac agcactttca tgttacacat gcatctgaga tgatttattt agaaatcgaa 21840
agggttctgc cgggtctcgg catggcccgc gggcagggac acgttgcgga actggtactt 21900
ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc 21960

ggtccacagc ttccgcgtca gttgcagggc gcccagcagg tcgggcgcgg agatcttgaa 22020
atcgcagttg ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg 22080
gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc 22140
cacgtcgagg tcctcggcgt tggccatccc gaagggggtc atcttgcagg tctgccttcc 22200
catggtgggc acgcacccgg gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat 22260
ctgggcctgg tcggcgttca tccccgggta catggccttc atgaaagcct ccaattgcct 22320
gaacgcctgc tgggccttgg ctccctcggt gaagaagacc ccgcaggact tgctagagaa 22380

ctggttggtg gcgcacccgg cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg 22440
caccacgctg cgcccccagc ggttctgggt gatcttggcc cggtcggggt tctccttcag 22500
cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat 22560
ggtggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag 22620
cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc 22680
ctgcaggaag cggcccatca tggtggtcag ggtcttgttg ctagtgaagg tcagcggaat 22740
gccgcggtgc tcctcgttga tgtacaggtg gcagatgcgg cggtacacct cgccctgctc 22800

gggcatcagc tggaagttgg ctttcaggtc ggtctccacg cggtagcggt ccatcagcat 22860
agtcatgatt tccataccct tctcccaggc cgagacgatg ggcaggctca tagggttctt 22920
caccatcatc ttagcgctag cagccgcggc cagggggtcg ctctcgtcca gggtctcaaa 22980
gctccgcttg ccgtccttct cggtgatccg caccgggggg tagctgaagc ccacggccgc 23040
cagctcctcc tcggcctgtc tttcgtcctc gctgtcctgg ctgacgtcct gcaggaccac 23100
atgcttggtc ttgcggggtt tcttcttggg cggcagcggc ggcggagatg ttggagatgg 23160
cgaggggggag cgcgagttct cgctcaccac tactatctct tcctcttctt ggtccgaggc 23220

cacgcggcgg taggtatgtc tcttcggggg cagaggcgga ggcgacgggc tctcgccgcc 23280

gcgacttggc ggatggctgg cagagcccct tccgcgttcg ggggtgcgct cccggcggcg 23340

ctctgactga cttcctccgc ggccggccat tgtgttctcc tagggaggaa caacaagcat 23400

ggagactcag ccatcgccaa cctcgccatc tgcccccacc gccgacgaga agcagcagca 23460

gcagaatgaa agcttaaccg ccccgccgcc cagccccgcc acctccgacg cggccgtccc 23520

agacatgcaa gagatggagg aatccatcga gattgacctg ggctatgtga cgcccgcgga 23580

gcacgaggag gagctggcag tgcgcttttc acaagaagag atacaccaag aacagccaga 23640

gcaggaagca gagaatgagc agagtcaggc tgggctcgag catgacggcg actacctcca 23700

cctgagcggg ggggaggacg cgctcatcaa gcatctggcc cggcaggcca ccatcgtcaa 23760

ggatgcgctg ctcgaccgca ccgaggtgcc cctcagcgtg gaggagctca gccgcgccta 23820

cgagttgaac ctcttctcgc cgcgcgtgcc ccccaagcgc cagcccaatg gcacctgcga 23880

gcccaacccg cgcctcaact tctacccggt cttcgcggtg cccgaggccc tggccaccta 23940

ccacatcttt ttcaagaacc aaaagatccc cgtctcctgc cgcgccaacc gcacccgcgc 24000

cgacgccctt ttcaacctgg gtcccggcgc ccgcctacct gatatcgcct ccttggaaga 24060

ggttcccaag atcttcgagg gtctgggcag cgacgagact cgggccgcga acgctctgca 24120

aggagaagga ggagagcatg agcaccacag cgccctggtc gagttggaag gcgacaacgc 24180

gcggctggcg gtgctcaaac gcacggtcga gctgacccat ttcgcctacc cggctctgaa 24240

cctgcccccc aaagtcatga gcgcggtcat ggaccaggtg ctcatcaagc gcgcgtcgcc 24300

catctccgag gacgagggca tgcaagactc cgaggagggc aagcccgtgg tcagcgacga 24360

gcagctggcc cggtggctgg gtcctaatgc tagtccccag agtttggaag agcggcgcaa 24420

actcatgatg gccgtggtcc tggtgaccgt ggagctggag tgcctgcgcc gcttcttcgc 24480

cgacgcggag accctgcgca aggtcgagga gaacctgcac tacctcttca ggcacgggtt 24540

cgtgcgccag gcctgcaaga tctccaacgt ggagctgacc aacctggtct cctacatggg 24600

catcttgcac gagaaccgcc tggggcagaa cgtgctgcac accaccctgc gcggggaggc 24660

ccggcgcgac tacatccgcg actgcgtcta cctctacctc tgccacacct ggcagacggg 24720

catgggcgtg tggcagcagt gtctggagga gcagaacctg aaagagctct gcaagctcct 24780

gcagaagaac ctcaagggtc tgtggaccgg gttcgacgag cgcaccaccg cctcggacct 24840

ggccgacctc attttccccg agcgcctcag gctgacgctg cgcaacggcc tgcccgactt 24900

tatgagccaa agcatgttgc aaaactttcg ctctttcatc ctcgaacgct ccggaatcct 24960

gcccgccacc tgctccgcgc tgccctcgga cttcgtgccg ctgaccttcc gcgagtgccc 25020

cccgccgctg tggagccact gctacctgct gcgcctggcc aactacctgg cctaccactc 25080

ggacgtgatc gaggacgtca gcggcgaggg cctgctcgag tgccactgcc gctgcaacct 25140

ctgcacgccg caccgctccc tggcctgcaa cccccagctg ctgagcgaga cccagatcat 25200

cggcaccttc gagttgcaag ggcccagcga aggcgagggt tcagccgcca agggggggtct 25260

gaaactcacc ccggggctgt ggacctcggc ctacttgcgc aagttcgtgc ccgaggacta 25320

ccatcccttc gagatcaggt tctacgagga ccaatcccat ccgcccaagg ccgagctgtc 25380

ggcctgcgtc atcacccagg gggcgatcct ggcccaattg caagccatcc agaaatcccg 25440

ccaagaattc ttgctgaaaa agggccgcgg ggtctacctc gacccccaga ccggtgagga 25500

gctcaacccc ggcttccccc aggatgcccc gaggaaacaa gaagctgaaa gtggagctgc 25560

cgcccgtgga ggatttggag gaagactggg agaacagcag tcaggcagag gaggaggaga 25620

tggaggaaga ctgggacagc actcaggcag aggaggacag cctgcaagac agtctggagg 25680

aagacgagga ggaggcagag gaggaggtgg aagaagcagc cgccgccaga ccgtcgtcct 25740

cggcgggggga gaaagcaagc agcacggata ccatctccgc tccgggtcgg ggtcccgctc 25800

gaccacacag tagatgggac gagaccggac gattcccgaa ccccaccacc cagaccggta 25860

agaaggagcg gcagggatac aagtcctggc gggggcacaa aaacgccatc gtctcctgct 25920

tgcaggcctg cgggggcaac atctccttca cccggcgcta cctgctcttc caccgcgggg 25980

tgaactttcc ccgcaacatc ttgcattact accgtcacct ccacagcccc tactacttcc 26040

aagaagaggc agcagcagca gaaaaagacc agcagaaaac cagcagctag aaaatccaca 26100

gcggcggcag caggtggact gaggatcgcg gcgaacgagc cggcgcaaac ccgggagctg 26160

aggaaccgga tctttcccac cctctatgcc atcttccagc agagtcgggg gcaggagcag 26220

gaactgaaag tcaagaaccg ttctctgcgc tcgctcaccc gcagttgtct gtatcacaag 26280

agcgaagacc aacttcagcg cactctcgag gacgccgagg ctctcttcaa caagtactgc 26340

gcgctcactc ttaaagagta gcccgcgccc gcccagtcgc agaaaaaggc gggaattacg 26400

tcacctgtgc ccttcgccct agccgcctcc acccatcatc atgagcaaag agattcccac 26460

gccttacatg tggagctacc agccccagat gggcctggcc gccggtgccg cccaggacta 26520

ctccacccgc atgaattggc tcagcgccgg gcccgcgatg atctcacggg tgaatgacat 26580

ccgcgcccac cgaaaccaga tactcctaga acagtcagcg ctcaccgcca cgccccgcaa 26640

tcacctcaat ccgcgtaatt ggcccgccgc cctggtgtac caggaaattc cccagcccac 26700

gaccgtacta cttccgcgag acgcccaggc cgaagtccag ctgactaact caggtgtcca 26760

gctggcgggc ggcgccaccc tgtgtcgtca ccgccccgct cagggtataa agcggctggt 26820

gatccggggc agaggcacac agctcaacga cgaggtggtg agctcttcgc tgggtctgcg 26880

acctgacgga gtcttccaac tcgccggatc ggggagatct tccttcacgc ctcgtcaggc 26940

cgtcctgact ttggagagtt cgtcctcgca gccccgctcg ggtggcatcg gcactctcca 27000

gttcgtggag gagttcactc cctcggtcta cttcaacccc ttctccggct ccccccggcca 27060
ctacccggac gagttcatcc cgaacttcga cgccatcagc gagtcggtgg acggctacga 27120
ttgaatgtcc catggtggcg cagctgacct agctcggctt cgacacctgg accactgccg 27180
ccgcttccgc tgcttcgctc gggatctcgc cgagtttgcc tactttgagc tgcccgagga 27240
gcaccctcag ggcccggccc acggagtgcg gatcgtcgtc gaaggggggcc tcgactccca 27300
cctgcttcgg atcttcagcc agcgtccgat cctggtcgag cgcgagcaag gacagaccct 27360
tctgactctg tactgcatct gcaaccaccc cggcctgcat gaaagtcttt gttgtctgct 27420

gtgtactgag tataataaaa gctgagatca gcgactactc cggacttccg tgtgttcctg 27480
aatccatcaa ccagtctttg ttcttcaccg ggaacgagac cgagctccag ctccagtgta 27540
agccccacaa gaagtacctc acctggctgt tccagggctc cccgatcgcc gttgtcaacc 27600
actgcgacaa cgacggagtc ctgctgagcg gccctgccaa ccttactttt tccacccgca 27660
gaagcaagct ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac 27720
cctgccatca caccttccac ctgatcccga ataccacagc gtcgctcccc gctactaaca 27780
accaaactaa cctccaccaa cgccaccgtc gcgacctttc tgaatctaat actaccaccc 27840

acaccggagg tgagctccga ggtcaaccaa cctctgggat ttactacggc ccctgggagg 27900
tggttgggtt aatagcgcta ggcctagttg cgggtgggct tttggttctc tgctacctat 27960
acctcccttg ctgttcgtac ttagtggtgc tgtgttgctg gtttaagaaa tggggaagat 28020
caccctagtg agctgcggtg cgctggtggc ggtgttgctt tcgattgtgg gactgggcgg 28080
tgcggctgta gtgaaggaga aggccgatcc ctgcttgcat ttcaatccca acaaatgcca 28140
gctgagtttt cagcccgatg gcaatcggtg cgcggtactg atcaagtgcg gatgggaatg 28200
cgagaacgtg agaatcgagt acaataacaa gactcggaac aatactctcg cgtccgtgtg 28260

gcagcccggg gaccccgagt ggtacaccgt ctctgtcccc ggtgctgacg gctccccgcg 28320
caccgtgaat aatactttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa 28380
gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag 28440
cctgtgcacg gcgctaatca ccgctatcgt gtgcctgagc attcacatgc tcatcgctat 28500
tcgccccaga aataatgccg aaaaagaaaa acagccataa cgtttttttt cacacctttt 28560
tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca 28620
tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaaggtcc 28680

agaaaaagcc acagaagttt catggtattg ttattttaat gaatcagatg tatctactga 28740

actctgtgga aacaataaca aaaaaaatga gagcattact ctcatcaagt ttcaatgtgg 28800

atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac 28860

agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag 28920

aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat 28980

ttttgccatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat 29040

tcccaaatcc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatcgcctt 29100

gtgcatggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca 29160

cttactaagt gttgaatttt aatttttag aaccatgaag atcctaggcc ttttaatttt 29220

ttctatcatt acctctgctc tatgcaattc tgacaatgag gacgttactg tcgttgtcgg 29280

atcaaattat acactgaaag gtccagcgaa gggtatgctt tcgtggtatt gctattttgg 29340

atctgacact acagaaactg aattatgcaa tcttaagaat ggcaaaattc aaaattctaa 29400

aattaacaat tatatatgca atggtactga tctgatactc ctcaatatca cgaaatcata 29460

tgctggcagt tacacctgcc ctggagatga tgctgacagt atgatttttt acaaagtaac 29520

tgttgttgat cccactactc cacctccacc caccacaact actcacacca cacacacaga 29580

tcaaaccgca gcagaggagg cagcaaagtt agccttgcag gtccaagaca gttcatttgt 29640

tggcattacc cctacacctg atcagcggtg tccggggctg ctagtcagcg gcattgtcgg 29700

tgtgctttcg ggattagcag tcataatcat ctgcatgttc atttttgctt gctgctatag 29760

aaggctttac cgacaaaaat cagacccact gctgaacctc tatgtttaat tttttccaga 29820

gtcatgaagg cagttagcgc tctagttttt tgttctttga ttggcattgt tttttgcaat 29880

cctattccta aagttagctt tattaaagat gtgaatgtta ctgaggggggg caatgtgaca 29940

ctggtaggtg tagagggtgc tgaaaacacc acctggacaa aataccacct caatgggtgg 30000

aaagatattt gcaattggag tgtattagtt tatacatgtg agggagttaa tcttaccatt 30060

gtcaatgcca cctcagctca aaatggtaga attcaaggac aaagtgtcag tgtatctaat 30120

gggtatttta cccaacatac ttttatctat gacgttaaag tcataccact gcctacgcct 30180

agcccaccta gcactaccac acagacaacc cacactacac agacaaccac atacagtaca 30240

ttaaatcagc ctaccaccac tacagcagca gaggttgcca gctcgtctgg ggtccgagtg 30300

gcatttttga tgtgggcccc atctagcagt cccactgcta gtaccaatga gcagactact 30360

gaatttttgt ccactgtcga gagccacacc acagctacct ccagtgcctt ctctagcacc 30420

gccaatctct cctcgctttc ctctacacca atcagtcccg ctactactcc tagccccgct 30480

cctcttccca ctcccctgaa gcaaacagac ggcggcatgc aatggcagat caccctgctc 30540

attgtgatcg ggttggtcat cctggccgtg ttgctctact acatcttctg ccgccgcatt 30600

cccaacgcgc accgcaagcc ggtctacaag cccatcattg tcgggcagcc ggagccgctt 30660

caggtggaag ggggtctaag gaatcttctc ttctctttta cagtatggtg attgaactat 30720

gattcctaga caattcttga tcactattct tatctgcctc ctccaagtct gtgccaccct 30780

cgctctggtg gccaacgcca gtccagactg tattgggccc ttcgcctcct acgtgctctt 30840

tgccttcacc acctgcatct gctgctgtag catagtctgc ctgcttatca ccttcttcca 30900

gttcattgac tggatctttg tgcgcatcgc ctacctgcgc caccaccccc agtaccgcga 30960

ccagcgagtg gcgcggctgc tcaggctcct ctgataagca tgcgggctct gctacttctc 31020

gcgcttctgc tgttagtgct cccccgtccc gtcgaccccc ggtcccccac ccagtccccc 31080

gaggaggtcc gcaaatgcaa attccaagaa ccctggaaat tcctcaaatg ctaccgccaa 31140

aaatcagaca tgcatcccag ctggatcatg atcattggga tcgtgaacat tctggcctgc 31200

accctcatct cctttgtgat ttacccctgc tttgactttg gttggaactc gccagaggcg 31260

ctctatctcc cgcctgaacc tgacacacca ccacagcaac ctcaggcaca cgcactacca 31320

ccactacagc ctaggccaca atacatgccc atattagact atgaggccga gccacagcga 31380

cccatgctcc ccgctattag ttacttcaat ctaaccggcg gagatgactg acccactggc 31440

caacaacaac gtcaacgacc ttctcctgga catggacggc cgcgcctcgg agcagcgact 31500

cgcccaactt cgcattcgcc agcagcagga gagagccgtc aaggagctgc aggatgcggt 31560

ggccatccac cagtgcaaga gaggcatctt ctgcctggtg aaacaggcca agatctccta 31620

cgaggtcact ccaaacgacc atcgcctctc ctacgagctc ctgcagcagc gccagaagtt 31680

cacctgcctg gtcggagtca accccatcgt catcacccag cagtctggcg ataccaaggg 31740

gtgcatccac tgctcctgcg actcccccga ctgcgtccac actctgatca agaccctctg 31800

cggcctccgc gacctcctcc ccatgaacta atcacccccc tatccagtga aataaagatc 31860

atattgatga tgattttaca gaaataaaaa ataatcattt gatttgaaat aaagatacaa 31920

tcatattgat gatttgagtt taacaaaaaa ataaagaatc acttacttga aatctgatac 31980

caggtctctg tccatgtttt ctgccaacac cacttcactc ccctcttccc agctctggta 32040

ctgcaggccc cggcgggctg caaacttcct ccacacgctg aaggggatgt caaattcctc 32100

ctgtccctca atcttcattt tatcttctat cagatgtcca aaaagcgcgt ccgggtggat 32160

gatgacttcg accccgtcta cccctacgat gcagacaacg caccgaccgt gcccttcatc 32220

aacccccccc tcgtctcttc agatggattc caagagaagc ccctgggggt gttgtccctg 32280

cgactggccg accccgtcac caccaagaac ggggaaatca ccctcaagct gggagagggg 32340

gtggacctcg attcctcggg aaaactcatc tccaacacgg ccaccaaggc cgccgccccc 32400

ctcagttttt ccaacaacac catttccctt aacatggatc acccctttta cactaaagat 32460

ggaaaattat ccttacaagt ttctccacca ttaaatatac tgagaacaag cattctaaac 32520
acactagctt taggttttgg atcaggttta ggactccgtg gctctgcctt ggcagtacag 32580
ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt 32640
ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa 32700
tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt 32760
acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc 32820
tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg 32880

acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca 32940
ctttgcttga ctaaatgtgg tagtcaaata ctggccactg tgtcagtctt agttgtagga 33000
agtggaaacc taaaccccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt 33060
gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg 33120
cagggagata gcatagatgg cactccatat accaatgctg taggattcat gcccaattta 33180
aaagcttatc caaagtcaca aagttctact actaaaaata atatagtagg gcaagtatac 33240
atgaatggag atgtttcaaa acctatgctt ctcactataa ccctcaatgg tactgatgac 33300

agcaacagta catattcaat gtcattttca tacacctgga ctaatggaag ctatgttgga 33360
gcaacatttg gggctaactc ttataccttc tcatacatcg cccaagaatg aacactgtat 33420
cccaccctgc atgccaaccc ttcccacccc actctgtgga acaaactctg aaacacaaaa 33480
taaaataaag ttcaagtgtt ttattgattc aacagtttta caggattcga gcagttattt 33540
ttcctccacc ctcccaggac atggaataca ccaccctctc cccccgcaca gccttgaaca 33600
tctgaatgcc attggtgatg gacatgcttt tggtctccac gttccacaca gtttcagagc 33660
gagccagtct cgggtcggtc agggagatga aaccctccgg gcactcccgc atctgcacct 33720

cacagctcaa cagctgagga ttgtcctcgg tggtcgggat cacggttatc tggaagaagc 33780
agaagagcgg cggtgggaat catagtccgc gaacgggatc ggccggtggt gtcgcatcag 33840
gccccgcagc agtcgctgcc gccgccgctc cgtcaagctg ctgctcaggg ggtccgggtc 33900
cagggactcc ctcagcatga tgcccacggc cctcagcatc agtcgtctgg tgcggcgggc 33960
gcagcagcgc atgcggatct cgctcaggtc gctgcagtac gtgcaacaca gaaccaccag 34020
gttgttcaac agtccatagt tcaacacgct ccagccgaaa ctcatcgcgg gaaggatgct 34080
acccacgtgg ccgtcgtacc agatcctcag gtaaatcaag tggtgccccc tccagaacac 34140

gctgcccacg tacatgatct ccttgggcat gtggcggttc accacctccc ggtaccacat 34200

caccctctgg ttgaacatgc agccccggat gatcctgcgg aaccacaggg ccagcaccgc 34260

cccgcccgcc atgcagcgaa gagaccccgg gtcccggcaa tggcaatgga ggacccaccg 34320

ctcgtacccg tggatcatct gggagctgaa caagtctatg ttggcacagc acaggcatat 34380

gctcatgcat ctcttcagca ctctcaactc ctcgggggtc aaaaccatat cccagggcac 34440

ggggaactct tgcaggacag cgaaccccgc agaacagggc aatcctcgca cagaacttac 34500

attgtgcatg gacagggtat cgcaatcagg cagcaccggg tgatcctcca ccagagaagc 34560

gcgggtctcg gtctcctcac agcgtggtaa gggggccggc cgatacgggt gatggcggga 34620

cgcggctgat cgtgttcgcg accgtgtcat gatgcagttg ctttcggaca ttttcgtact 34680

tgctgtagca gaacctggtc cgggcgctgc acaccgatcg ccggcggcgg tctcggcgct 34740

tggaacgctc ggtgttgaaa ttgtaaaaca gccactctct cagaccgtgc agcagatcta 34800

gggcctcagg agtgatgaag atcccatcat gcctgatggc tctgatcaca tcgaccaccg 34860

tggaatgggc cagacccagc cagatgatgc aattttgttg ggtttcggtg acggcggggg 34920

agggaagaac aggaagaacc atgattaact tttaatccaa acggtctcgg agtacttcaa 34980

aatgaagatc gcggagatgg cacctctcgc ccccgctgtg ttggtggaaa ataacagcca 35040

ggtcaaaggt gatacggttc tcgagatgtt ccacggtggc ttccagcaaa gcctccacgc 35100

gcacatccag aaacaagaca atagcgaaag cgggagggtt ctctaattcc tcaatcatca 35160

tgttacactc ctgcaccatc cccagataat tttcattttt ccagccttga atgattcgaa 35220

ctagttcctg aggtaaatcc aagccagcca tgataaagag ctcgcgcaga gcgccctcca 35280

ccggcattct taagcacacc ctcataattc caagatattc tgctcctggt tcacctgcag 35340

cagattgaca agcggaatat caaaatctct gccgcgatcc ctgagctcct ccctcagcaa 35400

taactgtaag tactctttca tatcctctcc gaaattttta gccataggac caccaggaat 35460

aagattaggg caagccacag tacagataaa ccgaagtcct ccccagtgag cattgccaaa 35520

tgcaagactg ctataagcat gctggctaga cccggtgata tcttccagat aactggacag 35580

aaaatcgccc aggcaatttt taagaaaatc aacaaaagaa aaatcctcca ggtggacgtt 35640

tagagcctcg ggaacaacga tgaagtaaat gcaagcggtg cgttccagca tggttagtta 35700

gctgatctgt agaaaaaaca aaaatgaaca ttaaaccatg ctagcctggc gaacaggtgg 35760

gtaaatcgtt ctctccagca ccaggcaggc cacggggtct ccggcgcgac cctcgtaaaa 35820

attgtcgcta tgattgaaaa ccatcacaga gagacgttcc cggtggccgg cgtgaatgat 35880

tcgacaagat gaatacaccc ccggaacatt ggcgtccgcg agtgaaaaaa agcgcccgag 35940

gaagcaataa ggcactacaa tgctcagtct caagtccagc aaagcgatgc catgcggatg 36000

aagcacaaaa ttctcaggtg cgtacaaaat gtaattactc ccctcctgca caggcagcaa 36060

agcccccgat ccctccaggt acacatacaa agcctcagcg tccatagctt accgagcagc 36120

agcacacaac aggcgcaaga gtcagagaaa ggctgagctc taacctgtcc acccgctctc 36180

tgctcaatat atagcccaga tctacactga cgtaaaggcc aaagtctaaa aatacccgcc 36240

aaataatcac acacgcccag cacacgccca gaaaccggtg acacactcaa aaaaatacgc 36300

gcacttcctc aaacgcccaa aactgccgtc atttccgggt tcccacgcta cgtcatcaaa 36360

acacgacttt caaattccgt cgaccgttaa aaacgtcacc cgccccgccc ctaacggtcg 36420

cccgtctctc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt 36480

aacgcgcaca aaaagtttga ggtatattat tgatgatgg 36519

<210> 2

<211> 31588

<212> DNA

<213> Artificial Sequence

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

<400> 2

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

ccgggatgtt ccaggcactg tccgaaggct gcacacccta tgatattaac cagatgctga 1320

atgtcctggg agaccaccag gtctctggcc tggagcagct ggagagcatc atcaacttcg 1380

agaagctgac cgagtggaca agctccaatg tgatgcctat cctgtcccca ctgaccaagg 1440

gcatcctggg cttcgtgttt accctgacag tgccttctga gcggggcctg tcttgcatca 1500

gcgaggcaga cgcaaccaca ccagagtccg ccaatctggg cgaggagatc ctgtctcagc 1560

tgtacctgtg gccccgggtg acatatcact ccccttctta cgcctatcac cagttcgagc 1620

ggagagccaa gtacaagaga cacttcccag gctttggcca gtctctgctg ttcggctacc 1680

ccgtgtacgt gttcggcgat tgcgtgcagg gcgactggga tgccatccgg tttagatact 1740

gcgcaccacc tggatatgca ctgctgaggt gtaacgacac caattattcc gccctgctgg 1800

cagtgggcgc cctggagggc cctcgcaatc aggattggct gggcgtgcca aggcagctgg 1860

tgacacgcat gcaggccatc cagaacgcag gcctgtgcac cctggtggca atgctggagg 1920

agacaatctt ctggctgcag gcctttctga tggccctgac cgacagcggc cccaagacaa 1980

acatcatcgt ggattcccag tacgtgatgg gcatctccaa gccttctttc caggagtttg 2040

tggactggga gaacgtgagc ccagagctga attccaccga tcagccattc tggcaggcag 2100

gaatcctggc aaggaacctg gtgcctatgg tggccacagt gcagggccag aatctgaagt 2160

accagggcca gagcctggtc atcagcgcct ccatcatcgt gtttaacctg ctggagctgg 2220

agggcgacta tcgggacgat ggcaacgtgt gggtgcacac cccactgagc cccagaacac 2280

tgaacgcctg ggtgaaggcc gtggaggaga agaagggcat cccagtgcac ctggagctgg 2340

cctccatgac caatatggag ctgatgtcta gcatcgtgca ccagcaggtg aggacatacg 2400

gacccgtgtt catgtgcctg ggaggcctgc tgaccatggt ggcaggagcc gtgtggctga 2460

cagtgcgggt gctggagctg ttcagagccg cccagctggc caacgatgtg gtgctgcaga 2520

tcatggagct gtgcggagca gcctttcgcc aggtgtgcca caccacagtg ccatggccca 2580

atgcctccct gacccccaag tggaacaatg agacaacaca gcctcagatc gccaactgta 2640

gcgtgtacga cttcttcgtg tggctgcact actatagcgt gagggatacc ctgtggcccc 2700

gcgtgacata ccacatgaat aagtacgcct atcacatgct ggagaggcgc gccaagtata 2760

agagaggccc tggcccaggc gcaaagtttg tggcagcatg gaccctgaag gccgccgccg 2820

gccccggccc cggccagtat atcaaggcta acagtaagtt cattggaatc acagagctgg 2880

gacccggacc tggataatga gtttaaactc ccatttaaat gtgagggtta atgcttcgag 2940

cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 3000

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 3060

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 3120

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaataact ataacggtcc 3180

taaggtagcg agtgagtagt gttctggggc gggggaggac ctgcatgagg gccagaataa 3240

ctgaaatctg tgcttttctg tgtgttgcag cagcatgagc ggaagcggct cctttgaggg 3300

aggggtattc agcccttatc tgacggggcg tctcccctcc tgggcgggag tgcgtcagaa 3360

tgtgatggga tccacggtgg acggccggcc cgtgcagccc gcgaactctt caaccctgac 3420

ctatgcaacc ctgagctctt cgtcgttgga cgcagctgcc gccgcagctg ctgcatctgc 3480

cgccagcgcc gtgcgcggaa tggccatggg cgccggctac tacggcactc tggtggccaa 3540

ctcgagttcc accaataatc ccgccagcct gaacgaggag aagctgttgc tgctgatggc 3600

ccagctcgag gccttgaccc agcgcctggg cgagctgacc cagcaggtgg ctcagctgca 3660

ggagcagacg cgggccgcgg ttgccacggt gaaatccaaa taaaaaatga atcaataaat 3720

aaacggagac ggttgttgat tttaacacag agtctgaatc tttatttgat ttttcgcgcg 3780

cggtaggccc tggaccaccg gtctcgatca ttgagcaccc ggtggatctt ttccaggacc 3840

cggtagaggt gggcttggat gttgaggtac atgggcatga gcccgtcccg ggggtggagg 3900

tagctccatt gcagggcctc gtgctcgggg gtggtgttgt aaatcaccca gtcatagcag 3960

gggcgcaggg catggtgttg cacaatatct ttgaggagga gactgatggc cacgggcagc 4020

cctttggtgt aggtgtttac aaatctgttg agctgggagg gatgcatgcg gggggagatg 4080

aggtgcatct tggcctggat cttgagattg gcgatgttac cgcccagatc ccgcctgggg 4140

ttcatgttgt gcaggaccac cagcacggtg tatccggtgc acttggggaa tttatcatgc 4200

aacttggaag ggaaggcgtg aaagaatttg gcgacgcctt tgtgcccgcc caggttttcc 4260

atgcactcat ccatgatgat ggcgatgggc ccgtgggcgg cggcctgggc aaagacgttt 4320

cgggggtcgg acacatcata gttgtggtcc tgggtgaggt catcataggc cattttaatg 4380

aatttggggc ggagggtgcc ggactggggg acaaaggtac cctcgatccc gggggcgtag 4440

ttcccctcac agatctgcat ctcccaggct ttgagctcgg agggggggat catgtccacc 4500

tgcggggcga taaagaacac ggtttccggg gcgggggaga tgagctgggc cgaaagcaag 4560

ttccggagca gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc 4620

tgcaggtggt agttgaggga gagacagctg ccgtcctccc ggaggagggg ggccacctcg 4680
ttcatcatct cgcgcacgtg catgttctcg cgcaccagtt ccgccaggag gcgctctccc 4740
cccagggata ggagctcctg gagcgaggcg aagtttttca gcggcttgag tccgtcggcc 4800
atgggcattt tggagagggt ttgttgcaag agttccaggc ggtcccagag ctcggtgatg 4860
tgctctacgg catctcgatc cagcagacct cctcgtttcg cgggttggga cggctgcggg 4920
agtagggcac cagacgatgg gcgtccagcg cagccagggt ccggtccttc cagggtcgca 4980
gcgtccgcgt cagggtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg 5040

cgagggtgcg cttcaggctc atccggctgg tcgaaaaccg ctcccgatcg gcgccctgcg 5100
cgtcggccag gtagcaattg accatgagtt cgtagttgag cgcctcggcc gcgtggcctt 5160
tggcgcggag cttacctttg gaagtctgcc cgcaggcggg acagaggagg gacttgaggg 5220
cgtagagctt gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagtggg 5280
cgcagacggt ctcgcactcc acgagccagg tgaggtcggg ctggtcgggg tcaaaaacca 5340
gtttcccgcc gttctttttg atgcgtttct taccttggt ctccatgagc tcgtgtcccc 5400
gctgggtgac aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga 5460

gcggtgtgcc gcggtcctcc tcgtagagga accccgccca ctccgagacg aaagcccggg 5520
tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt 5580
ccacctttc cagggtatgc aaacacatgt ccccctcgtc cacatccagg aaggtgattg 5640
gcttgtaagt gtaggccacg tgaccggggg tcccggccgg gggggtataa aagggtgcgg 5700
gtccctgctc gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta 5760
ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg 5820
aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct 5880

ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt 5940
tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct 6000
tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg 6060
tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt 6120
ccacactggt ggccacctcg ccgcgcaggg gctcattagt ccagcagagg cgtccgccct 6180
tgcgcgagca gaaggggggc aggggtcca gcatgacctc gtcgggggggg tcggcatcga 6240
tggtgaagat gccgggcagg aggtcggggt caaagtagct gatggaagtg gccagatcgt 6300

ccagggcagc ttgccattcg cgcacggcca gcgcgcgctc gtagggactg aggggcgtgc 6360

cccagggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga 6420

ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc 6480

gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cgggcccagg ttggtgcgac 6540

tgggcttttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg 6600

tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggatgaagt 6660

gggcgtagga gtcttgcagc ttggcgacga gctcggcggt gactaggacg tccagagcgc 6720

agtagtcgag ggtctcctgg atgatgtcat acttgagctg tccctttgt ttccacagct 6780

cgcggttgag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aacccgtcct 6840

gatctgcacg gtaagagcct agcatgtaga actggttgac ggccttgtag gcgcagcagc 6900

ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg 6960

cgaaagtgtc cctgaccatg accttgagga actggtgctt gaagtcgata tcgtcgcagc 7020

ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga 7080

aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga 7140

aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga 7200

agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt 7260

ggggcagttt cttgagctcc tcgtaggtga gctcgtcggg gtcgctgagc ccgtgctgct 7320

cgagcgccca gtcggcgaga tgggggttgg cgcggaggaa ggaagtccag agatccacgg 7380

ccagggcggt ttgcagacgg tcccggtact gacggaactg ctgcccgacg gccatttttt 7440

cgggggtgac gcagtagaag gtgcgggggt ccccgtgcca gcgatcccat ttgagctgga 7500

gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtttc atgaccagca 7560

tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg 7620

tgaggaagag cctttcggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc 7680

accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac 7740

actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt 7800

gctgcacgag ctgtacctga gttcctttga cgaggaattt cagtgggaag tggagtcgtg 7860

gcgcctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga 7920

tggtggtcat gctgacgagc ccgcgcggga ggcaggtcca gacctcggcg cgagcgggtc 7980

ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag 8040

tcaggtcagt gggcagcggc ggcgcgcggt tgacttgcag gagtttttcc agggcgcgcg 8100

ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca 8160

gggtcccgtg cccctggggt gtgaccaccg tcccccgttt cttcttgggc ggctggggcg 8220

acgggggcgg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag 8280

gggcggctcg gggcccggag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag 8340

gttctggtac tgcgcccgga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg 8400

gatctgacgc ctctgggtga aggccacggg acccgtgagt ttgaacctga aagagagttc 8460

gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc 8520

cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc 8580

tccgcggccg gcgcgctcca cggtggccgc gaggtcgttg gagatgcggc ccatgagctg 8640

cgagaaggcg ttcatgcccg cctcgttcca gacgcggctg tagaccacga cgccctcggg 8700

atcgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc 8760

gtagttgcag aggcgctggt agaggtagtt gagcgtggtg gcgatgtgct cggtgacgaa 8820

gaaatacatg atccagcggc ggagcggcat ctcgctgacg tcgcccagcg cctccaaacg 8880

ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac 8940

ggtcaactcc tcctccagaa gacggatgag ctcggcgatg gtggcgcgca cctcgcgctc 9000

gaaggccccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac 9060

ttcctcctca ggcggcagtg gtggcggggg agggggcctg cgtcgccggc ggcgcacggg 9120

cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac 9180

ggcgcgcccg tcctcgcggg gccgcagcgt gaagacgccg ccgcgcatct ccaggtggcc 9240

gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt 9300

agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac 9360

gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc 9420

atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct 9480

gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag 9540

acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg 9600

catgagccgc tccacgggca cctcctcctc gcccgcgcgg ccgtgcatgc gcgtgagccc 9660

gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc 9720

ttgctggatc tgggtgaggg tggtctggaa gtcatcaaag tcgacgaagc ggtggtaggc 9780

tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc 9840

cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc 9900

gttgcaggtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggta 9960

gagcggccat cgctcggtgg cgggggcgcc gggcgcgagg tcctcgagca tggtgcggtg 10020

gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg 10080

gaactcgcgg acgcggttcc agatgttgcg cagcggcagg aagtagttca tggtgggcac 10140
ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg 10200
tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc 10260
cggttcgaat ctcgaatcag gctggagccg cagctaacgt ggtattggca ctcccgtctc 10320
gacccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt 10380
ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta 10440
gtctggagaa gaatcgccag ggttgcgttg cggtgtgccc cggttcgagg ccggccggat 10500

tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact 10560
tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcatcccgt 10620
actgcggcag atgcgccccc accaccctcc accgcaacaa cagccccctc cacagccggc 10680
gcttctgccc ccgccccagc agcaacttcc agccacgacc gccgcggccg ccgtgagcgg 10740
ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgcct 10800
gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctcgcgaggc 10860
ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg 10920

cgcggcccgg ttccacgcgg ggcgggagct gcggcgcggc ctggaccgaa agagggtgct 10980
gagggacgag gatttcgagg cggacgagct gacggggatc agccccgcgc gcgcgcacgt 11040
ggccgcggcc aacctggtca cggcgtacga gcagaccgtg aaggaggaga gcaacttcca 11100
aaaatccttc aacaaccacg tgcgcaccct gatcgcgcgc gaggaggtga ccctgggcct 11160
gatgcacctg tgggacctgc tggaggccat cgtgcagaac cccaccagca agccgctgac 11220
ggcgcagctg ttcctggtgg tgcagcatag tcgggacaac gaagcgttca gggaggcgct 11280
gctgaatatc accgagcccg agggccgctg gctcctggac ctggtgaaca ttctgcagag 11340

catcgtggtg caggagcgcg ggctgccgct gtccgagaag ctggcggcca tcaacttctc 11400
ggtgctgagt ttgggcaagt actacgctag gaagatctac aagaccccgt acgtgcccat 11460
agacaaggag gtgaagatcg acgggtttta catgcgcatg accctgaaag tgctgaccct 11520
gagcgacgat ctgggggtgt accgcaacga caggatgcac cgtgcggtga gcgccagcag 11580
gcggcgcgag ctgagcgacc aggagctgat gcatagtctg cagcgggccc tgaccggggc 11640
cgggaccgag ggggagagct actttgacat gggcgcggac ctgcactggc agcccagccg 11700
ccgggccttg gaggcggcgg caggacccta cgtagaagag gtggacgatg aggtggacga 11760

ggagggcgag tacctggaag actgatggcg cgaccgtatt tttgctagat gcaacaacaa 11820

cagccacctc ctgatcccgc gatgcgggcg gcgctgcaga gccagccgtc cggcattaac 11880

tcctcggacg attggaccca ggccatgcaa cgcatcatgg cgctgacgac ccgcaacccc 11940

gaagccttta gacagcagcc ccaggccaac cggctctcgg ccatcctgga ggccgtggtg 12000

ccctcgcgct ccaaccccac gcacgagaag gtcctggcca tcgtgaacgc gctggtggag 12060

aacaaggcca tccgcggcga cgaggccggc ctggtgtaca acgcgctgct ggagcgcgtg 12120

gcccgctaca acagcaccaa cgtgcagacc aacctggacc gcatggtgac cgacgtgcgc 12180

gaggccgtgg cccagcgcga gcggttccac cgcgagtcca acctgggatc catggtggcg 12240

ctgaacgcct tcctcagcac ccagcccgcc aacgtgcccc ggggccagga ggactacacc 12300

aacttcatca gcgccctgcg cctgatggtg accgaggtgc cccagagcga ggtgtaccag 12360

tccgggccgg actacttctt ccagaccagt cgccagggct tgcagaccgt gaacctgagc 12420

caggctttca agaacttgca gggcctgtgg ggcgtgcagg ccccggtcgg ggaccgcgcg 12480

acggtgtcga gcctgctgac gccgaactcg cgcctgctgc tgctgctggt ggcccccttc 12540

acggacagcg gcagcatcaa ccgcaactcg tacctgggct acctgattaa cctgtaccgc 12600

gaggccatcg gccaggcgca cgtggacgag cagacctacc aggagatcac ccacgtgagc 12660

cgcgccctgg gccaggacga cccgggcaac ctggaagcca ccctgaactt tttgctgacc 12720

aaccggtcgc agaagatccc gccccagtac gcgctcagca ccgaggagga gcgcatcctg 12780

cgttacgtgc agcagagcgt gggcctgttc ctgatgcagg agggggccac ccccagcgcc 12840

gcgctcgaca tgaccgcgcg caacatggag cccagcatgt acgccagcaa ccgcccgttc 12900

atcaataaac tgatggacta cttgcatcgg gcggccgcca tgaactctga ctatttcacc 12960

aacgccatcc tgaatcccca ctggctcccg ccgccggggt tctacacggg cgagtacgac 13020

atgcccgacc ccaatgacgg gttcctgtgg gacgatgtgg acagcagcgt gttctccccc 13080

cgaccgggtg ctaacgagcg ccccttgtgg aagaaggaag gcagcgaccg acgcccgtcc 13140

tcggcgctgt ccggccgcga gggtgctgcc gcggcggtgc ccgaggccgc cagtcctttc 13200

ccgagcttgc ccttctcgct gaacagtatc cgcagcagcg agctgggcag gatcacgcgc 13260

ccgcgcttgc tgggcgaaga ggagtacttg aatgactcgc tgttgagacc cgagcgggag 13320

aagaacttcc ccaataacgg gatagaaagc ctggtggaca agatgagccg ctggaagacg 13380

tatgcgcagg agcacaggga cgatccccgg gcgtcgcagg gggccacgag ccggggcagc 13440

gccgcccgta aacgccggtg gcacgacagg cagcggggac agatgtggga cgatgaggac 13500

tccgccgacg acagcagcgt gttggacttg ggtgggagtg gtaacccgtt cgctcacctg 13560

cgcccccgta tcgggcgcat gatgtaagag aaaccgaaaa taaatgatac tcaccaaggc 13620

catggcgacc agcgtgcgtt cgtttcttct ctgttgttgt tgtatctagt atgatgaggc 13680

gtgcgtaccc ggagggtcct cctccctcgt acgagagcgt gatgcagcag gcgatggcgg 13740

cggcggcgat gcagcccccg ctggaggctc cttacgtgcc cccgcggtac ctggcgccta 13800

cggaggggcg gaacagcatt cgttactcgg agctggcacc cttgtacgat accacccggt 13860

tgtacctggt ggacaacaag tcggcggaca tcgcctcgct gaactaccag aacgaccaca 13920

gcaacttcct gaccaccgtg gtgcagaaca atgacttcac ccccacggag gccagcaccc 13980

agaccatcaa ctttgacgag cgctcgcggt ggggcggcca gctgaaaacc atcatgcaca 14040

ccaacatgcc caacgtgaac gagttcatgt acagcaacaa gttcaaggcg cgggtgatgg 14100

tctcccgcaa gacccccaat ggggtgacag tgacagagga ttatgatggt agtcaggatg 14160

agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga 14220

ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg 14280

gggtgctgga gagcgacatc ggcgtgaagt tcgacactag gaacttcagg ctgggctggg 14340

accccgtgac cgagctggtc atgcccgggg tgtacaccaa cgaggctttc catcccgata 14400

ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg 14460

gcattcgcaa gaggcagccc ttccaggaag gcttccagat catgtacgag gatctggagg 14520

ggggcaacat ccccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag 14580

cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg 14640

caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcattc 14700

agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca 14760

ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtgcgct 14820

cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc 14880

tgcccgacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc 14940

cggtggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg 15000

tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg 15060

agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc 15120

ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg 15180

tgaccgttac tgacgccaga cgccgcacct gcccctacgt ctacaaggcc ctgggcatag 15240

tcgcgccgcg cgtcctctcg agccgcacct tctaaatgtc cattctcatc tcgcccagta 15300

ataacaccgg ttggggcctg cgcgcgccca gcaagatgta cggaggcgct cgccaacgct 15360

ccacgcaaca ccccgtgcgc gtgcgcgggc acttccgcgc tccctggggc gccctcaagg 15420

gccgcgtgcg gtcgcgcacc accgtcgacg acgtgatcga ccaggtggtg gccgacgcgc 15480

gcaactacac ccccgccgcc gcgcccgtct ccaccgtgga cgccgtcatc gacagcgtgg 15540

tggccgacgc gcgccggtac gcccgcgcca agagccggcg gcggcgcatc gcccggcggc 15600
accggagcac ccccgccatg cgcgcggcgc gagccttgct gcgcagggcc aggcgcacgg 15660
gacgcagggc catgctcagg gcggccagac gcgcggcttc aggcgccagc gccggcagga 15720
cccggagacg cgcggccacg gcggcggcag cggccatcgc cagcatgtcc cgcccgcggc 15780
gagggaacgt gtactgggtg cgcgacgccg ccaccggtgt gcgcgtgccc gtgcgcaccc 15840
gcccccctcg cacttgaaga tgttcacttc gcgatgttga tgtgtcccag cggcgaggag 15900
gatgtccaag cgcaaattca aggaagagat gctccaggtc atcgcgcctg agatctacgg 15960

ccctgcggtg gtgaaggagg aaagaaagcc ccgcaaaatc aagcgggtca aaaaggacaa 16020
aaaggaagaa gaaagtgatg tggacggatt ggtggagttt gtgcgcgagt tcgccccccg 16080
gcggcgcgtg cagtggcgcg ggcggaaggt gcaaccggtg ctgagacccg gcaccaccgt 16140
ggtcttcacg cccggcgagc gctccggcac cgcttccaag cgctcctacg acgaggtgta 16200
cggggatgat gatattctgg agcaggcggc cgagcgcctg ggcgagtttg cttacggcaa 16260
gcgcagccgt tccgcaccga aggaagaggc ggtgtccatc ccgctggacc acggcaaccc 16320
cacgccgagc ctcaagcccg tgaccttgca gcaggtgctg ccgaccgcgg cgccgcgccg 16380

ggggttcaag cgcgagggcg aggatctgta ccccaccatg cagctgatgg tgcccaagcg 16440
ccagaagctg gaagacgtgc tggagaccat gaaggtggac ccggacgtgc agcccgaggt 16500
caaggtgcgg cccatcaagc aggtggcccc gggcctgggc gtgcagaccg tggacatcaa 16560
gattcccacg gagcccatgg aaacgcagac cgagcccatg atcaagccca gcaccagcac 16620
catggaggtg cagacggatc cctggatgcc atcggctcct agtcgaagac cccggcgcaa 16680
gtacggcgcg gccagcctgc tgatgcccaa ctacgcgctg catccttcca tcatccccac 16740
gccgggctac cgcggcacgc gcttctaccg cggtcatacc agcagccgcc gccgcaagac 16800

caccactcgc cgccgccgtc gccgcaccgc cgctgcaacc acccctgccg ccctggtgcg 16860
gagagtgtac cgccgcggcc gcgcacctct gaccctgccg cgcgcgcgct accacccgag 16920
catcgccatt taaactttcg cctgctttgc agatcaatgg ccctcacatg ccgccttcgc 16980
gttcccatta cgggctaccg aggaagaaaa ccgcgccgta gaaggctggc ggggaacggg 17040
atgcgtcgcc accaccaccg gcggcggcgc gccatcagca agcggttggg gggaggcttc 17100
ctgcccgcgc tgatccccat catcgccgcg gcgatcgggg cgatccccgg cattgcttcc 17160
gtggcggtgc aggcctctca gcgccactga gacacacttg gaaacatctt gtaataaacc 17220

aatggactct gacgctcctg gtcctgtgat gtgttttcgt agacagatgg aagacatcaa 17280

tttttcgtcc ctggctccgc gacacggcac gcggccgttc atgggcacct ggagcgacat 17340

cggcaccagc caactgaacg ggggcgcctt caattggagc agtctctgga gcgggcttaa 17400

gaatttcggg tccacgctta aaacctatgg cagcaaggcg tggaacagca ccacagggca 17460

ggcgctgagg gataagctga aagagcagaa cttccagcag aaggtggtcg atgggctcgc 17520

ctcgggcatc aacggggtgg tggacctggc caaccaggcc gtgcagcggc agatcaacag 17580

ccgcctggac ccggtgccgc ccgccggctc cgtggagatg ccgcaggtgg aggaggagct 17640

gcctcccctg gacaagcggg gcgagaagcg accccgcccc gatgcggagg agacgctgct 17700

gacgcacacg gacgagccgc ccccgtacga ggaggcggtg aaactgggtc tgcccaccac 17760

gcggcccatc gcgcccctgg ccaccggggt gctgaaaccc gaaaagcccg cgaccctgga 17820

cttgcctcct ccccagcctt cccgcccctc tacagtggct aagcccctgc cgccggtggc 17880

cgtggcccgc gcgcgacccg ggggcaccgc ccgccctcat gcgaactggc agagcactct 17940

gaacagcatc gtgggtctgg gagtgcagag tgtgaagcgc cgccgctgct attaaaccta 18000

ccgtagcgct taacttgctt gtctgtgtgt gtatgtatta tgtcgccgcc gccgctgtcc 18060

accagaagga ggagtgaaga ggcgcgtcgc cgagttgcaa gatggccacc ccatcgatgc 18120

tgccccagtg ggcgtacatg cacatcgccg gacaggacgc ttcggagtac ctgagtccgg 18180

gtctggtgca gtttgcccgc gccacagaca cctacttcag tctggggaac aagtttagga 18240

accccacggt ggcgcccacg cacgatgtga ccaccgaccg cagccagcgg ctgacgctgc 18300

gcttcgtgcc cgtggaccgc gaggacaaca cctactcgta caaagtgcgc tacacgctgg 18360

ccgtgggcga caaccgcgtg ctggacatgg ccagcaccta ctttgacatc cgcggcgtgc 18420

tggatcgggg ccctagcttc aaaccctact ccggcaccgc ctacaacagt ctggccccca 18480

agggagcacc caacacttgt cagtggacat ataaagccga tggtgaaact gccacagaaa 18540

aaacctatac atatggaaat gcacccgtgc agggcattaa catcacaaaa gatggtattc 18600

aacttggaac tgacaccgat gatcagccaa tctacgcaga taaaacctat cagcctgaac 18660

ctcaagtggg tgatgctgaa tggcatgaca tcactggtac tgatgaaaag tatggaggca 18720

gagctcttaa gcctgatacc aaaatgaagc cttgttatgg ttcttttgcc aagcctacta 18780

ataaagaagg aggtcaggca aatgtgaaaa caggaacagg cactactaaa gaatatgaca 18840

tagacatggc tttctttgac aacagaagtg cggctgctgc tggcctagct ccagaaattg 18900

ttttgtatac tgaaaatgtg gatttggaaa ctccagatac ccatattgta tacaaagcag 18960

gcacagatga cagcagctct tctattaatt tgggtcagca agccatgccc aacagaccta 19020

actacattgg tttcagagac aactttatcg ggctcatgta ctacaacagc actggcaata 19080

tgggggtgct ggccggtcag gcttctcagc tgaatgctgt ggttgacttg caagacagaa 19140

acaccgagct gtcctaccag ctcttgcttg actctctggg tgacagaacc cggtatttca 19200

gtatgtggaa tcaggcggtg gacagctatg atcctgatgt gcgcattatt gaaaatcatg 19260

gtgtggagga tgaacttccc aactattgtt tccctctgga tgctgttggc agaacagata 19320

cttatcaggg aattaaggct aatggaactg atcaaaccac atggaccaaa gatgacagtg 19380

tcaatgatgc taatgagata ggcaagggta atccattcgc catggaaatc aacatccaag 19440

ccaacctgtg gaggaacttc ctctacgcca acgtggccct gtacctgccc gactcttaca 19500

agtacacgcc ggccaatgtt accctgccca ccaacaccaa cacctacgat tacatgaacg 19560

gccgggtggt ggcgccctcg ctggtggact cctacatcaa catcggggcg cgctggtcgc 19620

tggatcccat ggacaacgtg aaccccttca accaccaccg caatgcgggg ctgcgctacc 19680

gctccatgct cctgggcaac gggcgctacg tgcccttcca catccaggtg ccccagaaat 19740

ttttcgccat caagagcctc ctgctcctgc ccgggtccta cacctacgag tggaacttcc 19800

gcaaggacgt caacatgatc ctgcagagct ccctcggcaa cgacctgcgc acggacgggg 19860

cctccatctc cttcaccagc atcaacctct acgccacctt cttccccatg gcgcacaaca 19920

cggcctccac gctcgaggcc atgctgcgca acgacaccaa cgaccagtcc ttcaacgact 19980

acctctcggc ggccaacatg ctctacccca tcccggccaa cgccaccaac gtgcccatct 20040

ccatcccctc gcgcaactgg gccgccttcc gcggctggtc cttcacgcgt ctcaagacca 20100

aggagacgcc ctcgctgggc tccgggttcg acccctactt cgtctactcg ggctccatcc 20160

cctacctcga cggcaccttc tacctcaacc acaccttcaa gaaggtctcc atcaccttcg 20220

actcctccgt cagctggccc ggcaacgacc ggctcctgac gcccaacgag ttcgaaatca 20280

agcgcaccgt cgacggcgag ggctacaacg tggcccagtg caacatgacc aaggactggt 20340

tcctggtcca gatgctggcc cactacaaca tcggctacca gggcttctac gtgcccgagg 20400

gctacaagga ccgcatgtac tccttcttcc gcaacttcca gcccatgagc cgccaggtgg 20460

tggacgaggt caactacaag gactaccagg ccgtcaccct ggcctaccag cacaacaact 20520

cgggcttcgt cggctacctc gcgcccacca tgcgccaggg ccagccctac cccgccaact 20580

acccctaccc gctcatcggc aagagcgccg tcaccagcgt cacccagaaa aagttcctct 20640

gcgacagggt catgtggcgc atccccttct ccagcaactt catgtccatg ggcgcgctca 20700

ccgacctcgg ccagaacatg ctctatgcca actccgccca cgcgctagac atgaatttcg 20760

aagtcgaccc catggatgag tccacccttc tctatgttgt cttcgaagtc ttcgacgtcg 20820

tccgagtgca ccagccccac cgcggcgtca tcgaggccgt ctacctgcgc acccccttct 20880

cggccggtaa cgccaccacc taagctcttg cttcttgcaa gccatggccg cgggctccgg 20940

cgagcaggag ctcagggcca tcatccgcga cctgggctgc gggccctact tcctgggcac 21000

cttcgataag cgcttcccgg gattcatggc cccgcacaag ctggcctgcg ccatcgtcaa 21060
cacggccggc cgcgagaccg ggggcgagca ctggctggcc ttcgcctgga acccgcgctc 21120
gaacacctgc tacctcttcg acccccttcgg gttctcggac gagcgcctca agcagatcta 21180
ccagttcgag tacgagggcc tgctgcgccg cagcgccctg gccaccgagg accgctgcgt 21240
caccctggaa aagtccaccc agaccgtgca gggtccgcgc tcggccgcct gcgggctctt 21300
ctgctgcatg ttcctgcacg ccttcgtgca ctggcccgac cgccccatgg acaagaaccc 21360
caccatgaac ttgctgacgg gggtgcccaa cggcatgctc cagtcgcccc aggtggaacc 21420

caccctgcgc cgcaaccagg aggcgctcta ccgcttcctc aactcccact ccgcctactt 21480
tcgctcccac cgcgcgcgca tcgagaaggc caccgccttc gaccgcatga atcaagacat 21540
gtaaaccgtg tgtgtatgtt aaatgtcttt aataaacagc actttcatgt tacacatgca 21600
tctgagatga tttatttaga aatcgaaagg gttctgccgg gtctcggcat ggcccgcggg 21660
cagggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt 21720
gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cgcgtcagtt gcagggcgcc 21780
cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga 21840

gttgcggtac acggggttgc agcactggaa caccatcagg gccgggtgct tcacgctcgc 21900
cagcaccgtc gcgtcggtga tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa 21960
gggggtcatc ttgcaggtct gccttcccat ggtgggcacg cacccgggct tgtggttgca 22020
atcgcagtgc agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat 22080
ggccttcatg aaagcctcca attgcctgaa cgcctgctgg gccttggctc cctcggtgaa 22140
gaagaccccg caggacttgc tagagaactg gttggtggcg cacccggcgt cgtgcacgca 22200
gcagcgcgcg tcgttgttgg ccagctgcac cacgctgcgc ccccagcggt tctgggtgat 22260

cttggcccgg tcggggttct ccttcagcgc gcgctgcccg ttctcgctcg ccacatccat 22320
ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc aggcaccgca gcttgccctc 22380
ggcctcggtg cacccgtgca gccacagcgc gcacccggtg cactcccagt tcttgtgggc 22440
gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt 22500
cttgttgcta gtgaaggtca gcggaatgcc gcggtgctcc tcgttgatgt acaggtggca 22560
gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcggt 22620
ctccacgcgg tagcggtcca tcagcatagt catgatttcc ataccttct cccaggccga 22680

gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag 22740

ggggtcgctc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac 22800

cgggggggtag ctgaagccca cggccgccag ctcctcctcg gcctgtcttt cgtcctcgct 22860

gtcctggctg acgtcctgca ggaccacatg cttggtcttg cggggtttct tcttgggcgg 22920

cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac 22980

tatctcttcc tcttcttggt ccgaggccac gcggcggtag gtatgtctct tcgggggcag 23040

aggcggaggc gacgggctct cgccgccgcg acttggcgga tggctggcag agccccttcc 23100

gcgttcgggg gtgcgctccc ggcggcgctc tgactgactt cctccgcggc cggccattgt 23160

gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc 23220

ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgcccag 23280

ccccgccacc tccgacgcgg ccgtcccaga catgcaagag atggaggaat ccatcgagat 23340

tgacctgggc tatgtgacgc ccgcggagca cgaggaggag ctggcagtgc gcttttcaca 23400

agaagagata caccaagaac agccagagca ggaagcagag aatgagcaga gtcaggctgg 23460

gctcgagcat gacggcgact acctccacct gagcgggggg gaggacgcgc tcatcaagca 23520

tctggcccgg caggccacca tcgtcaagga tgcgctgctc gaccgcaccg aggtgcccct 23580

cagcgtggag gagctcagcc gcgcctacga gttgaacctc ttctcgccgc gcgtgccccc 23640

caagcgccag cccaatggca cctgcgagcc caacccgcgc ctcaacttct acccggtctt 23700

cgcggtgccc gaggccctgg ccacctacca catctttttc aagaaccaaa agatccccgt 23760

ctcctgccgc gccaaccgca cccgcgccga cgccctttc aacctgggtc ccggcgcccg 23820

cctacctgat atcgcctcct tggaagaggt tcccaagatc ttcgagggtc tgggcagcga 23880

cgagactcgg gccgcgaacg ctctgcaagg agaaggagga gagcatgagc accacagcgc 23940

cctggtcgag ttggaaggcg acaacgcgcg gctggcggtg ctcaaacgca cggtcgagct 24000

gacccatttc gcctacccgg ctctgaacct gccccccaaa gtcatgagcg cggtcatgga 24060

ccaggtgctc atcaagcgcg cgtcgcccat ctccgaggac gagggcatgc aagactccga 24120

ggagggcaag cccgtggtca gcgacgagca gctggcccgg tggctgggtc ctaatgctag 24180

tccccagagt ttggaagagc ggcgcaaact catgatggcc gtggtcctgg tgaccgtgga 24240

gctggagtgc ctgcgccgct tcttcgccga cgcggagacc ctgcgcaagg tcgaggagaa 24300

cctgcactac ctcttcaggc acgggttcgt gcgccaggcc tgcaagatct ccaacgtgga 24360

gctgaccaac ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt 24420

gctgcacacc accctgcgcg gggaggcccg gcgcgactac atccgcgact gcgtctacct 24480

ctacctctgc cacacctggc agacgggcat gggcgtgtgg cagcagtgtc tggaggagca 24540

gaacctgaaa gagctctgca agctcctgca gaagaacctc aagggtctgt ggaccgggtt 24600

cgacgagcgc accaccgcct cggacctggc cgacctcatt ttccccgagc gcctcaggct 24660

gacgctgcgc aacggcctgc ccgactttat gagccaaagc atgttgcaaa actttcgctc 24720

tttcatcctc gaacgctccg gaatcctgcc cgccacctgc tccgcgctgc cctcggactt 24780

cgtgccgctg accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg 24840

cctggccaac tacctggcct accactcgga cgtgatcgag gacgtcagcg gcgagggcct 24900

gctcgagtgc cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc 24960

ccagctgctg agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccagcgaagg 25020

cgagggttca gccgccaagg ggggtctgaa actcaccccg gggctgtgga cctcggccta 25080

cttgcgcaag ttcgtgcccg aggactacca tcccttcgag atcaggttct acgaggacca 25140

atcccatccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg cgatcctggc 25200

ccaattgcaa gccatccaga aatcccgcca agaattcttg ctgaaaaagg gccgcggggt 25260

ctacctcgac ccccagaccg gtgaggagct caaccccggc ttcccccagg atgccccgag 25320

gaaacaagaa gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga 25380

acagcagtca ggcagaggag gaggagatgg aggaagactg ggacagcact caggcagagg 25440

aggacagcct gcaagacagt ctggaggaag acgaggagga ggcagaggag gaggtggaag 25500

aagcagccgc cgccagaccg tcgtcctcgg cgggggagaa agcaagcagc acggatacca 25560

tctccgctcc gggtcggggt cccgctcgac cacacagtag atgggacgag accggacgat 25620

tcccgaaccc caccacccag accggtaaga aggagcggca gggatacaag tcctggcggg 25680

ggcacaaaaa cgccatcgtc tcctgcttgc aggcctgcgg gggcaacatc tccttcaccc 25740

ggcgctacct gctcttccac cgcggggtga actttccccg caacatcttg cattactacc 25800

gtcacctcca cagcccctac tacttccaag aagaggcagc agcagcagaa aaagaccagc 25860

agaaaaccag cagctagaaa atccacagcg gcggcagcag gtggactgag gatcgcggcg 25920

aacgagccgg cgcaaacccg ggagctgagg aaccggatct ttcccaccct ctatgccatc 25980

ttccagcaga gtcgggggca ggagcaggaa ctgaaagtca agaaccgttc tctgcgctcg 26040

ctcacccgca gttgtctgta tcacaagagc gaagaccaac ttcagcgcac tctcgaggac 26100

gccgaggctc tcttcaacaa gtactgcgcg ctcactctta aagagtagcc cgcgcccgcc 26160

cagtcgcaga aaaaggcggg aattacgtca cctgtgccct tcgccctagc cgcctccacc 26220

catcatcatg agcaaagaga ttcccacgcc ttacatgtgg agctaccagc cccagatggg 26280

cctggccgcc ggtgccgccc aggactactc cacccgcatg aattggctca gcgccgggcc 26340

cgcgatgatc tcacgggtga atgacatccg cgcccaccga aaccagatac tcctagaaca 26400

gtcagcgctc accgccacgc cccgcaatca cctcaatccg cgtaattggc ccgccgccct 26460

ggtgtaccag gaaattcccc agcccacgac cgtactactt ccgcgagacg cccaggccga 26520
agtccagctg actaactcag gtgtccagct ggcgggcggc gccaccctgt gtcgtcaccg 26580
ccccgctcag ggtataaagc ggctggtgat ccggggcaga ggcacacagc tcaacgacga 26640
ggtggtgagc tcttcgctgg gtctgcgacc tgacggagtc ttccaactcg ccggatcggg 26700
gagatcttcc ttcacgcctc gtcaggccgt cctgactttg gagagttcgt cctcgcagcc 26760
ccgctcgggt ggcatcggca ctctccagtt cgtggaggag ttcactccct cggtctactt 26820
caaccccttc tccggctccc ccggccacta cccggacgag ttcatcccga acttcgacgc 26880

catcagcgag tcggtggacg gctacgattg aaactaatca ccccccttatc cagtgaaata 26940
aagatcatat tgatgatgat tttacagaaa taaaaaataa tcatttgatt tgaaataaag 27000
atacaatcat attgatgatt tgagtttaac aaaaaaataa agaatcactt acttgaaatc 27060
tgataccagg tctctgtcca tgttttctgc caacaccact tcactcccct cttcccagct 27120
ctggtactgc aggccccggc gggctgcaaa cttcctccac acgctgaagg ggatgtcaaa 27180
ttcctcctgt ccctcaatct tcattttatc ttctatcaga tgtccaaaaa gcgcgtccgg 27240
gtggatgatg acttcgaccc cgtctacccc tacgatgcag acaacgcacc gaccgtgccc 27300

ttcatcaacc ccccttcgt ctcttcagat ggattccaag agaagcccct gggggtgttg 27360
tccctgcgac tggccgaccc cgtcaccacc aagaacgggg aaatcaccct caagctggga 27420
gagggggtgg acctcgattc ctcgggaaaa ctcatctcca acacggccac caaggccgcc 27480
gcccctctca gtttttccaa caacaccatt tcccttaaca tggatcaccc cttttacact 27540
aaagatggaa aattatcctt acaagtttct ccaccattaa atatactgag aacaagcatt 27600
ctaaacacac tagctttagg ttttggatca ggtttaggac tccgtggctc tgccttggca 27660
gtacagttag tctctccact tacatttgat actgatggaa acataaagct taccttagac 27720

agaggtttgc atgttacaac aggagatgca attgaaagca acataagctg ggctaaaggt 27780
ttaaaatttg aagatggagc catagcaacc aacattggaa atgggttaga gtttggaagc 27840
agtagtacag aaacaggtgt tgatgatgct tacccaatcc aagttaaact tggatctggc 27900
cttagctttg acagtacagg agccataatg gctggtaaca aagaagacga taaactcact 27960
ttgtggacaa cacctgatcc atcaccaaac tgtcaaatac tcgcagaaaa tgatgcaaaa 28020
ctaacacttt gcttgactaa atgtggtagt caaatactgg ccactgtgtc agtcttagtt 28080
gtaggaagtg gaaacctaaa ccccattact ggcaccgtaa gcagtgctca ggtgtttcta 28140

cgttttgatg caaacggtgt tcttttaaca gaacattcta cactaaaaaa atactggggg 28200

tataggcagg gagatagcat agatggcact ccatatacca atgctgtagg attcatgccc 28260

aatttaaaag cttatccaaa gtcacaaagt tctactacta aaaataatat agtagggcaa 28320

gtatacatga atggagatgt ttcaaaacct atgcttctca ctataaccct caatggtact 28380

gatgacagca acagtacata ttcaatgtca ttttcataca cctggactaa tggaagctat 28440

gttggagcaa catttggggc taactcttat accttctcat acatcgccca agaatgaaca 28500

ctgtatccca ccctgcatgc caacccttcc caccccactc tgtggaacaa actctgaaac 28560

acaaaataaa ataaagttca agtgttttat tgattcaaca gttttacagg attcgagcag 28620

ttatttttcc tccaccctcc caggacatgg aatacaccac cctctccccc cgcacagcct 28680

tgaacatctg aatgccattg gtgatggaca tgcttttggt ctccacgttc cacacagttt 28740

cagagcgagc cagtctcggg tcggtcaggg agatgaaacc ctccgggcac tcccgcatct 28800

gcacctcaca gctcaacagc tgaggattgt cctcggtggt cgggatcacg gttatctgga 28860

agaagcagaa gagcggcggt gggaatcata gtccgcgaac gggatcggcc ggtggtgtcg 28920

catcaggccc cgcagcagtc gctgccgccg ccgctccgtc aagctgctgc tcagggggtc 28980

cgggtccagg gactccctca gcatgatgcc cacggccctc agcatcagtc gtctggtgcg 29040

gcgggcgcag cagcgcatgc ggatctcgct caggtcgctg cagtacgtgc aacacagaac 29100

caccaggttg ttcaacagtc catagttcaa cacgctccag ccgaaactca tcgcgggaag 29160

gatgctaccc acgtggccgt cgtaccagat cctcaggtaa atcaagtggt gccccctcca 29220

gaacacgctg cccacgtaca tgatctcctt gggcatgtgg cggttcacca cctcccggta 29280

ccacatcacc ctctggttga acatgcagcc ccggatgatc ctgcggaacc acagggccag 29340

caccgccccg cccgccatgc agcgaagaga ccccgggtcc cggcaatggc aatggaggac 29400

ccaccgctcg tacccgtgga tcatctggga gctgaacaag tctatgttgg cacagcacag 29460

gcatatgctc atgcatctct tcagcactct caactcctcg ggggtcaaaa ccatatccca 29520

gggcacgggg aactcttgca ggacagcgaa ccccgcagaa cagggcaatc ctcgcacaga 29580

acttacattg tgcatggaca gggtatcgca atcaggcagc accgggtgat cctccaccag 29640

agaagcgcgg gtctcggtct cctcacagcg tggtaagggg gccggccgat acgggtgatg 29700

gcgggacgcg gctgatcgtg ttcgcgaccg tgtcatgatg cagttgcttt cggacatttt 29760

cgtacttgct gtagcagaac ctggtccggg cgctgcacac cgatcgccgg cggcggtctc 29820

ggcgcttgga acgctcggtg ttgaaattgt aaaacagcca ctctctcaga ccgtgcagca 29880

gatctagggc ctcaggagtg atgaagatcc catcatgcct gatggctctg atcacatcga 29940

ccaccgtgga atgggccaga cccagccaga tgatgcaatt ttgttgggtt tcggtgacgg 30000

cgggggaggg aagaacagga agaaccatga ttaactttta atccaaacgg tctcggagta 30060

cttcaaaatg aagatcgcgg agatggcacc tctcgccccc gctgtgttgg tggaaaataa 30120

cagccaggtc aaaggtgata cggttctcga gatgttccac ggtggcttcc agcaaagcct 30180

ccacgcgcac atccagaaac aagacaatag cgaaagcggg agggttctct aattcctcaa 30240

tcatcatgtt acactcctgc accatcccca gataattttc atttttccag ccttgaatga 30300

ttcgaactag ttcctgaggt aaatccaagc cagccatgat aaagagctcg cgcagagcgc 30360

cctccaccgg cattcttaag cacaccctca taattccaag atattctgct cctggttcac 30420

ctgcagcaga ttgacaagcg gaatatcaaa atctctgccg cgatccctga gctcctccct 30480

cagcaataac tgtaagtact ctttcatatc ctctccgaaa tttttagcca taggaccacc 30540

aggaataaga ttagggcaag ccacagtaca gataaaccga agtcctcccc agtgagcatt 30600

gccaaatgca agactgctat aagcatgctg gctagacccg gtgatatctt ccagataact 30660

ggacagaaaa tcgcccaggc aatttttaag aaaatcaaca aaagaaaaat cctccaggtg 30720

gacgtttaga gcctcgggaa caacgatgaa gtaaatgcaa gcggtgcgtt ccagcatggt 30780

tagttagctg atctgtagaa aaaacaaaaa tgaacattaa accatgctag cctggcgaac 30840

aggtgggtaa atcgttctct ccagcaccag gcaggccacg gggtctccgg cgcgaccctc 30900

gtaaaaattg tcgctatgat tgaaaaccat cacagagaga cgttcccggt ggccggcgtg 30960

aatgattcga caagatgaat acacccccgg aacattggcg tccgcgagtg aaaaaaagcg 31020

cccgaggaag caataaggca ctacaatgct cagtctcaag tccagcaaag cgatgccatg 31080

cggatgaagc acaaaattct caggtgcgta caaaatgtaa ttactcccct cctgcacagg 31140

cagcaaagcc cccgatccct ccaggtacac atacaaagcc tcagcgtcca tagcttaccg 31200

agcagcagca cacaacaggc gcaagagtca gagaaaggct gagctctaac ctgtccaccc 31260

gctctctgct caatatatag cccagatcta cactgacgta aaggccaaag tctaaaaata 31320

cccgccaaat aatcacacac gcccagcaca cgcccagaaa ccggtgacac actcaaaaaa 31380

atacgcgcac ttcctcaaac gcccaaaact gccgtcattt ccgggttccc acgctacgtc 31440

atcaaaacac gactttcaaa ttccgtcgac cgttaaaaac gtcacccgcc ccgcccctaa 31500

cggtcgcccg tctctcagcc aatcagcgcc ccgcatcccc aaattcaaac acctcatttg 31560

catattaacg cgcacaaaaa gtttgagg 31588

<210> 3

<211> 11447

<212> DNA

<213> Venezuelan equine encephalitis virus

<400> 3

atgggcggcg catgagagaa gcccagacca ttacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccaccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960
acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320
acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380
acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560
atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620
caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740
tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980
caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160
aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220
aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa 7560

gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc 7620

ggccccgcgc aggccctggt tccccagaac cgacccttttt ctggcgatgc aggtgcagga 7680

attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg 7740

gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg 7800

gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc 7860

acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat 7920

gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc 7980

ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga 8040

cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt 8100

gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta 8160

cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt 8220

tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat 8280

tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga 8340

gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac 8400

catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag 8460

aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga 8520

gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt 8580

taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag 8640

ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag 8700

acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat 8760

gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac 8820

atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc 8880

aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc 8940

gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg 9000

agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga 9060

gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt 9120
caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa 9180
gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg 9240
cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc 9300
agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg 9360
caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa 9420
actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta 9480

cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg 9540
ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg 9600
aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc 9660
caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac 9720
ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc 9780
taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac 9840
ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct 9900

gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt 9960
gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat 10020
gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact 10080
ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt 10140
cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga 10200
atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt 10260
catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta 10320

cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc 10380
ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta 10440
tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc 10500
cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta 10560
tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac 10620
agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg 10680
agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa 10740

agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg 10800

cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt 10860

caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt 10920

gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa 10980

gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac 11040

cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct 11100

ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat 11160

tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg 11220

gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct 11280

ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca 11340

attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt 11400

ttatttttttc ttttcttttc cgaatcggat tttgttttta atatttc 11447

<210> 4

<211> 9577

<212> DNA

<213> Artificial Sequence

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

<400> 4

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900
tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960
cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080
tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140
tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200
tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320
acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380
atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440
caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560
tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620
tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740
ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800
taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860
tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920
ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980
gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040
aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160
cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220
gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280
aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340
ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400
ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460
ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360
cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480
aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260
gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500
gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560

ttaagccacc atggcaggca tgtttcaggc gctgagcgaa ggctgcaccc cgtatgatat 7620
taaccagatg ctgaacgtgc tgggcgatca tcaggtctca ggccttgagc agcttgagag 7680
tataatcaac tttgaaaaac tgactgaatg gaccagttct aatgttatgc ctatcctgtc 7740
tcctctgaca aagggcatcc tgggcttcgt gtttaccctg accgtgcctt ctgagagagg 7800
acttagctgc attagcgaag cggatgcgac caccccggaa agcgcgaacc tgggcgaaga 7860
aattctgagc cagctgtatc tttggccaag ggtgacctac cattccccta gttatgctta 7920
ccaccaattt gaaagacgag ccaaatataa aagacacttc cccggctttg gccagagcct 7980

gctgtttggc taccctgtgt acgtgttcgg cgattgcgtg cagggcgatt gggatgcgat 8040

tcgctttcgc tattgcgcgc cgccgggcta tgcgctgctg cgctgcaacg ataccaacta 8100

tagcgctctg ctggctgtgg gggcccctaga aggacccagg aatcaggact ggcttggtgt 8160

cccaagacaa cttgtaactc ggatgcaggc tattcagaat gccggcctgt gtaccctggt 8220

ggccatgctg gaagagacaa tcttctggct gcaagcgttt ctgatggcgc tgaccgatag 8280

cggcccgaaa accaacatta ttgtggatag ccagtatgtg atgggcatta gcaaaccgag 8340

ctttcaggaa tttgtggatt gggaaaacgt gagcccggaa ctgaacagca ccgatcagcc 8400

gttttggcaa gccggaatcc tggccagaaa tctggtgcct atggtggcca cagtgcaggg 8460

ccagaacctg aagtaccagg gtcagtcact agtcatctct gcttctatca ttgtcttcaa 8520

cctgctggaa ctggaaggtg attatcgaga tgatggcaac gtgtgggtgc ataccccgct 8580

gagcccgcgc accctgaacg cgtgggtgaa agcggtggaa gaaaaaaaag gtattccagt 8640

tcacctagag ctggccagta tgaccaacat ggagctcatg agcagtattg tgcatcagca 8700

ggtcagaaca tacggccccg tgttcatgtg tctcggcgga ctgcttacaa tggtggctgg 8760

tgctgtgtgg ctgacagtgc gagtgctcga gctgttccgg gccgcgcagc tggccaacga 8820

cgtggtcctc cagatcatgg agctttgtgg tgcagcgttt cgccaggtgt gccataccac 8880

cgtgccgtgg ccgaacgcga gcctgacccc gaaatggaac aacgaaacca cccagcccca 8940

gatcgccaac tgcagcgtgt atgacttttt tgtgtggctc cattattatt ctgttcgaga 9000

cacactttgg ccaagggtga cctaccatat gaacaaatat gcgtatcata tgctggaaag 9060

acgagccaaa tataaaagag gaccaggacc tggcgctaaa tttgtggccg cctggacact 9120

gaaagccgct gctggtcctg gacctggcca gtacatcaag gccaacagca agttcatcgg 9180

catcaccgaa ctcggacccg gaccaggctg atgattcgaa cggccgtatc acgcccaaac 9240

atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta acatccctgc tgggaggatc 9300

agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac 9360

caaccagaaa cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg 9420

cgattggcat gccgccttaa aatttttatt ttatttttc ttttcttttc cgaatcggat 9480

tttgttttta atatttcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 9540

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 9577

<210> 5

<211> 11447

<212> DNA

<213> Venezuelan equine encephalitis virus

<400> 5

atgggcggcg catgagagaa gcccagacca ttacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccaccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520
caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640
ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700
catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760
cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940
ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000
tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060
cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120
ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180
ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240
ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360
cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480
aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540
taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600
aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660
cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780
tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa 7560

gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc 7620

ggccccgcgc aggccctggt tccccagaac cgacccttttt ctggcgatgc aggtgcagga 7680

attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg 7740

gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg 7800

gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc 7860

acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat 7920

gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc 7980

ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga 8040

cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt 8100

gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta 8160

cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt 8220

tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat 8280

tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga 8340

gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac 8400

catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag 8460

aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga 8520

gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt 8580

taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag 8640

ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag 8700

acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat 8760

gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac 8820

atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc 8880

aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc 8940

gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg 9000

agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga 9060

gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt 9120

caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa 9180

gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg 9240

cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc 9300

agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg 9360

caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa 9420

actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta 9480

cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg 9540

ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg 9600

aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc 9660

caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac 9720

ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc 9780

taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac 9840

ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct 9900

gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt 9960

gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat 10020

gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact 10080

ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt 10140

cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga 10200

atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt 10260

catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta 10320

cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc 10380

ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta 10440

tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc 10500

cacagcttgg acacccttta atcgcaaaat cgtgcagtat gccggggaga tctataatta 10560

tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac 10620

agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg 10680

agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa 10740

agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg 10800

cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt 10860

caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt 10920

gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa 10980

gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac 11040

cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct 11100

ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat 11160

tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg 11220

gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct 11280

ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca 11340

attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt 11400

ttatttttttc ttttcttttc cgaatcggat tttgttttta atatttc 11447

<210> 6

<211> 7894

<212> DNA

<213> Artificial Sequence

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

<400> 6

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760
aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820
ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880
tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940
taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000
cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060
agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240
gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300
cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360
cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420
gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480
tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900
cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320
acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt 7560

tacagccgcg gtgtcaaaaa ccgcgtggac gtggttaaca tccctgctgg gaggatcagc 7620

cgtaattatt ataattggct tggtgctggc tactattgtg gccatgtacg tgctgaccaa 7680

ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctcgcggcga 7740

ttggcatgcc gccttaaaat ttttatttta ttttttcttt tcttttccga atcggatttt 7800

gtttttaata tttcaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 7894

<210> 7

<211> 7893

<212> DNA

<213> Artificial Sequence

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

<400> 7

ataggcggcg catgagagaa gcccagacca ttacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgatggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540
gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660
tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720
agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780
tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840
gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960
acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080
gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140
tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200
tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380
acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440
cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500
ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560
atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620
caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680
atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800
aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860
gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980
caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040
acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100
cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt 7560

tacagccgcg gtgtcaaaaa ccgcgtggac gtggttaaca tccctgctgg gaggatcagc 7620

cgtaattatt ataattggct tggtgctggc tactattgtg gccatgtacg tgctgaccaa 7680

ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctcgcggcga 7740

ttggcatgcc gccttaaaat ttttatttta ttttttcttt cttttccgaa tcggattttg 7800

tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 7893

<210> 8

<211> 7927

<212> DNA

<213> Artificial Sequence

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

<400> 8

taatacgact cactatagga tgggcggcgc atgagagaag cccagaccaa ttacctaccc 60

aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag 120

cggagcttcc cgcagtttga ggtagaagcc aagcaggtca ctgataatga ccatgctaat 180

gccagagcgt tttcgcatct ggcttcaaaa ctgatcgaaa cggaggtgga cccatccgac 240

acgatccttg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat 300

tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag 360

ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc 420

gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag 480

tcgtgtcgct acgaagggca agtcgctgtt taccaggatg tatacgcggt tgacggaccg 540

acaagtctct atcaccaagc caataaggga gttagagtcg cctactggat aggctttgac 600

accacccctt ttatgtttaa gaacttggct ggagcatatc catcatactc taccaactgg 660

gccgacgaaa ccgtgttaac ggctcgtaac ataggcctat gcagctctga cgttatggag 720

cggtcacgta gagggatgtc cattcttaga aagaagtatt tgaaaccatc caacaatgtt 780

ctattctctg ttggctcgac catctaccac gagaagaggg acttactgag gagctggcac 840

ctgccgtctg tatttcactt acgtggcaag caaaattaca catgtcggtg tgagactata 900

gttagttgcg acgggtacgt cgttaaaaga atagctatca gtccaggcct gtatgggaag 960

ccttcaggct atgctgctac gatgcaccgc gagggattct tgtgctgcaa agtgacagac 1020

acattgaacg gggagagggt ctcttttccc gtgtgcacgt atgtgccagc tacattgtgt 1080

gaccaaatga ctggcatact ggcaacagat gtcagtgcgg acgacgcgca aaaactgctg 1140

gttgggctca accagcgtat agtcgtcaac ggtcgcaccc agagaaacac caataccatg 1200

aaaaattacc ttttgcccgt agtggcccag gcatttgcta ggtgggcaaa ggaatataag 1260

gaagatcaag aagatgaaag gccactagga ctacgagata gacagttagt catggggtgt 1320

tgttgggctt ttagaaggca caagataaca tctatttata agcgcccgga tacccaaacc 1380

atcatcaaag tgaacagcga tttccactca ttcgtgctgc ccaggatagg cagtaacaca 1440

ttggagatcg ggctgagaac aagaatcagg aaaatgttag aggagcacaa ggagccgtca 1500

cctctcatta ccgccgagga cgtacaagaa gctaagtgcg cagccgatga ggctaaggag 1560

gtgcgtgaag ccgaggagtt gcgcgcagct ctaccacctt tggcagctga tgttgaggag 1620

cccactctgg aagccgatgt cgacttgatg ttacaagagg ctggggccgg ctcagtggag 1680

acacctcgtg gcttgataaa ggttaccagc tacgctggcg aggacaagat cggctcttac 1740

gctgtgcttt ctccgcaggc tgtactcaag agtgaaaaat tatcttgcat ccaccctctc 1800

gctgaacaag tcatagtgat aacacactct ggccgaaaag ggcgttatgc cgtggaacca 1860

taccatggta aagtagtggt gccagaggga catgcaatac ccgtccagga ctttcaagct 1920

ctgagtgaaa gtgccaccat tgtgtacaac gaacgtgagt tcgtaaacag gtacctgcac 1980

catattgcca cacatggagg agcgctgaac actgatgaag aatattacaa aactgtcaag 2040

cccagcgagc acgacggcga atacctgtac gacatcgaca ggaaacagtg cgtcaagaaa 2100

gaactagtca ctgggctagg gctcacaggc gagctggtgg atcctccctt ccatgaattc 2160

gcctacgaga gtctgagaac acgaccagcc gctccttacc aagtaccaac catggggtg 2220

tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat 2280

ctagtggtga gcgccaagaa agaaaactgt gcagaaatta taagggacgt caagaaaatg 2340

aaagggctgg acgtcaatgc cagaactgtg gactcagtgc tcttgaatgg atgcaaacac 2400

cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg 2460

ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt 2520

ttttttaaca tgatgtgcct gaaagtgcat tttaaccacg agatttgcac acaagtcttc 2580

cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt 2640

tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc 2700

ggcagtacca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag 2760

cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggctg 2820

acccgtaaag gtgtgtatgc cgttcggtac aaggtgaatg aaaatcctct gtacgcaccc 2880

acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca 2940

ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc 3000

acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg 3060

gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgccg 3120

gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatttt 3180

gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt 3240

ggactcgatc tggactccgg tctattttct gcacccactg ttccgttatc cattaggaat 3300

aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtggtccgt 3360

cagctctctc gcaggtaccc acaactgcct cgggcagttg ccactggaag agtctatgac 3420

atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga 3480

agactgcctc atgctttagt cctccaccat aatgaacacc cacagagtga cttttcttca 3540

ttcgtcagca aattgaaggg cagaactgtc ctggtggtcg gggaaaagtt gtccgtccca 3600

ggcaaaatgg ttgactggtt gtcagaccgg cctgaggcta ccttcagagc tcggctggat 3660

ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggacccca 3720

tataaatacc atcactatca gcagtgtgaa gaccatgcca ttaagcttag catgttgacc 3780

aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac 3840

gctgacaggg ccagcgaaag catcattggt gctatagcgc ggcagttcaa gttttcccgg 3900

gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac 3960

gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat 4020

acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat 4080

attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc 4140

ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc 4200

gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga 4260

ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag 4320
tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc 4380
accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca 4440
gctttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg 4500
actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac 4560
tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct 4620
ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag 4680

tttcaccagg cggccaagga tatagcagaa attaatgcca tgtggcccgt tgcaacggag 4740
gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcgaaa 4800
tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc 4860
catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact 4920
gtgtgctcat cctttccatt gccgaagtat agaatcactg gtgtgcagaa gatccaatgc 4980
tcccagccta tattgttctc accgaaagtg cctgcgtata ttcatccaag gaagtatctc 5040
gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag 5100

gggacacctg aacaaccacc acttataacc gaggatgaga ccaggactag aacgcctgag 5160
ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc 5220
caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc 5280
tggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg 5340
gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag 5400
agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca 5460
catcccgctc cgcgcacaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc 5520

agcctagttt ccaccccgcc aggcgtgaat agggtgatca ctagagagga gctcgaggcg 5580
cttaccccgt cacgcactcc tagcaggtcg gtctcgagaa ccagcctggt ctccaacccg 5640
ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa 5700
tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacaa 5760
caaaaatcag taaggcaaac ggtgctatcc gaagtggtgt tggagaggac cgaattggag 5820
atttcgtatg ccccgcgcct cgaccaagaa aaagaagaat tactacgcaa gaaattacag 5880
ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaaggtgga gaacatgaaa 5940

gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa 6000

gtggagtgct accgaaccct gcatcctgtt cctttgtatt catctagtgt gaaccgtgcc 6060

ttttcaagcc ccaaggtcgc agtggaagcc tgtaacgcca tgttgaaaga gaactttccg 6120

actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga 6180

gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag 6240

aaacactcct atttggaacc cacaatacga tcggcagtgc cttcagcgat ccagaacacg 6300

ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa 6360

ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat 6420

aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta 6480

aattacatta ccaaattaaa aggaccaaaa gctgctgctc tttttgcgaa gacacataat 6540

ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg 6600

aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacaggt gatccaggct 6660

gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggt taggagatta 6720

aatgcggtcc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac 6780

gctattatag ccgagcactt ccagcctggg gattgtgttc tggaaactga catcgcgtcg 6840

tttgataaaa gtgaggacga cgccatggct ctgaccgcgt taatgattct ggaagactta 6900

ggtgtggacg cagagctgtt gacgctgatt gaggcggctt tcggcgaaat ttcatcaata 6960

catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc 7020

acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg 7080

ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa 7140

tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata 7200

gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatttt gtgtgactcc 7260

gtgaccggca cagcgtgccg tgtggcagac cccctaaaaa ggctgtttaa gcttggcaaa 7320

cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca 7380

acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat 7440

gaaaccgtag gaacttccat catagttatg gccatgacta ctctagctag cagtgttaaa 7500

tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg 7560

actatcacgc ccaaacattt acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat 7620

ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg 7680

ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct 7740

tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttattttat tttttctttt 7800

cttttccgaa tcggattttg tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaatacgtag 7920

tttaaac 7927

<210> 9

<211> 7926

<212> DNA

<213> Artificial Sequence

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

<400> 9

taatacgact cactatagga taggcggcgc atgagagaag cccagaccaa ttacctaccc 60

aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag 120

cggagcttcc cgcagtttga ggtagaagcc aagcaggtca ctgataatga ccatgctaat 180

gccagagcgt tttcgcatct ggcttcaaaa ctgatcgaaa cggaggtgga cccatccgac 240

acgatccttg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat 300

tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag 360

ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc 420

gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag 480

tcgtgtcgct acgaagggca agtcgctgtt taccaggatg tatacgcggt tgacggaccg 540

acaagtctct atcaccaagc caataaggga gttagagtcg cctactggat aggctttgac 600

accacccctt ttatgtttaa gaacttggct ggagcatatc catcatactc taccaactgg 660

gccgacgaaa ccgtgttaac ggctcgtaac ataggcctat gcagctctga cgttatggag 720

cggtcacgta gagggatgtc cattcttaga aagaagtatt tgaaaccatc caacaatgtt 780

ctattctctg ttggctcgac catctaccac gagaagaggg acttactgag gagctggcac 840

ctgccgtctg tatttcactt acgtggcaag caaaattaca catgtcggtg tgagactata 900

gttagttgcg acgggtacgt cgttaaaaga atagctatca gtccaggcct gtatgggaag 960

ccttcaggct atgctgctac gatgcaccgc gagggattct tgtgctgcaa agtgacagac 1020

acattgaacg gggagagggt ctcttttccc gtgtgcacgt atgtgccagc tacattgtgt 1080

gaccaaatga ctggcatact ggcaacagat gtcagtgcgg acgacgcgca aaaactgctg 1140

gttgggctca accagcgtat agtcgtcaac ggtcgcaccc agagaaacac caataccatg 1200

aaaaattacc ttttgcccgt agtggcccag gcatttgcta ggtgggcaaa ggaatataag 1260

gaagatcaag aagatgaaag gccactagga ctacgagata gacagttagt catggggtgt 1320

tgttgggctt ttagaaggca caagataaca tctatttata agcgcccgga tacccaaacc 1380

atcatcaaag tgaacagcga tttccactca ttcgtgctgc ccaggatagg cagtaacaca 1440

ttggagatcg ggctgagaac aagaatcagg aaaatgttag aggagcacaa ggagccgtca 1500

cctctcatta ccgccgagga cgtacaagaa gctaagtgcg cagccgatga ggctaaggag 1560

gtgcgtgaag ccgaggagtt gcgcgcagct ctaccacctt tggcagctga tgttgaggag 1620

cccactctgg aagccgatgt cgacttgatg ttacaagagg ctggggccgg ctcagtggag 1680

acacctcgtg gcttgataaa ggttaccagc tacgatggcg aggacaagat cggctcttac 1740

gctgtgcttt ctccgcaggc tgtactcaag agtgaaaaat tatcttgcat ccaccctctc 1800

gctgaacaag tcatagtgat aacacactct ggccgaaaag ggcgttatgc cgtggaacca 1860

taccatggta aagtagtggt gccagaggga catgcaatac ccgtccagga ctttcaagct 1920

ctgagtgaaa gtgccaccat tgtgtacaac gaacgtgagt tcgtaaacag gtacctgcac 1980

catattgcca cacatggagg agcgctgaac actgatgaag aatattacaa aactgtcaag 2040

cccagcgagc acgacggcga atacctgtac gacatcgaca ggaaacagtg cgtcaagaaa 2100

gaactagtca ctgggctagg gctcacaggc gagctggtgg atcctccctt ccatgaattc 2160

gcctacgaga gtctgagaac acgaccagcc gctccttacc aagtaccaac cataggggtg 2220

tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat 2280

ctagtggtga gcgccaagaa agaaaactgt gcagaaatta taagggacgt caagaaaatg 2340

aaagggctgg acgtcaatgc cagaactgtg gactcagtgc tcttgaatgg atgcaaacac 2400

cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg 2460

ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt 2520

ttttttaaca tgatgtgcct gaaagtgcat tttaaccacg agatttgcac acaagtcttc 2580

cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt 2640

tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc 2700

ggcagtacca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag 2760

cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggctg 2820

acccgtaaag gtgtgtatgc cgttcggtac aaggtgaatg aaaatcctct gtacgcaccc 2880

acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca 2940

ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc 3000

acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg 3060

gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgccg 3120

gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatttt 3180

gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt 3240

ggactcgatc tggactccgg tctattttct gcacccactg ttccgttatc cattaggaat 3300

aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtggtccgt 3360

cagctctctc gcaggtaccc acaactgcct cgggcagttg ccactggaag agtctatgac 3420

atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga 3480

agactgcctc atgctttagt cctccaccat aatgaacacc cacagagtga cttttcttca 3540

ttcgtcagca aattgaaggg cagaactgtc ctggtggtcg gggaaaagtt gtccgtccca 3600

ggcaaaatgg ttgactggtt gtcagaccgg cctgaggcta ccttcagagc tcggctggat 3660

ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggacccca 3720

tataaatacc atcactatca gcagtgtgaa gaccatgcca ttaagcttag catgttgacc 3780

aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac 3840

gctgacaggg ccagcgaaag catcattggt gctatagcgc ggcagttcaa gttttcccgg 3900

gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac 3960

gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat 4020

acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat 4080

attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc 4140

ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc 4200

gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga 4260

ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag 4320

tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc 4380

accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca 4440

gctttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg 4500

actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac 4560

tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct 4620

ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag 4680

tttcaccagg cggccaagga tatagcagaa attaatgcca tgtggcccgt tgcaacggag 4740

gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcgaaa 4800

tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc 4860

catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact 4920

gtgtgctcat cctttccatt gccgaagtat agaatcactg gtgtgcagaa gatccaatgc 4980

tcccagccta tattgttctc accgaaagtg cctgcgtata ttcatccaag gaagtatctc 5040
gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag 5100
gggacacctg aacaaccacc acttataacc gaggatgaga ccaggactag aacgcctgag 5160
ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc 5220
caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc 5280
tggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg 5340
gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag 5400

agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca 5460
catcccgctc cgcgcacaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc 5520
agcctagttt ccaccccgcc aggcgtgaat agggtgatca ctagagagga gctcgaggcg 5580
cttaccccgt cacgcactcc tagcaggtcg gtctcgagaa ccagcctggt ctccaacccg 5640
ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa 5700
tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacaa 5760
caaaaatcag taaggcaaac ggtgctatcc gaagtggtgt tggagaggac cgaattggag 5820

atttcgtatg ccccgcgcct cgaccaagaa aaagaagaat tactacgcaa gaaattacag 5880
ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaaggtgga gaacatgaaa 5940
gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa 6000
gtggagtgct accgaaccct gcatcctgtt cctttgtatt catctagtgt gaaccgtgcc 6060
ttttcaagcc ccaaggtcgc agtggaagcc tgtaacgcca tgttgaaaga gaactttccg 6120
actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga 6180
gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag 6240

aaacactcct atttggaacc cacaatacga tcggcagtgc cttcagcgat ccagaacacg 6300
ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa 6360
ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat 6420
aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta 6480
aattacatta ccaaattaaa aggaccaaaa gctgctgctc tttttgcgaa gacacataat 6540
ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg 6600
aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacaggt gatccaggct 6660

gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggt taggagatta 6720

aatgcggtcc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac 6780

gctattatag ccgagcactt ccagcctggg gattgtgttc tggaaactga catcgcgtcg 6840

tttgataaaa gtgaggacga cgccatggct ctgaccgcgt taatgattct ggaagactta 6900

ggtgtggacg cagagctgtt gacgctgatt gaggcggctt tcggcgaaat ttcatcaata 6960

catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc 7020

acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg 7080

ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa 7140

tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata 7200

gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatttt gtgtgactcc 7260

gtgaccggca cagcgtgccg tgtggcagac cccctaaaaa ggctgtttaa gcttggcaaa 7320

cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca 7380

acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat 7440

gaaaccgtag gaacttccat catagttatg gccatgacta ctctagctag cagtgttaaa 7500

tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg 7560

actatcacgc ccaaacattt acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat 7620

ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg 7680

ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct 7740

tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttattttat ttttcttttc 7800

ttttccgaat cggattttgt ttttaatatt tcaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aatacgtagt 7920

ttaaac 7926

<210> 10

<211> 36519

<212> DNA

<213> Artificial Sequence

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

<400> 10

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc 600

gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactggtggt aaatgccatg 660

atgggcgacg accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg 720

tatgatctgg aggtggatgt gcccgaggac gatcccaatg aggaggcggt aaatgatttt 780

tttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac 840

tcttcactgc ataccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg 900

gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag 960

caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg 1020

gactgcccgc ctctgcccgg acacggctgt aagtcttgtg aatttcatcg catgaatact 1080

ggagataaag ctgtgttgtg tgcactttgc tatatgagag cttacaacca ttgtgtttac 1140

agtaagtgtg attaagttga actttagagg gaggcagaga gcagggtgac tgggcgatga 1200

ctggtttatt tatgtatata tgttctttat ataggtcccg tctctgacgc agatgatgag 1260

acccccacta caaagtccac ttcgtcaccc ccagaaattg gcacatctcc acctgagaat 1320

attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat 1380

gacttgctac agggtggggt tgaacctttg gacttgtgta cccggaaacg ccccaggcac 1440

taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc 1500

aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt 1560

atataagcag gtgcagacct gtgtggttag ctcagagcgg catggagatt tggacggtct 1620

tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc 1680

tgtggagatt ctgcttcggt ggcgacctag ctaggctagt ctacagggcc aaacaggatt 1740

atagtgaaca atttgaggtt attttgagag agtgttctgg tctttttgac gctcttaact 1800

tgggccatca gtctcacttt aaccagagga tttcgagagc ccttgatttt actactcctg 1860

gcagaaccac tgcagcagta gccttttttg cttttattct tgacaaatgg agtcaagaaa 1920

cccatttcag cagggattac cagctggatt tcttagcagt agctttgtgg agaacatgga 1980

agtgccagcg cctgaatgca atctccggct acttgccggt acagccgcta gacactctga 2040

ggatcctgaa tctccaggag agtcccaggg cacgccaacg tcgccagcag cagcagcagg 2100

aggaggatca agaagagaac ccgagagccg gcctggaccc tccggcggag gaggaggagt 2160

agctgacctg tttcctgaac tgcgccgggt gctgactagg tcttcgagtg gtcgggagag 2220

ggggattaag cgggagaggc atgatgagac taatcacaga actgaactga ctgtgggtct 2280

gatgagtcgc aagcgcccag aaacagtgtg gtggcatgag gtgcagtcga ctggcacaga 2340

tgaggtgtcg gtgatgcatg agaggttttc tctagaacaa gtcaagactt gttggttaga 2400

gcctgaggat gattgggagg tagccatcag gaattatgcc aagctggctc tgaggccaga 2460

caagaagtac aagattacta agctgataaa tatcagaaat gcctgctaca tctcagggaa 2520

tggggctgaa gtggagatct gtctccagga aagggtggct ttcagatgct gcatgatgaa 2580

tatgtacccg ggagtggtgg gcatggatgg ggttaccttt atgaacatga ggttcagggg 2640

agatgggtat aatggcacgg tctttatggc caataccaag ctgacagtcc atggctgctc 2700

cttctttggg tttaataaca cctgcatcga ggcctggggt caggtcggtg tgaggggctg 2760

cagtttttca gccaactgga tgggggtcgt gggcaggacc aagagtatgc tgtccgtgaa 2820

gaaatgcttg tttgagaggt gccacctggg ggtgatgagc gagggcgaag ccagaatccg 2880

ccactgcgcc tctaccgaga cgggctgctt tgtgctgtgc aagggcaatg ctaagatcaa 2940

gcataatatg atctgtggag cctcggacga gcgcggctac cagatgctga cctgcgccgg 3000

cgggaacagc catatgctgg ccaccgtaca tgtggcttcc catgctcgca agccctggcc 3060

cgagttcgag cacaatgtca tgaccaggtg caatatgcat ctggggtccc gccgaggcat 3120

gttcatgccc taccagtgca acctgaatta tgtgaaggtg ctgctggagc ccgatgccat 3180

gtccagagtg agcctgacgg gggtgtttga catgaatgtg gaggtgtgga agattctgag 3240

atatgatgaa tccaagacca ggtgccgagc ctgcgagtgc ggagggaagc atgccaggtt 3300

ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg 3360

caccgggacg gagttcggtt ccagcgggga agaatctgac tagagtgagt agtgttctgg 3420

ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg 3480

cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg 3540

gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg tggacggccg 3600

gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcgtt 3660

ggacgcagct gccgccgcag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat 3720

gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atcccgccag 3780

cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct 3840

gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac 3900

ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttgtt gattttaaca 3960

cagagtctga atctttattt gatttttcgc gcgcggtagg ccctggacca ccggtctcga 4020

tcattgagca cccggtggat cttttccagg acccggtaga ggtgggcttg gatgttgagg 4080

tacatgggca tgagcccgtc ccgggggtgg aggtagctcc attgcagggc ctcgtgctcg 4140

ggggtggtgt tgtaaatcac ccagtcatag caggggcgca gggcatggtg ttgcacaata 4200

tctttgagga ggagactgat ggccacgggc agccctttgg tgtaggtgtt tacaaatctg 4260

ttgagctggg agggatgcat gcgggggggag atgaggtgca tcttggcctg gatcttgaga 4320

ttggcgatgt taccgcccag atcccgcctg gggttcatgt tgtgcaggac caccagcacg 4380

gtgtatccgg tgcacttggg gaatttatca tgcaacttgg aagggaaggc gtgaaagaat 4440

ttggcgacgc ctttgtgccc gcccaggttt tccatgcact catccatgat gatggcgatg 4500

ggcccgtggg cggcggcctg ggcaaagacg tttcgggggt cggacacatc atagttgtgg 4560

tcctgggtga ggtcatcata ggccatttta atgaatttgg ggcggagggt gccggactgg 4620

gggacaaagg taccctcgat cccgggggcg tagttcccct cacagatctg catctcccag 4680

gctttgagct cggaggggggg gatcatgtcc acctgcgggg cgataaagaa cacggtttcc 4740

ggggcggggg agatgagctg ggccgaaagc aagttccgga gcagctggga cttgccgcag 4800

ccggtggggc cgtagatgac cccgatgacc ggctgcaggt ggtagttgag ggagagacag 4860

ctgccgtcct cccggaggag gggggccacc tcgttcatca tctcgcgcac gtgcatgttc 4920

tcgcgcacca gttccgccag gaggcgctct ccccccaggg ataggagctc ctggagcgag 4980

gcgaagtttt tcagcggctt gagtccgtcg gccatgggca ttttggagag ggtttgttgc 5040

aagagttcca ggcggtccca gagctcggtg atgtgctcta cggcatctcg atccagcaga 5100

cctcctcgtt tcgcgggttg ggacggctgc gggagtaggg caccagacga tgggcgtcca 5160

gcgcagccag ggtccggtcc ttccagggtc gcagcgtccg cgtcagggtg gtctccgtca 5220

cggtgaaggg gtgcgcgccg ggctgggcgc ttgcgagggt gcgcttcagg ctcatccggc 5280

tggtcgaaaa ccgctcccga tcggcgccct gcgcgtcggc caggtagcaa ttgaccatga 5340

gttcgtagtt gagcgcctcg gccgcgtggc ctttggcgcg gagcttacct ttggaagtct 5400

gcccgcaggc gggacagagg agggacttga gggcgtagag cttgggggcg aggaagacgg 5460

actcgggggc gtaggcgtcc gcgccgcagt gggcgcagac ggtctcgcac tccacgagcc 5520

aggtgaggtc gggctggtcg gggtcaaaaa ccagtttccc gccgttcttt ttgatgcgtt 5580

tcttaccttt ggtctccatg agctcgtgtc cccgctgggt gacaaagagg ctgtccgtgt 5640

ccccgtagac cgactttatg ggccggtcct cgagcggtgt gccgcggtcc tcctcgtaga 5700

ggaaccccgc ccactccgag acgaaagccc gggtccaggc cagcacgaag gaggccacgt 5760
gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggta tgcaaacaca 5820
tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg 5880
gggtcccggc cgggggggta taaaagggtg cgggtccctg ctcgtcctca ctgtcttccg 5940
gatcgctgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga 6000
cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg 6060
cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt 6120

cgagcttggt ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca 6180
tggtctggtt tttttccttg tcggcgcgct ccttggcggc gatgttgagc tgcacgtact 6240
cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctcgtcgggc acgattctga 6300
cctgccagcc ccgattatgc agggtgatga ggtccacact ggtggccacc tcgccgcgca 6360
ggggctcatt agtccagcag aggcgtccgc ccttgcgcga gcagaagggg ggcagggggt 6420
ccagcatgac ctcgtcgggg gggtcggcat cgatggtgaa gatgccgggc aggaggtcgg 6480
ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg 6540

ccagcgcgcg ctcgtaggga ctgaggggcg tgccccaggg catgggatgg gtaagcgcgg 6600
aggcgtacat gccgcagatg tcgtagacgt agaggggctc ctcgaggatg ccgatgtagg 6660
tggggtagca gcgccccccg cggatgctgg cgcgcacgta gtcatacagc tcgtgcgagg 6720
gggcgaggag ccccgggccc aggttggtgc gactgggctt ttcggcgcgg tagacgatct 6780
ggcggaaaat ggcatgcgag ttggaggaga tggtgggcct ttggaagatg ttgaagtggg 6840
cgtggggcag tccgaccgag tcgcggatga agtgggcgta ggagtcttgc agcttggcga 6900
cgagctcggc ggtgactagg acgtccagag cgcagtagtc gagggtctcc tggatgatgt 6960

catacttgag ctgtcccttt tgtttccaca gctcgcggtt gagaaggaac tcttcgcggt 7020
ccttccagta ctcttcgagg gggaacccgt cctgatctgc acggtaagag cctagcatgt 7080
agaactggtt gacggccttg taggcgcagc agcccttctc cacggggagg gcgtaggcct 7140
gggcggcctt gcgcagggag gtgtgcgtga gggcgaaagt gtccctgacc atgaccttga 7200
ggaactggtg cttgaagtcg atatcgtcgc agccccccctg ctcccagagc tggaagtccg 7260
tgcgcttctt gtaggcgggg ttgggcaaag cgaaagtaac atcgttgaag aggatcttgc 7320
ccgcgcgggg cataaagttg cgagtgatgc ggaaaggttg gggcacctcg gcccggttgt 7380

tgatgacctg ggcggcgagc acgatctcgt cgaagccgtt gatgttgtgg cccacgatgt 7440

agagttccac gaatcgcgga cggcccttga cgtggggcag tttcttgagc tcctcgtagg 7500

tgagctcgtc ggggtcgctg agcccgtgct gctcgagcgc ccagtcggcg agatgggggt 7560

tggcgcggag gaaggaagtc cagagatcca cggccagggc ggtttgcaga cggtcccggt 7620

actgacggaa ctgctgcccg acggccattt tttcgggggt gacgcagtag aaggtgcggg 7680

ggtccccgtg ccagcgatcc catttgagct ggagggcgag atcgagggcg agctcgacga 7740

gccggtcgtc cccggagagt ttcatgacca gcatgaaggg gacgagctgc ttgccgaagg 7800

accccatcca ggtgtaggtt tccacatcgt aggtgaggaa gagcctttcg gtgcgaggat 7860

gcgagccgat ggggaagaac tggatctcct gccaccaatt ggaggaatgg ctgttgatgt 7920

gatggaagta gaaatgccga cggcgcgccg aacactcgtg cttgtgttta tacaagcggc 7980

cacagtgctc gcaacgctgc acgggatgca cgtgctgcac gagctgtacc tgagttcctt 8040

tgacgaggaa tttcagtggg aagtggagtc gtggcgcctg catctcgtgc tgtactacgt 8100

cgtggtggtc ggcctggccc tcttctgcct cgatggtggt catgctgacg agcccgcgcg 8160

ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc 8220

cggagctgtc cagggtcctg agacgctgcg gagtcaggtc agtgggcagc ggcggcgcgc 8280

ggttgacttg caggagtttt tccagggcgc gcgggaggtc cagatggtac ttgatctcca 8340

ccgcgccatt ggtggcgacg tcgatggctt gcagggtccc gtgcccctgg ggtgtgacca 8400

ccgtcccccg tttcttcttg ggcggctggg gcgacggggg cggtgcctct tccatggtta 8460

gaagcggcgg cgaggacgcg cgccgggcgg caggggcggc tcggggcccg gaggcagggg 8520

cggcaggggc acgtcggcgc cgcgcgcggg taggttctgg tactgcgccc ggagaagact 8580

ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgcctctggg tgaaggccac 8640

gggacccgtg agtttgaacc tgaaagagag ttcgacagaa tcaatctcgg tatcgttgac 8700

ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcggt 8760

catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc 8820

cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt 8880

ccagacgcgg ctgtagacca cgacgccctc gggatcgcgg gcgcgcatga ccacctgggc 8940

gaggttgagc tccacgtggc gcgtgaagac cgcgtagttg cagaggcgct ggtagaggta 9000

gttgagcgtg gtggcgatgt gctcggtgac gaagaaatac atgatccagc ggcggagcgg 9060

catctcgctg acgtcgccca gcgcctccaa acgttccatg gcctcgtaaa agtccacggc 9120

gaagttgaaa aactgggagt tgcgcgccga gacggtcaac tcctcctcca gaagacggat 9180

gagctcggcg atggtggcgc gcacctcgcg ctcgaaggcc cccgggagtt cctccacttc 9240

ctcttcttcc tcctccacta acatctcttc tacttcctcc tcaggcggca gtggtggcgg 9300

gggaggggggc ctgcgtcgcc ggcggcgcac gggcagacgg tcgatgaagc gctcgatggt 9360

ctcgccgcgc cggcgtcgca tggtctcggt gacggcgcgc ccgtcctcgc ggggccgcag 9420

cgtgaagacg ccgccgcgca tctccaggtg gccgggggggg tccccgttgg gcagggagag 9480

ggcgctgacg atgcatctta tcaattgccc cgtagggact ccgcgcaagg acctgagcgt 9540

ctcgagatcc acgggatctg aaaaccgctg aacgaaggct tcgagccagt cgcagtcgca 9600

aggtaggctg agcacggttt cttctggcgg gtcatgttgg ttgggagcgg ggcgggcgat 9660

gctgctggtg atgaagttga aataggcggt tctgagacgg cggatggtgg cgaggagcac 9720

caggtctttg ggcccggctt gctggatgcg cagacggtcg gccatgcccc aggcgtggtc 9780

ctgacacctg gccaggtcct tgtagtagtc ctgcatgagc cgctccacgg gcacctcctc 9840

ctcgcccgcg cggccgtgca tgcgcgtgag cccgaagccg cgctggggct ggacgagcgc 9900

caggtcggcg acgacgcgct cggcgaggat ggcttgctgg atctgggtga gggtggtctg 9960

gaagtcatca aagtcgacga agcggtggta ggctccggtg ttgatggtgt aggagcagtt 10020

ggccatgacg gaccagttga cggtctggtg gcccggacgc acgagctcgt ggtacttgag 10080

gcgcgagtag gcgcgcgtgt cgaagatgta gtcgttgcag gtgcgcacca ggtactggta 10140

gccgatgagg aagtgcggcg gcggctggcg gtagagcggc catcgctcgg tggcgggggc 10200

gccgggcgcg aggtcctcga gcatggtgcg gtggtagccg tagatgtacc tggacatcca 10260

ggtgatgccg gcggcggtgg tggaggcgcg cgggaactcg cggacgcggt tccagatgtt 10320

gcgcagcggc aggaagtagt tcatggtggg cacggtctgg cccgtgaggc gcgcgcagtc 10380

gtggatgctc tatacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtattg gcactcccgt ctcgacccaa gcctgcacca accctccagg 10560

atacggaggc gggtcgtttt gcaacttttt tttggaggcc ggatgagact agtaagcgcg 10620

gaaagcggcc gaccgcgatg gctcgctgcc gtagtctgga gaagaatcgc cagggttgcg 10680

ttgcggtgtg ccccggttcg aggccggccg gattccgcgg ctaacgaggg cgtggctgcc 10740

ccgtcgtttc caagacccca tagccagccg acttctccag ttacggagcg agcccctctt 10800

ttgttttgtt tgtttttgcc agatgcatcc cgtactgcgg cagatgcgcc cccaccaccc 10860

tccaccgcaa caacagcccc ctccacagcc ggcgcttctg cccccgcccc agcagcaact 10920

tccagccacg accgccgcgg ccgccgtgag cggggctgga cagagttatg atcaccagct 10980

ggccttggaa gagggcgagg ggctggcgcg cctgggggcg tcgtcgccgg agcggcaccc 11040

gcgcgtgcag atgaaaaggg acgctcgcga ggcctacgtg cccaagcaga acctgttcag 11100

agacaggagc ggcgaggagc ccgaggagat gcgcgcggcc cggttccacg cggggcggga 11160

gctgcggcgc ggcctggacc gaaagagggt gctgagggac gaggatttcg aggcggacga 11220
gctgacgggg atcagccccg cgcgcgcgca cgtggccgcg gccaacctgg tcacggcgta 11280
cgagcagacc gtgaaggagg agagcaactt ccaaaaatcc ttcaacaacc acgtgcgcac 11340
cctgatcgcg cgcgaggagg tgaccctggg cctgatgcac ctgtgggacc tgctggaggc 11400
catcgtgcag aaccccacca gcaagccgct gacggcgcag ctgttcctgg tggtgcagca 11460
tagtcgggac aacgaagcgt tcagggaggc gctgctgaat atcaccgagc ccgagggccg 11520
ctggctcctg gacctggtga acattctgca gagcatcgtg gtgcaggagc gcgggctgcc 11580

gctgtccgag aagctggcgg ccatcaactt ctcggtgctg agtttgggca agtactacgc 11640
taggaagatc tacaagaccc cgtacgtgcc catagacaag gaggtgaaga tcgacgggtt 11700
ttacatgcgc atgaccctga aagtgctgac cctgagcgac gatctggggg tgtaccgcaa 11760
cgacaggatg caccgtgcgg tgagcgccag caggcggcgc gagctgagcg accaggagct 11820
gatgcatagt ctgcagcggg ccctgaccgg ggccgggacc gaggggggaga gctactttga 11880
catgggcgcg gacctgcact ggcagcccag ccgccgggcc ttggaggcgg cggcaggacc 11940
ctacgtagaa gaggtggacg atgaggtgga cgaggagggc gagtacctgg aagactgatg 12000

gcgcgaccgt atttttgcta gatgcaacaa caacagccac ctcctgatcc cgcgatgcgg 12060
gcggcgctgc agagccagcc gtccggcatt aactcctcgg acgattggac ccaggccatg 12120
caacgcatca tggcgctgac gacccgcaac cccgaagcct ttagacagca gccccaggcc 12180
aaccggctct cggccatcct ggaggccgtg gtgccctcgc gctccaaccc cacgcacgag 12240
aaggtcctgg ccatcgtgaa cgcgctggtg gagaacaagg ccatccgcgg cgacgaggcc 12300
ggcctggtgt acaacgcgct gctggagcgc gtggcccgct acaacagcac caacgtgcag 12360
accaacctgg accgcatggt gaccgacgtg cgcgaggccg tggcccagcg cgagcggttc 12420

caccgcgagt ccaacctggg atccatggtg gcgctgaacg ccttcctcag cacccagccc 12480
gccaacgtgc cccggggcca ggaggactac accaacttca tcagcgccct gcgcctgatg 12540
gtgaccgagg tgccccagag cgaggtgtac cagtccgggc cggactactt cttccagacc 12600
agtcgccagg gcttgcagac cgtgaacctg agccaggctt tcaagaactt gcagggcctg 12660
tggggcgtgc aggccccggt cggggaccgc gcgacggtgt cgagcctgct gacgccgaac 12720
tcgcgcctgc tgctgctgct ggtggccccc ttcacggaca gcggcagcat caaccgcaac 12780
tcgtacctgg gctacctgat taacctgtac cgcgaggcca tcggccaggc gcacgtggac 12840

gagcagacct accaggagat cacccacgtg agccgcgccc tgggccagga cgacccgggc 12900

aacctggaag ccaccctgaa ctttttgctg accaaccggt cgcagaagat cccgccccag 12960

tacgcgctca gcaccgagga ggagcgcatc ctgcgttacg tgcagcagag cgtgggcctg 13020

ttcctgatgc aggagggggc cacccccagc gccgcgctcg acatgaccgc gcgcaacatg 13080

gagcccagca tgtacgccag caaccgcccg ttcatcaata aactgatgga ctacttgcat 13140

cgggcggccg ccatgaactc tgactatttc accaacgcca tcctgaatcc ccactggctc 13200

ccgccgccgg ggttctacac gggcgagtac gacatgcccg accccaatga cgggttcctg 13260

tgggacgatg tggacagcag cgtgttctcc ccccgaccgg gtgctaacga gcgccccttg 13320

tggaagaagg aaggcagcga ccgacgcccg tcctcggcgc tgtccggccg cgagggtgct 13380

gccgcggcgg tgcccgaggc cgccagtcct ttcccgagct tgcccttctc gctgaacagt 13440

atccgcagca gcgagctggg caggatcacg cgcccgcgct tgctgggcga agaggagtac 13500

ttgaatgact cgctgttgag acccgagcgg gagaagaact tccccaataa cgggatagaa 13560

agcctggtgg acaagatgag ccgctggaag acgtatgcgc aggagcacag ggacgatccc 13620

cgggcgtcgc agggggccac gagccggggc agcgccgccc gtaaacgccg gtggcacgac 13680

aggcagcggg gacagatgtg ggacgatgag gactccgccg acgacagcag cgtgttggac 13740

ttgggtggga gtggtaaccc gttcgctcac ctgcgccccc gtatcgggcg catgatgtaa 13800

gagaaaccga aaataaatga tactcaccaa ggccatggcg accagcgtgc gttcgtttct 13860

tctctgttgt tgttgtatct agtatgatga ggcgtgcgta cccggagggt cctcctccct 13920

cgtacgagag cgtgatgcag caggcgatgg cggcggcggc gatgcagccc ccgctggagg 13980

ctccttacgt gcccccgcgg tacctggcgc ctacggaggg gcggaacagc attcgttact 14040

cggagctggc acccttgtac gataccaccc ggttgtacct ggtggacaac aagtcggcgg 14100

acatcgcctc gctgaactac cagaacgacc acagcaactt cctgaccacc gtggtgcaga 14160

acaatgactt cacccccacg gaggccagca cccagaccat caactttgac gagcgctcgc 14220

ggtggggcgg ccagctgaaa accatcatgc acaccaacat gcccaacgtg aacgagttca 14280

tgtacagcaa caagttcaag gcgcgggtga tggtctcccg caagaccccc aatggggtga 14340

cagtgacaga ggattatgat ggtagtcagg atgagctgaa gtatgaatgg gtggaatttg 14400

agctgcccga aggcaacttc tcggtgacca tgaccatcga cctgatgaac aacgccatca 14460

tcgacaatta cttggcggtg gggcggcaga acggggtgct ggagagcgac atcggcgtga 14520

agttcgacac taggaacttc aggctgggct gggaccccgt gaccgagctg gtcatgcccg 14580

gggtgtacac caacgaggct ttccatcccg atattgtctt gctgcccggc tgcggggtgg 14640

acttcaccga gagccgcctc agcaacctgc tgggcattcg caagaggcag cccttccagg 14700

aaggcttcca gatcatgtac gaggatctgg agggggggcaa catccccgcg ctcctggatg 14760

tcgacgccta tgagaaaagc aaggaggatg cagcagctga agcaactgca gccgtagcta 14820

ccgcctctac cgaggtcagg ggcgataatt ttgcaagcgc cgcagcagtg gcagcggccg 14880

aggcggctga aaccgaaagt aagatagtca ttcagccggt ggagaaggat agcaagaaca 14940

ggagctacaa cgtactaccg gacaagataa acaccgccta ccgcagctgg tacctagcct 15000

acaactatgg cgaccccgag aagggcgtgc gctcctggac gctgctcacc acctcggacg 15060

tcacctgcgg cgtggagcaa gtctactggt cgctgcccga catgatgcaa gacccggtca 15120

ccttccgctc cacgcgtcaa gttagcaact acccggtggt gggcgccgag ctcctgcccg 15180

tctactccaa gagcttcttc aacgagcagg ccgtctactc gcagcagctg cgcgccttca 15240

cctcgcttac gcacgtcttc aaccgcttcc ccgagaacca gatcctcgtc cgcccgcccg 15300

cgcccaccat taccaccgtc agtgaaaacg ttcctgctct cacagatcac gggaccctgc 15360

cgctgcgcag cagtatccgg ggagtccagc gcgtgaccgt tactgacgcc agacgccgca 15420

cctgccccta cgtctacaag gccctgggca tagtcgcgcc gcgcgtcctc tcgagccgca 15480

ccttctaaat gtccattctc atctcgccca gtaataacac cggttggggc ctgcgcgcgc 15540

ccagcaagat gtacggaggc gctcgccaac gctccacgca acaccccgtg cgcgtgcgcg 15600

ggcacttccg cgctccctgg ggcgccctca agggccgcgt gcggtcgcgc accaccgtcg 15660

acgacgtgat cgaccaggtg gtggccgacg cgcgcaacta cacccccgcc gccgcgcccg 15720

tctccaccgt ggacgccgtc atcgacagcg tggtggccga cgcgcgccgg tacgcccgcg 15780

ccaagagccg gcggcggcgc atcgcccggc ggcaccggag cacccccgcc atgcgcgcgg 15840

cgcgagcctt gctgcgcagg gccaggcgca cgggacgcag ggccatgctc agggcggcca 15900

gacgcgcggc ttcaggcgcc agcgccggca ggacccggag acgcgcggcc acggcggcgg 15960

cagcggccat cgccagcatg tcccgcccgc ggcgagggaa cgtgtactgg gtgcgcgacg 16020

ccgccaccgg tgtgcgcgtg cccgtgcgca cccgcccccc tcgcacttga agatgttcac 16080

ttcgcgatgt tgatgtgtcc cagcggcgag gaggatgtcc aagcgcaaat tcaaggaaga 16140

gatgctccag gtcatcgcgc ctgagatcta cggccctgcg gtggtgaagg aggaaagaaa 16200

gccccgcaaa atcaagcggg tcaaaaagga caaaaaggaa gaagaaagtg atgtggacgg 16260

attggtggag tttgtgcgcg agttcgcccc ccggcggcgc gtgcagtggc gcgggcggaa 16320

ggtgcaaccg gtgctgagac ccggcaccac cgtggtcttc acgcccggcg agcgctccgg 16380

caccgcttcc aagcgctcct acgacgaggt gtacggggat gatgatattc tggagcaggc 16440

ggccgagcgc ctgggcgagt ttgcttacgg caagcgcagc cgttccgcac cgaaggaaga 16500

ggcggtgtcc atcccgctgg accacggcaa ccccacgccg agcctcaagc ccgtgacctt 16560

gcagcaggtg ctgccgaccg cggcgccgcg ccgggggttc aagcgcgagg gcgaggatct 16620

gtaccccacc atgcagctga tggtgcccaa gcgccagaag ctggaagacg tgctggagac 16680
catgaaggtg gacccggacg tgcagcccga ggtcaaggtg cggcccatca agcaggtggc 16740
cccgggcctg ggcgtgcaga ccgtggacat caagattccc acggagccca tggaaacgca 16800
gaccgagccc atgatcaagc ccagcaccag caccatggag gtgcagacgg atccctggat 16860
gccatcggct cctagtcgaa gaccccggcg caagtacggc gcggccagcc tgctgatgcc 16920
caactacgcg ctgcatcctt ccatcatccc cacgccgggc taccgcggca cgcgcttcta 16980
ccgcggtcat accagcagcc gccgccgcaa gaccaccact cgccgccgcc gtcgccgcac 17040

cgccgctgca accacccctg ccgccctggt gcggagagtg taccgccgcg gccgcgcacc 17100
tctgaccctg ccgcgcgcgc gctaccaccc gagcatcgcc atttaaactt tcgcctgctt 17160
tgcagatcaa tggccctcac atgccgcctt cgcgttccca ttacgggcta ccgaggaaga 17220
aaaccgcgcc gtagaaggct ggcggggaac gggatgcgtc gccaccacca ccggcggcgg 17280
cgcgccatca gcaagcggtt gggggggaggc ttcctgcccg cgctgatccc catcatcgcc 17340
gcggcgatcg gggcgatccc cggcattgct tccgtggcgg tgcaggcctc tcagcgccac 17400
tgagacacac ttggaaacat cttgtaataa accaatggac tctgacgctc ctggtcctgt 17460

gatgtgtttt cgtagacaga tggaagacat caatttttcg tccctggctc cgcgacacgg 17520
cacgcggccg ttcatgggca cctggagcga catcggcacc agccaactga acgggggcgc 17580
cttcaattgg agcagtctct ggagcgggct taagaatttc gggtccacgc ttaaaaccta 17640
tggcagcaag gcgtggaaca gcaccacagg gcaggcgctg agggataagc tgaaagagca 17700
gaacttccag cagaaggtgg tcgatgggct cgcctcgggc atcaacgggg tggtggacct 17760
ggccaaccag gccgtgcagc ggcagatcaa cagccgcctg gacccggtgc cgcccgccgg 17820
ctccgtggag atgccgcagg tggaggagga gctgcctccc ctggacaagc ggggcgagaa 17880

gcgaccccgc cccgatgcgg aggagacgct gctgacgcac acggacgagc cgcccccgta 17940
cgaggaggcg gtgaaactgg gtctgcccac cacgcggccc atcgcgcccc tggccaccgg 18000
ggtgctgaaa cccgaaaagc ccgcgaccct ggacttgcct cctccccagc cttcccgccc 18060
ctctacagtg gctaagcccc tgccgccggt ggccgtggcc cgcgcgcgac ccgggggcac 18120
cgcccgccct catgcgaact ggcagagcac tctgaacagc atcgtgggtc tgggagtgca 18180
gagtgtgaag cgccgccgct gctattaaac ctaccgtagc gcttaacttg cttgtctgtg 18240
tgtgtatgta ttatgtcgcc gccgccgctg tccaccagaa ggaggagtga agaggcgcgt 18300

cgccgagttg caagatggcc accccatcga tgctgcccca gtgggcgtac atgcacatcg 18360

ccggacagga cgcttcggag tacctgagtc cgggtctggt gcagtttgcc cgcgccacag 18420

acacctactt cagtctgggg aacaagttta ggaaccccac ggtggcgccc acgcacgatg 18480

tgaccaccga ccgcagccag cggctgacgc tgcgcttcgt gcccgtggac cgcgaggaca 18540

acacctactc gtacaaagtg cgctacacgc tggccgtggg cgacaaccgc gtgctggaca 18600

tggccagcac ctactttgac atccgcggcg tgctggatcg gggccctagc ttcaaaccct 18660

actccggcac cgcctacaac agtctggccc ccaagggagc acccaacact tgtcagtgga 18720

catataaagc cgatggtgaa actgccacag aaaaaaccta tacatatgga aatgcacccg 18780

tgcagggcat taacatcaca aaagatggta ttcaacttgg aactgacacc gatgatcagc 18840

caatctacgc agataaaacc tatcagcctg aacctcaagt gggtgatgct gaatggcatg 18900

acatcactgg tactgatgaa aagtatggag gcagagctct taagcctgat accaaaatga 18960

agccttgtta tggttctttt gccaagccta ctaataaaga aggaggtcag gcaaatgtga 19020

aaacaggaac aggcactact aaagaatatg acatagacat ggctttcttt gacaacagaa 19080

gtgcggctgc tgctggccta gctccagaaa ttgttttgta tactgaaaat gtggatttgg 19140

aaactccaga tacccatatt gtatacaaag caggcacaga tgacagcagc tcttctatta 19200

atttgggtca gcaagccatg cccaacagac ctaactacat tggtttcaga gacaacttta 19260

tcgggctcat gtactacaac agcactggca atatgggggt gctggccggt caggcttctc 19320

agctgaatgc tgtggttgac ttgcaagaca gaaacaccga gctgtcctac cagctcttgc 19380

ttgactctct gggtgacaga acccggtatt tcagtatgtg gaatcaggcg gtggacagct 19440

atgatcctga tgtgcgcatt attgaaaatc atggtgtgga ggatgaactt cccaactatt 19500

gtttccctct ggatgctgtt ggcagaacag atacttatca gggaattaag gctaatggaa 19560

ctgatcaaac cacatggacc aaagatgaca gtgtcaatga tgctaatgag ataggcaagg 19620

gtaatccatt cgccatggaa atcaacatcc aagccaacct gtggaggaac ttcctctacg 19680

ccaacgtggc cctgtacctg cccgactctt acaagtacac gccggccaat gttaccctgc 19740

ccaccaacac caacacctac gattacatga acggccgggt ggtggcgccc tcgctggtgg 19800

actcctacat caacatcggg gcgcgctggt cgctggatcc catggacaac gtgaacccct 19860

tcaaccacca ccgcaatgcg gggctgcgct accgctccat gctcctgggc aacgggcgct 19920

acgtgccctt ccacatccag gtgccccaga aatttttcgc catcaagagc ctcctgctcc 19980

tgcccgggtc ctacacctac gagtggaact tccgcaagga cgtcaacatg atcctgcaga 20040

gctccctcgg caacgacctg cgcacggacg gggcctccat ctccttcacc agcatcaacc 20100

tctacgccac cttcttcccc atggcgcaca acacggcctc cacgctcgag gccatgctgc 20160

gcaacgacac caacgaccag tccttcaacg actacctctc ggcggccaac atgctctacc 20220

ccatcccggc caacgccacc aacgtgccca tctccatccc ctcgcgcaac tgggccgcct 20280

tccgcggctg gtccttcacg cgtctcaaga ccaaggagac gccctcgctg ggctccgggt 20340

tcgaccccta cttcgtctac tcgggctcca tcccctacct cgacggcacc ttctacctca 20400

accacacctt caagaaggtc tccatcacct tcgactcctc cgtcagctgg cccggcaacg 20460

accggctcct gacgcccaac gagttcgaaa tcaagcgcac cgtcgacggc gagggctaca 20520

acgtggccca gtgcaacatg accaaggact ggttcctggt ccagatgctg gcccactaca 20580

acatcggcta ccagggcttc tacgtgcccg agggctacaa ggaccgcatg tactccttct 20640

tccgcaactt ccagcccatg agccgccagg tggtggacga ggtcaactac aaggactacc 20700

aggccgtcac cctggcctac cagcacaaca actcgggctt cgtcggctac ctcgcgccca 20760

ccatgcgcca gggccagccc taccccgcca actaccccta cccgctcatc ggcaagagcg 20820

ccgtcaccag cgtcacccag aaaaagttcc tctgcgacag ggtcatgtgg cgcatcccct 20880

tctccagcaa cttcatgtcc atgggcgcgc tcaccgacct cggccagaac atgctctatg 20940

ccaactccgc ccacgcgcta gacatgaatt tcgaagtcga ccccatggat gagtccaccc 21000

ttctctatgt tgtcttcgaa gtcttcgacg tcgtccgagt gcaccagccc caccgcggcg 21060

tcatcgaggc cgtctacctg cgcacccccct tctcggccgg taacgccacc acctaagctc 21120

ttgcttcttg caagccatgg ccgcgggctc cggcgagcag gagctcaggg ccatcatccg 21180

cgacctgggc tgcgggccct acttcctggg caccttcgat aagcgcttcc cgggattcat 21240

ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccgggggcga 21300

gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgacccctt 21360

cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg 21420

ccgcagcgcc ctggccaccg aggaccgctg cgtcaccctg gaaaagtcca cccagaccgt 21480

gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt 21540

gcactggccc gaccgcccca tggacaagaa ccccaccatg aacttgctga cgggggtgcc 21600

caacggcatg ctccagtcgc cccaggtgga acccaccctg cgccgcaacc aggaggcgct 21660

ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa 21720

ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaatgtc 21780

tttaataaac agcactttca tgttacacat gcatctgaga tgatttattt agaaatcgaa 21840

agggttctgc cgggtctcgg catggcccgc gggcagggac acgttgcgga actggtactt 21900

ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc 21960

ggtccacagc ttccgcgtca gttgcagggc gcccagcagg tcgggcgcgg agatcttgaa 22020

atcgcagttg ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg 22080

gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc 22140
cacgtcgagg tcctcggcgt tggccatccc gaaggggggtc atcttgcagg tctgccttcc 22200
catggtgggc acgcacccgg gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat 22260
ctgggcctgg tcggcgttca tccccgggta catggccttc atgaaagcct ccaattgcct 22320
gaacgcctgc tgggccttgg ctccctcggt gaagaagacc ccgcaggact tgctagagaa 22380
ctggttggtg gcgcacccgg cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg 22440
caccacgctg cgcccccagc ggttctgggt gatcttggcc cggtcggggt tctccttcag 22500

cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat 22560
ggtggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag 22620
cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc 22680
ctgcaggaag cggcccatca tggtggtcag ggtcttgttg ctagtgaagg tcagcggaat 22740
gccgcggtgc tcctcgttga tgtacaggtg gcagatgcgg cggtacacct cgccctgctc 22800
gggcatcagc tggaagttgg ctttcaggtc ggtctccacg cggtagcggt ccatcagcat 22860
agtcatgatt tccataccct tctcccaggc cgagacgatg ggcaggctca tagggttctt 22920

caccatcatc ttagcgctag cagccgcggc cagggggtcg ctctcgtcca gggtctcaaa 22980
gctccgcttg ccgtccttct cggtgatccg caccgggggg tagctgaagc ccacggccgc 23040
cagctcctcc tcggcctgtc tttcgtcctc gctgtcctgg ctgacgtcct gcaggaccac 23100
atgcttggtc ttgcggggtt tcttcttggg cggcagcggc ggcggagatg ttggagatgg 23160
cgagggggag cgcgagttct cgctcaccac tactatctct tcctcttctt ggtccgaggc 23220
cacgcggcgg taggtatgtc tcttcggggg cagaggcgga ggcgacgggc tctcgccgcc 23280
gcgacttggc ggatggctgg cagagcccct tccgcgttcg ggggtgcgct cccggcggcg 23340

ctctgactga cttcctccgc ggccggccat tgtgttctcc tagggaggaa caacaagcat 23400
ggagactcag ccatcgccaa cctcgccatc tgcccccacc gccgacgaga agcagcagca 23460
gcagaatgaa agcttaaccg ccccgccgcc cagccccgcc acctccgacg cggccgtccc 23520
agacatgcaa gagatggagg aatccatcga gattgacctg ggctatgtga cgcccgcgga 23580
gcacgaggag gagctggcag tgcgcttttc acaagaagag atacaccaag aacagccaga 23640
gcaggaagca gagaatgagc agagtcaggc tgggctcgag catgacggcg actacctcca 23700
cctgagcggg ggggaggacg cgctcatcaa gcatctggcc cggcaggcca ccatcgtcaa 23760

ggatgcgctg ctcgaccgca ccgaggtgcc cctcagcgtg gaggagctca gccgcgccta 23820

cgagttgaac ctcttctcgc cgcgcgtgcc ccccaagcgc cagcccaatg gcacctgcga 23880

gcccaacccg cgcctcaact tctacccggt cttcgcggtg cccgaggccc tggccaccta 23940

ccacatcttt ttcaagaacc aaaagatccc cgtctcctgc cgcgccaacc gcacccgcgc 24000

cgacgccctt ttcaacctgg gtcccggcgc ccgcctacct gatatcgcct ccttggaaga 24060

ggttcccaag atcttcgagg gtctgggcag cgacgagact cgggccgcga acgctctgca 24120

aggagaagga ggagagcatg agcaccacag cgccctggtc gagttggaag gcgacaacgc 24180

gcggctggcg gtgctcaaac gcacggtcga gctgacccat ttcgcctacc cggctctgaa 24240

cctgcccccc aaagtcatga gcgcggtcat ggaccaggtg ctcatcaagc gcgcgtcgcc 24300

catctccgag gacgagggca tgcaagactc cgaggagggc aagcccgtgg tcagcgacga 24360

gcagctggcc cggtggctgg gtcctaatgc tagtccccag agtttggaag agcggcgcaa 24420

actcatgatg gccgtggtcc tggtgaccgt ggagctggag tgcctgcgcc gcttcttcgc 24480

cgacgcggag accctgcgca aggtcgagga gaacctgcac tacctcttca ggcacgggtt 24540

cgtgcgccag gcctgcaaga tctccaacgt ggagctgacc aacctggtct cctacatggg 24600

catcttgcac gagaaccgcc tggggcagaa cgtgctgcac accaccctgc gcggggaggc 24660

ccggcgcgac tacatccgcg actgcgtcta cctctacctc tgccacacct ggcagacggg 24720

catgggcgtg tggcagcagt gtctggagga gcagaacctg aaagagctct gcaagctcct 24780

gcagaagaac ctcaagggtc tgtggaccgg gttcgacgag cgcaccaccg cctcggacct 24840

ggccgacctc attttccccg agcgcctcag gctgacgctg cgcaacggcc tgcccgactt 24900

tatgagccaa agcatgttgc aaaactttcg ctctttcatc ctcgaacgct ccggaatcct 24960

gcccgccacc tgctccgcgc tgccctcgga cttcgtgccg ctgaccttcc gcgagtgccc 25020

cccgccgctg tggagccact gctacctgct gcgcctggcc aactacctgg cctaccactc 25080

ggacgtgatc gaggacgtca gcggcgaggg cctgctcgag tgccactgcc gctgcaacct 25140

ctgcacgccg caccgctccc tggcctgcaa ccccagctg ctgagcgaga cccagatcat 25200

cggcaccttc gagttgcaag ggcccagcga aggcgagggt tcagccgcca agggggtct 25260

gaaactcacc ccggggctgt ggacctcggc ctacttgcgc aagttcgtgc ccgaggacta 25320

ccatcccttc gagatcaggt tctacgagga ccaatcccat ccgcccaagg ccgagctgtc 25380

ggcctgcgtc atcacccagg gggcgatcct ggcccaattg caagccatcc agaaatcccg 25440

ccaagaattc ttgctgaaaa agggccgcgg ggtctacctc gacccccaga ccggtgagga 25500

gctcaacccc ggcttccccc aggatgcccc gaggaaacaa gaagctgaaa gtggagctgc 25560

cgcccgtgga ggatttggag gaagactggg agaacagcag tcaggcagag gaggaggaga 25620

tggaggaaga ctgggacagc actcaggcag aggaggacag cctgcaagac agtctggagg 25680

aagacgagga ggaggcagag gaggaggtgg aagaagcagc cgccgccaga ccgtcgtcct 25740

cggcgggggа gaaagcaagc agcacggata ccatctccgc tccgggtcgg ggtcccgctc 25800

gaccacacag tagatgggac gagaccggac gattcccgaa ccccaccacc cagaccggta 25860

agaaggagcg gcagggatac aagtcctggc gggggcacaa aaacgccatc gtctcctgct 25920

tgcaggcctg cgggggcaac atctccttca cccggcgcta cctgctcttc caccgcgggg 25980

tgaactttcc ccgcaacatc ttgcattact accgtcacct ccacagcccc tactacttcc 26040

aagaagaggc agcagcagca gaaaaagacc agcagaaaac cagcagctag aaaatccaca 26100

gcggcggcag caggtggact gaggatcgcg gcgaacgagc cggcgcaaac ccgggagctg 26160

aggaaccgga tctttcccac cctctatgcc atcttccagc agagtcgggg gcaggagcag 26220

gaactgaaag tcaagaaccg ttctctgcgc tcgctcaccc gcagttgtct gtatcacaag 26280

agcgaagacc aacttcagcg cactctcgag gacgccgagg ctctcttcaa caagtactgc 26340

gcgctcactc ttaaagagta gcccgcgccc gcccagtcgc agaaaaaggc gggaattacg 26400

tcacctgtgc ccttcgccct agccgcctcc acccatcatc atgagcaaag agattcccac 26460

gccttacatg tggagctacc agccccagat gggcctggcc gccggtgccg cccaggacta 26520

ctccacccgc atgaattggc tcagcgccgg gcccgcgatg atctcacggg tgaatgacat 26580

ccgcgcccac cgaaaccaga tactcctaga acagtcagcg ctcaccgcca cgccccgcaa 26640

tcacctcaat ccgcgtaatt ggcccgccgc cctggtgtac caggaaattc cccagcccac 26700

gaccgtacta cttccgcgag acgcccaggc cgaagtccag ctgactaact caggtgtcca 26760

gctggcgggc ggcgccaccc tgtgtcgtca ccgccccgct cagggtataa agcggctggt 26820

gatccggggc agaggcacac agctcaacga cgaggtggtg agctcttcgc tgggtctgcg 26880

acctgacgga gtcttccaac tcgccggatc ggggagatct tccttcacgc ctcgtcaggc 26940

cgtcctgact ttggagagtt cgtcctcgca gccccgctcg ggtggcatcg gcactctcca 27000

gttcgtggag gagttcactc cctcggtcta cttcaacccc ttctccggct ccccggcca 27060

ctacccggac gagttcatcc cgaacttcga cgccatcagc gagtcggtgg acggctacga 27120

ttgaatgtcc catggtggcg cagctgacct agctcggctt cgacacctgg accactgccg 27180

ccgcttccgc tgcttcgctc gggatctcgc cgagtttgcc tactttgagc tgcccgagga 27240

gcaccctcag ggcccggccc acggagtgcg gatcgtcgtc gaaggggcc tcgactccca 27300

cctgcttcgg atcttcagcc agcgtccgat cctggtcgag cgcgagcaag gacagaccct 27360

tctgactctg tactgcatct gcaaccaccc cggcctgcat gaaagtcttt gttgtctgct 27420

gtgtactgag tataataaaa gctgagatca gcgactactc cggacttccg tgtgttcctg 27480

aatccatcaa ccagtctttg ttcttcaccg ggaacgagac cgagctccag ctccagtgta 27540

agccccacaa gaagtacctc acctggctgt tccagggctc cccgatcgcc gttgtcaacc 27600
actgcgacaa cgacggagtc ctgctgagcg gccctgccaa ccttactttt tccacccgca 27660
gaagcaagct ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac 27720
cctgccatca caccttccac ctgatcccga ataccacagc gtcgctcccc gctactaaca 27780
accaaactaa cctccaccaa cgccaccgtc gcgacctttc tgaatctaat actaccaccc 27840
acaccggagg tgagctccga ggtcaaccaa cctctgggat ttactacggc ccctgggagg 27900
tggttgggtt aatagcgcta ggcctagttg cgggtgggct tttggttctc tgctacctat 27960

acctcccttg ctgttcgtac ttagtggtgc tgtgttgctg gtttaagaaa tggggaagat 28020
caccctagtg agctgcggtg cgctggtggc ggtgttgctt tcgattgtgg gactgggcgg 28080
tgcggctgta gtgaaggaga aggccgatcc ctgcttgcat ttcaatccca acaaatgcca 28140
gctgagtttt cagcccgatg gcaatcggtg cgcggtactg atcaagtgcg gatgggaatg 28200
cgagaacgtg agaatcgagt acaataacaa gactcggaac aatactctcg cgtccgtgtg 28260
gcagcccggg gaccccgagt ggtacaccgt ctctgtcccc ggtgctgacg gctccccgcg 28320
caccgtgaat aatactttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa 28380

gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag 28440
cctgtgcacg gcgctaatca ccgctatcgt gtgcctgagc attcacatgc tcatcgctat 28500
tcgccccaga aataatgccg aaaaagaaaa acagccataa cgtttttttt cacacctttt 28560
tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca 28620
tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaaggtcc 28680
agaaaaagcc acagaagttt catggtattg ttattttaat gaatcagatg tatctactga 28740
actctgtgga aacaataaca aaaaaaatga gagcattact ctcatcaagt ttcaatgtgg 28800

atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac 28860
agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag 28920
aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat 28980
ttttgccatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat 29040
tcccaaatcc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatcgcctt 29100
gtgcatggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca 29160
cttactaagt gttgaatttt aatttttag aaccatgaag atcctaggcc ttttaatttt 29220

ttctatcatt acctctgctc tatgcaattc tgacaatgag gacgttactg tcgttgtcgg 29280

atcaaattat acactgaaag gtccagcgaa gggtatgctt tcgtggtatt gctattttgg 29340

atctgacact acagaaactg aattatgcaa tcttaagaat ggcaaaattc aaaattctaa 29400

aattaacaat tatatatgca atggtactga tctgatactc ctcaatatca cgaaatcata 29460

tgctggcagt tacacctgcc ctggagatga tgctgacagt atgatttttt acaaagtaac 29520

tgttgttgat cccactactc cacctccacc caccacaact actcacacca cacacacaga 29580

tcaaaccgca gcagaggagg cagcaaagtt agccttgcag gtccaagaca gttcatttgt 29640

tggcattacc cctacacctg atcagcggtg tccggggctg ctagtcagcg gcattgtcgg 29700

tgtgctttcg ggattagcag tcataatcat ctgcatgttc atttttgctt gctgctatag 29760

aaggctttac cgacaaaaat cagacccact gctgaacctc tatgtttaat tttttccaga 29820

gtcatgaagg cagttagcgc tctagttttt tgttctttga ttggcattgt tttttgcaat 29880

cctattccta aagttagctt tattaaagat gtgaatgtta ctgaggggggg caatgtgaca 29940

ctggtaggtg tagagggtgc tgaaaacacc acctggacaa aataccacct caatgggtgg 30000

aaagatattt gcaattggag tgtattagtt tatacatgtg agggagttaa tcttaccatt 30060

gtcaatgcca cctcagctca aaatggtaga attcaaggac aaagtgtcag tgtatctaat 30120

gggtatttta cccaacatac ttttatctat gacgttaaag tcataccact gcctacgcct 30180

agcccaccta gcactaccac acagacaacc cacactacac agacaaccac atacagtaca 30240

ttaaatcagc ctaccaccac tacagcagca gaggttgcca gctcgtctgg ggtccgagtg 30300

gcatttttga tgttggcccc atctagcagt cccactgcta gtaccaatga gcagactact 30360

gaatttttgt ccactgtcga gagccacacc acagctacct ccagtgcctt ctctagcacc 30420

gccaatctct cctcgctttc ctctacacca atcagtcccg ctactactcc tagccccgct 30480

cctcttccca ctcccctgaa gcaaacagac ggcggcatgc aatggcagat caccctgctc 30540

attgtgatcg ggttggtcat cctggccgtg ttgctctact acatcttctg ccgccgcatt 30600

cccaacgcgc accgcaagcc ggtctacaag cccatcattg tcgggcagcc ggagccgctt 30660

caggtggaag ggggtctaag gaatcttctc ttctctttta cagtatggtg attgaactat 30720

gattcctaga caattcttga tcactattct tatctgcctc ctccaagtct gtgccaccct 30780

cgctctggtg gccaacgcca gtccagactg tattgggccc ttcgcctcct acgtgctctt 30840

tgccttcacc acctgcatct gctgctgtag catagtctgc ctgcttatca ccttcttcca 30900

gttcattgac tggatctttg tgcgcatcgc ctacctgcgc caccaccccc agtaccgcga 30960

ccagcgagtg gcgcggctgc tcaggctcct ctgataagca tgcgggctct gctacttctc 31020

gcgcttctgc tgttagtgct cccccgtccc gtcgaccccc ggtcccccac ccagtccccc 31080

gaggaggtcc gcaaatgcaa attccaagaa ccctggaaat tcctcaaatg ctaccgccaa 31140

aaatcagaca tgcatcccag ctggatcatg atcattggga tcgtgaacat tctggcctgc 31200

accctcatct cctttgtgat ttacccctgc tttgactttg gttggaactc gccagaggcg 31260

ctctatctcc cgcctgaacc tgacacacca ccacagcaac ctcaggcaca cgcactacca 31320

ccactacagc ctaggccaca atacatgccc atattagact atgaggccga gccacagcga 31380

cccatgctcc ccgctattag ttacttcaat ctaaccggcg gagatgactg acccactggc 31440

caacaacaac gtcaacgacc ttctcctgga catggacggc cgcgcctcgg agcagcgact 31500

cgcccaactt cgcattcgcc agcagcagga gagagccgtc aaggagctgc aggatgcggt 31560

ggccatccac cagtgcaaga gaggcatctt ctgcctggtg aaacaggcca agatctccta 31620

cgaggtcact ccaaacgacc atcgcctctc ctacgagctc ctgcagcagc gccagaagtt 31680

cacctgcctg gtcggagtca accccatcgt catcacccag cagtctggcg ataccaaggg 31740

gtgcatccac tgctcctgcg actcccccga ctgcgtccac actctgatca agaccctctg 31800

cggcctccgc gacctcctcc ccatgaacta atcacccccc tatccagtga aataaagatc 31860

atattgatga tgattttaca gaaataaaaa ataatcattt gatttgaaat aaagatacaa 31920

tcatattgat gatttgagtt taacaaaaaa ataaagaatc acttacttga aatctgatac 31980

caggtctctg tccatgtttt ctgccaacac cacttcactc ccctcttccc agctctggta 32040

ctgcaggccc cggcgggctg caaacttcct ccacacgctg aaggggatgt caaattcctc 32100

ctgtccctca atcttcattt tatcttctat cagatgtcca aaaagcgcgt ccgggtggat 32160

gatgacttcg accccgtcta cccctacgat gcagacaacg caccgaccgt gcccttcatc 32220

aacccccccc tcgtctcttc agatggattc caagagaagc ccctgggggt gttgtccctg 32280

cgactggccg accccgtcac caccaagaac ggggaaatca ccctcaagct gggagagggg 32340

gtggacctcg attcctcggg aaaactcatc tccaacacgg ccaccaaggc cgccgccccc 32400

ctcagttttt ccaacaacac catttccctt aacatggatc acccctttta cactaaagat 32460

ggaaaattat ccttacaagt ttctccacca ttaaatatac tgagaacaag cattctaaac 32520

acactagctt taggttttgg atcaggttta ggactccgtg gctctgcctt ggcagtacag 32580

ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt 32640

ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa 32700

tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt 32760

acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc 32820

tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg 32880

acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca 32940

ctttgcttga ctaaatgtgg tagtcaaata ctggccactg tgtcagtctt agttgtagga 33000

agtggaaacc taaaccccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt 33060
gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg 33120
cagggagata gcatagatgg cactccatat accaatgctg taggattcat gcccaattta 33180
aaagcttatc caaagtcaca aagttctact actaaaaata atatagtagg gcaagtatac 33240
atgaatggag atgtttcaaa acctatgctt ctcactataa ccctcaatgg tactgatgac 33300
agcaacagta catattcaat gtcattttca tacacctgga ctaatggaag ctatgttgga 33360
gcaacatttg gggctaactc ttataccttc tcatacatcg cccaagaatg aacactgtat 33420

cccaccctgc atgccaaccc ttcccacccc actctgtgga acaaactctg aaacacaaaa 33480
taaaataaag ttcaagtgtt ttattgattc aacagtttta caggattcga gcagttattt 33540
ttcctccacc ctcccaggac atggaataca ccaccctctc cccccgcaca gccttgaaca 33600
tctgaatgcc attggtgatg gacatgcttt tggtctccac gttccacaca gtttcagagc 33660
gagccagtct cgggtcggtc agggagatga aaccctccgg gcactcccgc atctgcacct 33720
cacagctcaa cagctgagga ttgtcctcgg tggtcgggat cacggttatc tggaagaagc 33780
agaagagcgg cggtgggaat catagtccgc gaacgggatc ggccggtggt gtcgcatcag 33840

gccccgcagc agtcgctgcc gccgccgctc cgtcaagctg ctgctcaggg ggtccgggtc 33900
cagggactcc ctcagcatga tgcccacggc cctcagcatc agtcgtctgg tgcggcgggc 33960
gcagcagcgc atgcggatct cgctcaggtc gctgcagtac gtgcaacaca gaaccaccag 34020
gttgttcaac agtccatagt tcaacacgct ccagccgaaa ctcatcgcgg gaaggatgct 34080
acccacgtgg ccgtcgtacc agatcctcag gtaaatcaag tggtgccccc tccagaacac 34140
gctgcccacg tacatgatct ccttgggcat gtggcggttc accacctccc ggtaccacat 34200
caccctctgg ttgaacatgc agccccggat gatcctgcgg aaccacaggg ccagcaccgc 34260

cccgcccgcc atgcagcgaa gagaccccgg gtcccggcaa tggcaatgga ggacccaccg 34320
ctcgtacccg tggatcatct gggagctgaa caagtctatg ttggcacagc acaggcatat 34380
gctcatgcat ctcttcagca ctctcaactc ctcgggggtc aaaaccatat cccagggcac 34440
ggggaactct tgcaggacag cgaaccccgc agaacagggc aatcctcgca cagaacttac 34500
attgtgcatg gacagggtat cgcaatcagg cagcaccggg tgatcctcca ccagagaagc 34560
gcgggtctcg gtctcctcac agcgtggtaa gggggccggc cgatacgggt gatggcggga 34620
cgcggctgat cgtgttcgcg accgtgtcat gatgcagttg ctttcggaca ttttcgtact 34680

tgctgtagca gaacctggtc cgggcgctgc acaccgatcg ccggcggcgg tctcggcgct 34740

tggaacgctc ggtgttgaaa ttgtaaaaca gccactctct cagaccgtgc agcagatcta 34800

gggcctcagg agtgatgaag atcccatcat gcctgatggc tctgatcaca tcgaccaccg 34860

tggaatgggc cagacccagc cagatgatgc aattttgttg ggtttcggtg acggcggggg 34920

agggaagaac aggaagaacc atgattaact tttaatccaa acggtctcgg agtacttcaa 34980

aatgaagatc gcggagatgg cacctctcgc ccccgctgtg ttggtggaaa ataacagcca 35040

ggtcaaaggt gatacggttc tcgagatgtt ccacggtggc ttccagcaaa gcctccacgc 35100

gcacatccag aaacaagaca atagcgaaag cgggagggtt ctctaattcc tcaatcatca 35160

tgttacactc ctgcaccatc cccagataat tttcattttt ccagccttga atgattcgaa 35220

ctagttcctg aggtaaatcc aagccagcca tgataaagag ctcgcgcaga gcgccctcca 35280

ccggcattct taagcacacc ctcataattc caagatattc tgctcctggt tcacctgcag 35340

cagattgaca agcggaatat caaaatctct gccgcgatcc ctgagctcct ccctcagcaa 35400

taactgtaag tactctttca tatcctctcc gaaattttta gccataggac caccaggaat 35460

aagattaggg caagccacag tacagataaa ccgaagtcct ccccagtgag cattgccaaa 35520

tgcaagactg ctataagcat gctggctaga cccggtgata tcttccagat aactggacag 35580

aaaatcgccc aggcaatttt taagaaaatc aacaaaagaa aaatcctcca ggtggacgtt 35640

tagagcctcg ggaacaacga tgaagtaaat gcaagcggtg cgttccagca tggttagtta 35700

gctgatctgt agaaaaaaca aaaatgaaca ttaaaccatg ctagcctggc gaacaggtgg 35760

gtaaatcgtt ctctccagca ccaggcaggc cacggggtct ccggcgcgac cctcgtaaaa 35820

attgtcgcta tgattgaaaa ccatcacaga gagacgttcc cggtggccgg cgtgaatgat 35880

tcgacaagat gaatacaccc ccggaacatt ggcgtccgcg agtgaaaaaa agcgcccgag 35940

gaagcaataa ggcactacaa tgctcagtct caagtccagc aaagcgatgc catgcggatg 36000

aagcacaaaa ttctcaggtg cgtacaaaat gtaattactc ccctcctgca caggcagcaa 36060

agcccccgat ccctccaggt acacatacaa agcctcagcg tccatagctt accgagcagc 36120

agcacacaac aggcgcaaga gtcagagaaa ggctgagctc taacctgtcc acccgctctc 36180

tgctcaatat atagcccaga tctacactga cgtaaaggcc aaagtctaaa aatacccgcc 36240

aaataatcac acacgcccag cacacgccca gaaaccggtg acacactcaa aaaaatacgc 36300

gcacttcctc aaacgcccaa aactgccgtc atttccgggt tcccacgcta cgtcatcaaa 36360

acacgacttt caaattccgt cgaccgttaa aaacgtcacc cgccccgccc ctaacggtcg 36420

cccgtctctc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt 36480

aacgcgcaca aaaagtttga ggtatattat tgatgatgg 36519

<210> 11

<211> 31867

<212> DNA

<213> Artificial Sequence

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

<400> 11

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg 1320

acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca 1380

agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg 1440

tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac atgaagcagc 1500

acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca 1560

aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga 1620

accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc 1680

tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca 1740

tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc 1800

actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac aaccactacc 1860

tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc 1920

tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag 1980

tttaaactcc catttaaatg tgagggttaa tgcttcgagc agacatgata agatacattg 2040

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 2100

gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 2160

attgcattca ttttatgttt caggttcagg gggagatgtg ggaggttttt taaagcaagt 2220

aaaacctcta caaatgtggt aaaataacta taacggtcct aaggtagcga gtgagtagtg 2280

ttctggggcg ggggaggacc tgcatgaggg ccagaataac tgaaatctgt gcttttctgt 2340

gtgttgcagc agcatgagcg gaagcggctc ctttgaggga ggggtattca gcccttatct 2400

gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat ccacggtgga 2460

cggccggccc gtgcagcccg cgaactcttc aaccctgacc tatgcaaccc tgagctcttc 2520

gtcgttggac gcagctgccg ccgcagctgc tgcatctgcc gccagcgccg tgcgcggaat 2580

ggccatgggc gccggctact acggcactct ggtggccaac tcgagttcca ccaataatcc 2640

cgccagcctg aacgaggaga agctgttgct gctgatggcc cagctcgagg ccttgaccca 2700

gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gagcagacgc gggccgcggt 2760

tgccacggtg aaatccaaat aaaaaatgaa tcaataaata aacggagacg gttgttgatt 2820

ttaacacaga gtctgaatct ttatttgatt tttcgcgcgc ggtaggccct ggaccaccgg 2880

tctcgatcat tgagcacccg gtggatcttt tccaggaccc ggtagaggtg ggcttggatg 2940

ttgaggtaca tgggcatgag cccgtcccgg gggtggaggt agctccattg cagggcctcg 3000

tgctcggggg tggtgttgta aatcacccag tcatagcagg ggcgcagggc atggtgttgc 3060

acaatatctt tgaggaggag actgatggcc acgggcagcc ctttggtgta ggtgtttaca 3120

aatctgttga gctgggaggg atgcatgcgg ggggagatga ggtgcatctt ggcctggatc 3180

ttgagattgg cgatgttacc gcccagatcc cgcctggggt tcatgttgtg caggaccacc 3240

agcacggtgt atccggtgca cttggggaat ttatcatgca acttggaagg gaaggcgtga 3300

aagaatttgg cgacgccttt gtgcccgccc aggttttcca tgcactcatc catgatgatg 3360

gcgatgggcc cgtgggcggc ggcctgggca aagacgtttc gggggtcgga cacatcatag 3420

ttgtggtcct gggtgaggtc atcataggcc attttaatga atttggggcg gagggtgccg 3480

gactggggga caaaggtacc ctcgatcccg ggggcgtagt tcccctcaca gatctgcatc 3540

tcccaggctt tgagctcgga gggggggatc atgtccacct gcggggcgat aaagaacacg 3600

gtttccgggg cgggggagat gagctgggcc gaaagcaagt tccggagcag ctgggacttg 3660

ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgagggag 3720

agacagctgc cgtcctcccg gaggaggggg gccacctcgt tcatcatctc gcgcacgtgc 3780

atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg 3840

agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt 3900

tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc 3960

agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg 4020

cgtccagcgc agccagggtc cggtccttcc agggtcgcag cgtccgcgtc agggtggtct 4080

ccgtcacggt gaaggggtgc gcgccgggct gggcgcttgc gagggtgcgc ttcaggctca 4140

tccggctggt cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga 4200

ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcggagc ttacctttgg 4260

aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga 4320

agacggactc gggggcgtag gcgtccgcgc cgcagtgggc gcagacggtc tcgcactcca 4380

cgagccaggt gaggtcgggc tggtcggggt caaaaaccag tttcccgccg ttctttttga 4440

tgcgtttctt acctttggtc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt 4500

ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cggtgtgccg cggtcctcct 4560

cgtagaggaa ccccgcccac tccgagacga aagcccgggt ccaggccagc acgaaggagg 4620

ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc cacctttcc agggtatgca 4680

aacacatgtc cccctcgtcc acatccagga aggtgattgg cttgtaagtg taggccacgt 4740

gaccgggggt cccggccggg ggggtataaa agggtgcggg tccctgctcg tcctcactgt 4800

cttccggatc gctgtccagg agcgccagct gttggggtag gtattccctc tcgaaggcgg 4860

gcatgacctc ggcactcagg ttgtcagttt ctagaaacga ggaggatttg atattgacgg 4920

tgccggcgga gatgcctttc aagagcccct cgtccatctg gtcagaaaag acgatctttt 4980

tgttgtcgag cttggtggcg aaggagccgt agagggcgtt ggagaggagc ttggcgatgg 5040

agcgcatggt ctggtttttt tccttgtcgg cgcgctcctt ggcggcgatg ttgagctgca 5100

cgtactcgcg cgccacgcac ttccattcgg ggaagacggt ggtcagctcg tcgggcacga 5160

ttctgacctg ccagccccga ttatgcaggg tgatgaggtc cacactggtg gccacctcgc 5220
cgcgcagggg ctcattagtc cagcagaggc gtccgccctt gcgcgagcag aaggggggca 5280
gggggtccag catgacctcg tcgggggggt cggcatcgat ggtgaagatg ccgggcagga 5340
ggtcggggtc aaagtagctg atggaagtgg ccagatcgtc cagggcagct tgccattcgc 5400
gcacggccag cgcgctctcg tagggactga ggggcgtgcc ccagggcatg ggatgggtaa 5460
gcgcggaggc gtacatgccg cagatgtcgt agacgtagag gggctcctcg aggatgccga 5520
tgtaggtggg gtagcagcgc cccccgcgga tgctggcgcg cacgtagtca tacagctcgt 5580

gcgaggggggc gaggagcccc gggcccaggt tggtgcgact gggcttttcg gcgcggtaga 5640
cgatctggcg gaaaatggca tgcgagttgg aggagatggt gggcctttgg aagatgttga 5700
agtgggcgtg gggcagtccg accgagtcgc ggatgaagtg ggcgtaggag tcttgcagct 5760
tggcgacgag ctcggcggtg actaggacgt ccagagcgca gtagtcgagg gtctcctgga 5820
tgatgtcata cttgagctgt ccctttttgtt tccacagctc gcggttgaga aggaactctt 5880
cgcggtcctt ccagtactct tcgaggggga acccgtcctg atctgcacgg taagagccta 5940
gcatgtagaa ctggttgacg gccttgtagg cgcagcagcc cttctccacg gggagggcgt 6000

aggcctgggc ggccttgcgc agggaggtgt gcgtgagggc gaaagtgtcc ctgaccatga 6060
ccttgaggaa ctggtgcttg aagtcgatat cgtcgcagcc cccctgctcc cagagctgga 6120
agtccgtgcg cttcttgtag gcggggttgg gcaaagcgaa agtaacatcg ttgaagagga 6180
tcttgcccgc gcggggcata aagttgcgag tgatgcggaa aggttggggc acctcggccc 6240
ggttgttgat gacctgggcg gcgagcacga tctcgtcgaa gccgttgatg ttgtggccca 6300
cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct 6360
cgtaggtgag ctcgtcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat 6420

gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcggtt tgcagacggt 6480
cccggtactg acggaactgc tgcccgacgg ccatttttttc gggggtgacg cagtagaagg 6540
tgcgggggtc cccgtgccag cgatcccatt tgagctggag ggcgagatcg agggcgagct 6600
cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc 6660
cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcggtgc 6720
gaggatgcga gccgatgggg aagaactgga tctcctgcca ccaattggag gaatggctgt 6780
tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca 6840

agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag 6900

ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta 6960

ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcatg ctgacgagcc 7020

cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc 7080

gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggcagcggcg 7140

gcgcgcggtt gacttgcagg agtttttcca gggcgcgcgg gaggtccaga tggtacttga 7200

tctccaccgc gccattggtg gcgacgtcga tggcttgcag ggtcccgtgc ccctggggtg 7260

tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcggt gcctcttcca 7320

tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggcggctcgg ggcccggagg 7380

caggggcggc aggggcacgt cggcgccgcg cgcgggtagg ttctggtact gcgcccggag 7440

aagactggcg tgagcgacga cgcgacggtt gacgtcctgg atctgacgcc tctgggtgaa 7500

ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc 7560

gttgacggcg gcctgccgca ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat 7620

ctcggtcatg aactgctcga tctcctcctc ttgaaggtct ccgcggccgg cgcgctccac 7680

ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc 7740

ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac 7800

ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta 7860

gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg 7920

gagcggcatc tcgctgacgt cgcccagcgc ctccaaacgt tccatggcct cgtaaaagtc 7980

cacggcgaag ttgaaaaact gggagttgcg cgccgagacg gtcaactcct cctccagaag 8040

acggatgagc tcggcgatgg tggcgcgcac ctcgcgctcg aaggcccccg ggagttcctc 8100

cacttcctct tcttcctcct ccactaacat ctcttctact tcctcctcag gcggcagtgg 8160

tggcgggggg gggggcctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgctc 8220

gatggtctcg ccgcgccggc gtcgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg 8280

ccgcagcgtg aagacgccgc cgcgcatctc caggtggccg ggggggtccc cgttgggcag 8340

ggagagggcg ctgacgatgc atcttatcaa ttgccccgta gggactccgc gcaaggacct 8400

gagcgtctcg agatccacgg gatctgaaaa ccgctgaacg aaggcttcga gccagtcgca 8460

gtcgcaaggt aggctgagca cggtttcttc tggcgggtca tgttggttgg gagcggggcg 8520

ggcgatgctg ctggtgatga agttgaaata ggcggttctg agacggcgga tggtggcgag 8580

gagcaccagg tctttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc 8640

gtggtcctga cacctggcca ggtccttgta gtagtcctgc atgagccgct ccacgggcac 8700

ctcctcctcg cccgcgcggc cgtgcatgcg cgtgagcccg aagccgcgct ggggctggac 8760

gagcgccagg tcggcgacga cgcgctcggc gaggatggct tgctggatct gggtgagggt 8820

ggtctggaag tcatcaaagt cgacgaagcg gtggtaggct ccggtgttga tggtgtagga 8880

gcagttggcc atgacggacc agttgacggt ctggtggccc ggacgcacga gctcgtggta 8940

cttgaggcgc gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggtgc gcaccaggta 9000

ctggtagccg atgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc 9060

gggggcgccg ggcgcgaggt cctcgagcat ggtgcggtgg tagccgtaga tgtacctgga 9120

catccaggtg atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca 9180

gatgttgcgc agcggcagga agtagttcat ggtgggcacg gtctggcccg tgaggcgcgc 9240

gcagtcgtgg atgctctata cgggcaaaaa cgaaagcggt cagcggctcg actccgtggc 9300

ctggaggcta agcgaacggg ttgggctgcg cgtgtacccc ggttcgaatc tcgaatcagg 9360

ctggagccgc agctaacgtg gtattggcac tcccgtctcg acccaagcct gcaccaaccc 9420

tccaggatac ggaggcgggt cgttttgcaa ctttttttg gaggccggat gagactagta 9480

agcgcggaaa gcggccgacc gcgatggctc gctgccgtag tctggagaag aatcgccagg 9540

gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg 9600

gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc 9660

cctcttttgt tttgtttgtt tttgccagat gcatcccgta ctgcggcaga tgcgccccca 9720

ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca 9780

gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca 9840

ccagctggcc ttggaagagg gcgaggggct ggcgcgcctg ggggcgtcgt cgccggagcg 9900

gcacccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgccca agcagaacct 9960

gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggcccggt tccacgcggg 10020

gcgggagctg cggcgcggcc tggaccgaaa gagggtgctg agggacgagg atttcgaggc 10080

ggacgagctg acggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac 10140

ggcgtacgag cagaccgtga aggaggagag caacttccaa aaatccttca acaaccacgt 10200

gcgcaccctg atcgcgcgcg aggaggtgac cctgggcctg atgcacctgt gggacctgct 10260

ggaggccatc gtgcagaacc ccaccagcaa gccgctgacg gcgcagctgt tcctggtggt 10320

gcagcatagt cgggacaacg aagcgttcag ggaggcgctg ctgaatatca ccgagcccga 10380

gggccgctgg ctcctggacc tggtgaacat tctgcagagc atcgtggtgc aggagcgcgg 10440

gctgccgctg tccgagaagc tggcggccat caacttctcg gtgctgagtt tgggcaagta 10500

ctacgctagg aagatctaca agaccccgta cgtgcccata gacaaggagg tgaagatcga 10560

cgggttttac atgcgcatga ccctgaaagt gctgaccctg agcgacgatc tgggggtgta 10620

ccgcaacgac aggatgcacc gtgcggtgag cgccagcagg cggcgcgagc tgagcgacca 10680
ggagctgatg catagtctgc agcgggccct gaccggggcc gggaccgagg gggagagcta 10740
ctttgacatg ggcgcggacc tgcactggca gcccagccgc cgggccttgg aggcggcggc 10800
aggaccctac gtagaagagg tggacgatga ggtggacgag gagggcgagt acctggaaga 10860
ctgatggcgc gaccgtattt ttgctagatg caacaacaac agccacctcc tgatcccgcg 10920
atgcgggcgg cgctgcagag ccagccgtcc ggcattaact cctcggacga ttggacccag 10980
gccatgcaac gcatcatggc gctgacgacc cgcaaccccg aagcctttag acagcagccc 11040

caggccaacc ggctctcggc catcctggag gccgtggtgc cctcgcgctc caaccccacg 11100
cacgagaagg tcctggccat cgtgaacgcg ctggtggaga acaaggccat ccgcggcgac 11160
gaggccggcc tggtgtacaa cgcgctgctg gagcgcgtgg cccgctacaa cagcaccaac 11220
gtgcagacca acctggaccg catggtgacc gacgtgcgcg aggccgtggc ccagcgcgag 11280
cggttccacc gcgagtccaa cctgggatcc atggtggcgc tgaacgcctt cctcagcacc 11340
cagcccgcca acgtgccccg gggccaggag gactacacca acttcatcag cgccctgcgc 11400
ctgatggtga ccgaggtgcc ccagagcgag gtgtaccagt ccgggccgga ctacttcttc 11460

cagaccagtc gccagggctt gcagaccgtg aacctgagcc aggctttcaa gaacttgcag 11520
ggcctgtggg gcgtgcaggc cccggtcggg gaccgcgcga cggtgtcgag cctgctgacg 11580
ccgaactcgc gcctgctgct gctgctggtg gcccccttca cggacagcgg cagcatcaac 11640
cgcaactcgt acctgggcta cctgattaac ctgtaccgcg aggccatcgg ccaggcgcac 11700
gtggacgagc agacctacca ggagatcacc cacgtgagcc gcgccctggg ccaggacgac 11760
ccgggcaacc tggaagccac cctgaacttt ttgctgacca accggtcgca gaagatcccg 11820
ccccagtacg cgctcagcac cgaggaggag cgcatcctgc gttacgtgca gcagagcgtg 11880

ggcctgttcc tgatgcagga gggggccacc cccagcgccg cgctcgacat gaccgcgcgc 11940
aacatggagc ccagcatgta cgccagcaac cgcccgttca tcaataaact gatggactac 12000
ttgcatcggg cggccgccat gaactctgac tatttcacca acgccatcct gaatccccac 12060
tggctcccgc cgccggggtt ctacacgggc gagtacgaca tgcccgaccc caatgacggg 12120
ttcctgtggg acgatgtgga cagcagcgtg ttctcccccc gaccgggtgc taacgagcgc 12180
cccttgtgga agaaggaagg cagcgaccga cgcccgtcct cggcgctgtc cggccgcgag 12240
ggtgctgccg cggcggtgcc cgaggccgcc agtcctttcc cgagcttgcc cttctcgctg 12300

aacagtatcc gcagcagcga gctgggcagg atcacgcgcc cgcgcttgct gggcgaagag 12360

gagtacttga atgactcgct gttgagaccc gagcgggaga agaacttccc caataacggg 12420

atagaaagcc tggtggacaa gatgagccgc tggaagacgt atgcgcagga gcacagggac 12480

gatccccggg cgtcgcaggg ggccacgagc cggggcagcg ccgcccgtaa acgccggtgg 12540

cacgacaggc agcggggaca gatgtgggac gatgaggact ccgccgacga cagcagcgtg 12600

ttggacttgg gtgggagtgg taacccgttc gctcacctgc gcccccgtat cgggcgcatg 12660

atgtaagaga aaccgaaaat aaatgatact caccaaggcc atggcgacca gcgtgcgttc 12720

gtttcttctc tgttgttgtt gtatctagta tgatgaggcg tgcgtacccg gagggtcctc 12780

ctccctcgta cgagagcgtg atgcagcagg cgatggcggc ggcggcgatg cagcccccgc 12840

tggaggctcc ttacgtgccc ccgcggtacc tggcgcctac ggaggggcgg aacagcattc 12900

gttactcgga gctggcaccc ttgtacgata ccacccggtt gtacctggtg gacaacaagt 12960

cggcggacat cgcctcgctg aactaccaga acgaccacag caacttcctg accaccgtgg 13020

tgcagaacaa tgacttcacc cccacggagg ccagcaccca gaccatcaac tttgacgagc 13080

gctcgcggtg gggcggccag ctgaaaacca tcatgcacac caacatgccc aacgtgaacg 13140

agttcatgta cagcaacaag ttcaaggcgc gggtgatggt ctcccgcaag acccccaatg 13200

gggtgacagt gacagaggat tatgatggta gtcaggatga gctgaagtat gaatgggtgg 13260

aatttgagct gcccgaaggc aacttctcgg tgaccatgac catcgacctg atgaacaacg 13320

ccatcatcga caattacttg gcggtggggc ggcagaacgg ggtgctggag agcgacatcg 13380

gcgtgaagtt cgacactagg aacttcaggc tgggctggga ccccgtgacc gagctggtca 13440

tgcccggggt gtacaccaac gaggctttcc atcccgatat tgtcttgctg cccggctgcg 13500

gggtggactt caccgagagc cgcctcagca acctgctggg cattcgcaag aggcagccct 13560

tccaggaagg cttccagatc atgtacgagg atctggaggg gggcaacatc cccgcgctcc 13620

tggatgtcga cgcctatgag aaaagcaagg aggatgcagc agctgaagca actgcagccg 13680

tagctaccgc ctctaccgag gtcaggggcg ataattttgc aagcgccgca gcagtggcag 13740

cggccgaggc ggctgaaacc gaaagtaaga tagtcattca gccggtggag aaggatagca 13800

agaacaggag ctacaacgta ctaccggaca agataaacac cgcctaccgc agctggtacc 13860

tagcctacaa ctatggcgac cccgagaagg gcgtgcgctc ctggacgctg ctcaccacct 13920

cggacgtcac ctgcggcgtg gagcaagtct actggtcgct gcccgacatg atgcaagacc 13980

cggtcacctt ccgctccacg cgtcaagtta gcaactaccc ggtggtgggc gccgagctcc 14040

tgcccgtcta ctccaagagc ttcttcaacg agcaggccgt ctactcgcag cagctgcgcg 14100

ccttcacctc gcttacgcac gtcttcaacc gcttccccga gaaccagatc ctcgtccgcc 14160

cgcccgcgcc caccattacc accgtcagtg aaaacgttcc tgctctcaca gatcacggga 14220

ccctgccgct gcgcagcagt atccggggag tccagcgcgt gaccgttact gacgccagac 14280

gccgcacctg cccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtcctctcga 14340

gccgcacctt ctaaatgtcc attctcatct cgcccagtaa taacaccggt tggggcctgc 14400

gcgcgcccag caagatgtac ggaggcgctc gccaacgctc cacgcaacac cccgtgcgcg 14460

tgcgcgggca cttccgcgct ccctggggcg ccctcaaggg ccgcgtgcgg tcgcgcacca 14520

ccgtcgacga cgtgatcgac caggtggtgg ccgacgcgcg caactacacc cccgccgccg 14580

cgcccgtctc caccgtggac gccgtcatcg acagcgtggt ggccgacgcg cgccggtacg 14640

cccgcgccaa gagccggcgg cggcgcatcg cccggcggca ccggagcacc cccgccatgc 14700

gcgcggcgcg agccttgctg cgcagggcca ggcgcacggg acgcagggcc atgctcaggg 14760

cggccagacg cgcggcttca ggcgccagcg ccggcaggac ccggagacgc gcggccacgg 14820

cggcggcagc ggccatcgcc agcatgtccc gcccgcggcg agggaacgtg tactgggtgc 14880

gcgacgccgc caccggtgtg cgcgtgcccg tgcgcacccg ccccccctcgc acttgaagat 14940

gttcacttcg cgatgttgat gtgtcccagc ggcgaggagg atgtccaagc gcaaattcaa 15000

ggaagagatg ctccaggtca tcgcgcctga gatctacggc cctgcggtgg tgaaggagga 15060

aagaaagccc cgcaaaatca agcgggtcaa aaaggacaaa aaggaagaag aaagtgatgt 15120

ggacggattg gtggagtttg tgcgcgagtt cgcccccggg cggcgcgtgc agtggcgcgg 15180

gcggaaggtg caaccggtgc tgagacccgg caccaccgtg gtcttcacgc ccggcgagcg 15240

ctccggcacc gcttccaagc gctcctacga cgaggtgtac ggggatgatg atattctgga 15300

gcaggcggcc gagcgcctgg gcgagtttgc ttacggcaag cgcagccgtt ccgcaccgaa 15360

ggaagaggcg gtgtccatcc cgctggacca cggcaacccc acgccgagcc tcaagcccgt 15420

gaccttgcag caggtgctgc cgaccgcggc gccgcgccgg gggttcaagc gcgagggcga 15480

ggatctgtac cccaccatgc agctgatggt gcccaagcgc cagaagctgg aagacgtgct 15540

ggagaccatg aaggtggacc cggacgtgca gcccgaggtc aaggtgcggc ccatcaagca 15600

ggtggccccg ggcctgggcg tgcagaccgt ggacatcaag attcccacgg agcccatgga 15660

aacgcagacc gagcccatga tcaagcccag caccagcacc atggaggtgc agacggatcc 15720

ctggatgcca tcggctccta gtcgaagacc ccggcgcaag tacggcgcgg ccagcctgct 15780

gatgcccaac tacgcgctgc atccttccat catccccacg ccgggctacc gcggcacgcg 15840

cttctaccgc ggtcatacca gcagccgccg ccgcaagacc accactcgcc gccgccgtcg 15900

ccgcaccgcc gctgcaacca cccctgccgc cctggtgcgg agagtgtacc gccgcggccg 15960

cgcacctctg accctgccgc gcgcgcgcta ccacccgagc atcgccattt aaactttcgc 16020

ctgctttgca gatcaatggc cctcacatgc cgccttcgcg ttcccattac gggctaccga 16080

ggaagaaaac cgcgccgtag aaggctggcg gggaacggga tgcgtcgcca ccaccaccgg 16140

cggcggcgcg ccatcagcaa gcggttgggg ggaggcttcc tgcccgcgct gatccccatc 16200

atcgccgcgg cgatcggggc gatccccggc attgcttccg tggcggtgca ggcctctcag 16260

cgccactgag acacacttgg aaacatcttg taataaacca atggactctg acgctcctgg 16320

tcctgtgatg tgttttcgta gacagatgga agacatcaat ttttcgtccc tggctccgcg 16380

acacggcacg cggccgttca tgggcacctg gagcgacatc ggcaccagcc aactgaacgg 16440

gggcgccttc aattggagca gtctctggag cgggcttaag aatttcgggt ccacgcttaa 16500

aacctatggc agcaaggcgt ggaacagcac cacagggcag gcgctgaggg ataagctgaa 16560

agagcagaac ttccagcaga aggtggtcga tgggctcgcc tcgggcatca acggggtggt 16620

ggacctggcc aaccaggccg tgcagcggca gatcaacagc cgcctggacc cggtgccgcc 16680

cgccggctcc gtggagatgc cgcaggtgga ggaggagctg cctcccctgg acaagcgggg 16740

cgagaagcga ccccgccccg atgcggagga gacgctgctg acgcacacgg acgagccgcc 16800

cccgtacgag gaggcggtga aactgggtct gcccaccacg cggcccatcg cgcccctggc 16860

caccggggtg ctgaaacccg aaaagcccgc gaccctggac ttgcctcctc cccagccttc 16920

ccgcccctct acagtggcta agcccctgcc gccggtggcc gtggcccgcg cgcgacccgg 16980

gggcaccgcc cgccctcatg cgaactggca gagcactctg aacagcatcg tgggtctggg 17040

agtgcagagt gtgaagcgcc gccgctgcta ttaaacctac cgtagcgctt aacttgcttg 17100

tctgtgtgtg tatgtattat gtcgccgccg ccgctgtcca ccagaaggag gagtgaagag 17160

gcgcgtcgcc gagttgcaag atggccaccc catcgatgct gccccagtgg gcgtacatgc 17220

acatcgccgg acaggacgct tcggagtacc tgagtccggg tctggtgcag tttgcccgcg 17280

ccacagacac ctacttcagt ctggggaaca agtttaggaa ccccacggtg gcgcccacgc 17340

acgatgtgac caccgaccgc agccagcggc tgacgctgcg cttcgtgccc gtggaccgcg 17400

aggacaacac ctactcgtac aaagtgcgct acacgctggc cgtgggcgac aaccgcgtgc 17460

tggacatggc cagcacctac tttgacatcc gcggcgtgct ggatcggggc cctagcttca 17520

aaccctactc cggcaccgcc tacaacagtc tggcccccaa gggagcaccc aacacttgtc 17580

agtggacata taaagccgat ggtgaaactg ccacagaaaa aacctataca tatggaaatg 17640

cacccgtgca gggcattaac atcacaaaag atggtattca acttggaact gacaccgatg 17700

atcagccaat ctacgcagat aaaacctatc agcctgaacc tcaagtgggt gatgctgaat 17760

ggcatgacat cactggtact gatgaaaagt atggaggcag agctcttaag cctgatacca 17820

aaatgaagcc ttgttatggt tcttttgcca agcctactaa taaagaagga ggtcaggcaa 17880

atgtgaaaac aggaacaggc actactaaag aatatgacat agacatggct ttctttgaca 17940

acagaagtgc ggctgctgct ggcctagctc cagaaattgt tttgtatact gaaaatgtgg 18000

atttggaaac tccagatacc catattgtat acaaagcagg cacagatgac agcagctctt 18060

ctattaattt gggtcagcaa gccatgccca acagacctaa ctacattggt ttcagagaca 18120

actttatcgg gctcatgtac tacaacagca ctggcaatat gggggtgctg gccggtcagg 18180

cttctcagct gaatgctgtg gttgacttgc aagacagaaa caccgagctg tcctaccagc 18240

tcttgcttga ctctctgggt gacagaaccc ggtatttcag tatgtggaat caggcggtgg 18300

acagctatga tcctgatgtg cgcattattg aaaatcatgg tgtggaggat gaacttccca 18360

actattgttt ccctctggat gctgttggca gaacagatac ttatcaggga attaaggcta 18420

atggaactga tcaaaccaca tggaccaaag atgacagtgt caatgatgct aatgagatag 18480

gcaagggtaa tccattcgcc atggaaatca acatccaagc caacctgtgg aggaacttcc 18540

tctacgccaa cgtggccctg tacctgcccg actcttacaa gtacacgccg gccaatgtta 18600

ccctgcccac caacaccaac acctacgatt acatgaacgg ccgggtggtg gcgccctcgc 18660

tggtggactc ctacatcaac atcggggcgc gctggtcgct ggatcccatg gacaacgtga 18720

acccccttcaa ccaccaccgc aatgcggggc tgcgctaccg ctccatgctc ctgggcaacg 18780

ggcgctacgt gcccttccac atccaggtgc cccagaaatt tttcgccatc aagagcctcc 18840

tgctcctgcc cgggtcctac acctacgagt ggaacttccg caaggacgtc aacatgatcc 18900

tgcagagctc cctcggcaac gacctgcgca cggacggggc ctccatctcc ttcaccagca 18960

tcaacctcta cgccaccttc ttccccatgg cgcacaacac ggcctccacg ctcgaggcca 19020

tgctgcgcaa cgacaccaac gaccagtcct tcaacgacta cctctcggcg gccaacatgc 19080

tctaccccat cccggccaac gccaccaacg tgcccatctc catcccctcg cgcaactggg 19140

ccgccttccg cggctggtcc ttcacgcgtc tcaagaccaa ggagacgccc tcgctgggct 19200

ccgggttcga cccctacttc gtctactcgg gctccatccc ctacctcgac ggcaccttct 19260

acctcaacca caccttcaag aaggtctcca tcaccttcga ctcctccgtc agctggcccg 19320

gcaacgaccg gctcctgacg cccaacgagt tcgaaatcaa gcgcaccgtc gacggcgagg 19380

gctacaacgt ggcccagtgc aacatgacca aggactggtt cctggtccag atgctggccc 19440

actacaacat cggctaccag ggcttctacg tgcccgaggg ctacaaggac cgcatgtact 19500

ccttcttccg caacttccag cccatgagcc gccaggtggt ggacgaggtc aactacaagg 19560

actaccaggc cgtcaccctg gcctaccagc acaacaactc gggcttcgtc ggctacctcg 19620

cgcccaccat gcgccagggc cagccctacc ccgccaacta cccctacccg ctcatcggca 19680

agagcgccgt caccagcgtc acccagaaaa agttcctctg cgacagggtc atgtggcgca 19740

tccccttctc cagcaacttc atgtccatgg gcgcgctcac cgacctcggc cagaacatgc 19800

tctatgccaa ctccgcccac gcgctagaca tgaatttcga agtcgacccc atggatgagt 19860

ccacccttct ctatgttgtc ttcgaagtct tcgacgtcgt ccgagtgcac cagccccacc 19920

gcggcgtcat cgaggccgtc tacctgcgca ccccttctc ggccggtaac gccaccacct 19980

aagctcttgc ttcttgcaag ccatggccgc gggctccggc gagcaggagc tcagggccat 20040

catccgcgac ctgggctgcg ggccctactt cctgggcacc ttcgataagc gcttcccggg 20100

attcatggcc ccgcacaagc tggcctgcgc catcgtcaac acggccggcc gcgagaccgg 20160

gggcgagcac tggctggcct tcgcctggaa cccgcgctcg aacacctgct acctcttcga 20220

ccccttcggg ttctcggacg agcgcctcaa gcagatctac cagttcgagt acgagggcct 20280

gctgcgccgc agcgccctgg ccaccgagga ccgctgcgtc accctggaaa agtccaccca 20340

gaccgtgcag ggtccgcgct cggccgcctg cgggctcttc tgctgcatgt tcctgcacgc 20400

cttcgtgcac tggcccgacc gccccatgga caagaacccc accatgaact tgctgacggg 20460

ggtgcccaac ggcatgctcc agtcgcccca ggtggaaccc accctgcgcc gcaaccagga 20520

ggcgctctac cgcttcctca actcccactc cgcctacttt cgctcccacc gcgcgcgcat 20580

cgagaaggcc accgccttcg accgcatgaa tcaagacatg taaaccgtgt gtgtatgtta 20640

aatgtcttta ataaacagca ctttcatgtt acacatgcat ctgagatgat ttatttagaa 20700

atcgaaaggg ttctgccggg tctcggcatg gcccgcgggc agggacacgt tgcggaactg 20760

gtacttggcc agccacttga actcggggat cagcagtttg ggcagcgggg tgtcggggaa 20820

ggagtcggtc cacagcttcc gcgtcagttg cagggcgccc agcaggtcgg gcgcggagat 20880

cttgaaatcg cagttgggac ccgcgttctg cgcgcgggag ttgcggtaca cggggttgca 20940

gcactggaac accatcaggg ccgggtgctt cacgctcgcc agcaccgtcg cgtcggtgat 21000

gctctccacg tcgaggtcct cggcgttggc catcccgaag ggggtcatct tgcaggtctg 21060

ccttcccatg gtgggcacgc acccgggctt gtggttgcaa tcgcagtgca gggggatcag 21120

catcatctgg gcctggtcgg cgttcatccc cgggtacatg gccttcatga aagcctccaa 21180

ttgcctgaac gcctgctggg ccttggctcc ctcggtgaag aagaccccgc aggacttgct 21240

agagaactgg ttggtggcgc acccggcgtc gtgcacgcag cagcgcgcgt cgttgttggc 21300

cagctgcacc acgctgcgcc cccagcggtt ctgggtgatc ttggcccggt cggggttctc 21360

cttcagcgcg cgctgcccgt tctcgctcgc cacatccatc tcgatcatgt gctccttctg 21420

gatcatggtg gtcccgtgca ggcaccgcag cttgccctcg gcctcggtgc acccgtgcag 21480

ccacagcgcg cacccggtgc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac 21540

gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggtcag 21600
cggaatgccg cggtgctcct cgttgatgta caggtggcag atgcggcggt acacctcgcc 21660
ctgctcgggc atcagctgga agttggcttt caggtcggtc tccacgcggt agcggtccat 21720
cagcatagtc atgatttcca tacccttctc ccaggccgag acgatgggca ggctcatagg 21780
gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtcgctct cgtccagggt 21840
ctcaaagctc cgcttgccgt ccttctcggt gatccgcacc ggggggtagc tgaagcccac 21900
ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag 21960

gaccacatgc ttggtcttgc ggggtttctt cttgggcggc agcggcggcg gagatgttgg 22020
agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc 22080
cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcggaggcg acgggctctc 22140
gccgccgcga cttggcggat ggctggcaga gcccccttccg cgttcgggggg tgcgctcccg 22200
gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac 22260
aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgccg acgagaagca 22320
gcagcagcag aatgaaagct taaccgcccc gccgcccagc cccgccacct ccgacgcggc 22380

cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc 22440
cgcggagcac gaggaggagc tggcagtgcg cttttcacaa gaagagatac accaagaaca 22500
gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta 22560
cctccacctg agcgggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat 22620
cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg 22680
cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac 22740
ctgcgagccc aacccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc 22800

cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac 22860
ccgcgccgac gccctttca acctgggtcc cggcgcccgc ctacctgata tcgcctcctt 22920
ggaagaggtt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc 22980
tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga 23040
caacgcgcgg ctggcggtgc tcaaacgcac ggtcgagctg acccatttcg cctacccggc 23100
tctgaacctg ccccccaaag tcatgagcgc ggtcatggac caggtgctca tcaagcgcgc 23160
gtcgcccatc tccgaggacg agggcatgca agactccgag gagggcaagc ccgtggtcag 23220

cgacgagcag ctggcccggt ggctgggtcc taatgctagt ccccagagtt tggaagagcg 23280

gcgcaaactc atgatggccg tggtcctggt gaccgtggag ctggagtgcc tgcgccgctt 23340

cttcgccgac gcggagaccc tgcgcaaggt cgaggagaac ctgcactacc tcttcaggca 23400

cgggttcgtg cgccaggcct gcaagatctc caacgtggag ctgaccaacc tggtctccta 23460

catgggcatc ttgcacgaga accgcctggg gcagaacgtg ctgcacacca ccctgcgcgg 23520

ggaggcccgg cgcgactaca tccgcgactg cgtctacctc tacctctgcc acacctggca 23580

gacgggcatg ggcgtgtggc agcagtgtct ggaggagcag aacctgaaag agctctgcaa 23640

gctcctgcag aagaacctca agggtctgtg gaccgggttc gacgagcgca ccaccgcctc 23700

ggacctggcc gacctcattt tccccgagcg cctcaggctg acgctgcgca acggcctgcc 23760

cgactttatg agccaaagca tgttgcaaaa ctttcgctct ttcatcctcg aacgctccgg 23820

aatcctgccc gccacctgct ccgcgctgcc ctcggacttc gtgccgctga ccttccgcga 23880

gtgccccccg ccgctgtgga gccactgcta cctgctgcgc ctggccaact acctggccta 23940

ccactcggac gtgatcgagg acgtcagcgg cgagggcctg ctcgagtgcc actgccgctg 24000

caacctctgc acgccgcacc gctccctggc ctgcaacccc cagctgctga gcgagaccca 24060

gatcatcggc accttcgagt tgcaagggcc cagcgaaggc gagggttcag ccgccaaggg 24120

gggtctgaaa ctcaccccgg ggctgtggac ctcggcctac ttgcgcaagt tcgtgcccga 24180

ggactaccat cccttcgaga tcaggttcta cgaggaccaa tcccatccgc ccaaggccga 24240

gctgtcggcc tgcgtcatca cccaggggggc gatcctggcc caattgcaag ccatccagaa 24300

atcccgccaa gaattcttgc tgaaaaaggg ccgcggggtc tacctcgacc cccagaccgg 24360

tgaggagctc aaccccggct tcccccagga tgccccgagg aaacaagaag ctgaaagtgg 24420

agctgccgcc cgtggaggat ttggaggaag actgggagaa cagcagtcag gcagaggagg 24480

aggagatgga ggaagactgg gacagcactc aggcagagga ggacagcctg caagacagtc 24540

tggaggaaga cgaggaggag gcagaggagg aggtggaaga agcagccgcc gccagaccgt 24600

cgtcctcggc gggggagaaa gcaagcagca cggataccat ctccgctccg ggtcggggtc 24660

ccgctcgacc acacagtaga tgggacgaga ccggacgatt cccgaacccc accacccaga 24720

ccggtaagaa ggagcggcag ggatacaagt cctggcgggg gcacaaaaac gccatcgtct 24780

cctgcttgca ggcctgcggg ggcaacatct ccttcacccg gcgctacctg ctcttccacc 24840

gcggggtgaa ctttccccgc aacatcttgc attactaccg tcacctccac agcccctact 24900

acttccaaga agaggcagca gcagcagaaa aagaccagca gaaaaccagc agctagaaaa 24960

tccacagcgg cggcagcagg tggactgagg atcgcggcga acgagccggc gcaaacccgg 25020

gagctgagga accggatctt tcccaccctc tatgccatct tccagcagag tcgggggcag 25080

gagcaggaac tgaaagtcaa gaaccgttct ctgcgctcgc tcacccgcag ttgtctgtat 25140

cacaagagcg aagaccaact tcagcgcact ctcgaggacg ccgaggctct cttcaacaag 25200

tactgcgcgc tcactcttaa agagtagccc gcgcccgccc agtcgcagaa aaaggcggga 25260

attacgtcac ctgtgccctt cgccctagcc gcctccaccc atcatcatga gcaaagagat 25320

tcccacgcct tacatgtgga gctaccagcc ccagatgggc ctggccgccg gtgccgccca 25380

ggactactcc acccgcatga attggctcag cgccgggccc gcgatgatct cacgggtgaa 25440

tgacatccgc gcccaccgaa accagatact cctagaacag tcagcgctca ccgccacgcc 25500

ccgcaatcac ctcaatccgc gtaattggcc cgccgccctg gtgtaccagg aaattcccca 25560

gcccacgacc gtactacttc cgcgagacgc ccaggccgaa gtccagctga ctaactcagg 25620

tgtccagctg gcgggcggcg ccaccctgtg tcgtcaccgc cccgctcagg gtataaagcg 25680

gctggtgatc cggggcagag gcacacagct caacgacgag gtggtgagct cttcgctggg 25740

tctgcgacct gacggagtct tccaactcgc cggatcgggg agatcttcct tcacgcctcg 25800

tcaggccgtc ctgactttgg agagttcgtc ctcgcagccc cgctcgggtg gcatcggcac 25860

tctccagttc gtggaggagt tcactccctc ggtctacttc aacccccttct ccggctcccc 25920

cggccactac ccggacgagt tcatcccgaa cttcgacgcc atcagcgagt cggtggacgg 25980

ctacgattga atgtcccatg gtggcgcagc tgacctagct cggcttcgac acctggacca 26040

ctgccgccgc ttccgctgct tcgctcggga tctcgccgag tttgcctact ttgagctgcc 26100

cgaggagcac cctcagggcc cggcccacgg agtgcggatc gtcgtcgaag ggggcctcga 26160

ctcccacctg cttcggatct tcagccagcg tccgatcctg gtcgagcgcg agcaaggaca 26220

gacccttctg actctgtact gcatctgcaa ccaccccggc ctgcatgaaa gtctttgttg 26280

tctgctgtgt actgagtata ataaaagctg agatcagcga ctactccgga cttccgtgtg 26340

ttcctgaatc catcaaccag tctttgttct tcaccgggaa cgagaccgag ctccagctcc 26400

agtgtaagcc ccacaagaag tacctcacct ggctgttcca gggctccccg atcgccgttg 26460

tcaaccactg cgacaacgac ggagtcctgc tgagcggccc tgccaacctt acttttttcca 26520

cccgcagaag caagctccag ctcttccaac ccttcctccc cgggacctat cagtgcgtct 26580

cgggaccctg ccatcacacc ttccacctga tcccgaatac cacagcgtcg ctccccgcta 26640

ctaacaacca aactaacctc caccaacgcc accgtcgcga cggccacaat acatgcccat 26700

attagactat gaggccgagc cacagcgacc catgctcccc gctattagtt acttcaatct 26760

aaccggcgga gatgactgac ccactggcca acaacaacgt caacgacctt ctcctggaca 26820

tggacggccg cgcctcggag cagcgactcg cccaacttcg cattcgccag cagcaggaga 26880

gagccgtcaa ggagctgcag gatgcggtgg ccatccacca gtgcaagaga ggcatcttct 26940

gcctggtgaa acaggccaag atctcctacg aggtcactcc aaacgaccat cgcctctcct 27000

acgagctcct gcagcagcgc cagaagttca cctgcctggt cggagtcaac cccatcgtca 27060
tcacccagca gtctggcgat accaaggggt gcatccactg ctcctgcgac tcccccgact 27120
gcgtccacac tctgatcaag accctctgcg gcctccgcga cctcctcccc atgaactaat 27180
caccccctta tccagtgaaa taaagatcat attgatgatg attttacaga aataaaaaat 27240
aatcatttga tttgaaataa agatacaatc atattgatga tttgagttta acaaaaaaat 27300
aaagaatcac ttacttgaaa tctgatacca ggtctctgtc catgttttct gccaacacca 27360
cttcactccc ctcttcccag ctctggtact gcaggccccg gcgggctgca aacttcctcc 27420

acacgctgaa ggggatgtca aattcctcct gtccctcaat cttcatttta tcttctatca 27480
gatgtccaaa aagcgcgtcc gggtggatga tgacttcgac cccgtctacc cctacgatgc 27540
agacaacgca ccgaccgtgc ccttcatcaa ccccccccttc gtctcttcag atggattcca 27600
agagaagccc ctgggggtgt tgtccctgcg actggccgac cccgtcacca ccaagaacgg 27660
ggaaatcacc ctcaagctgg gagagggggt ggacctcgat tcctcgggaa aactcatctc 27720
caacacggcc accaaggccg ccgcccctct cagtttttcc aacaacacca tttcccttaa 27780
catggatcac ccctttttaca ctaaagatgg aaaattatcc ttacaagttt ctccaccatt 27840

aaatatactg agaacaagca ttctaaacac actagcttta ggttttggat caggtttagg 27900
actccgtggc tctgccttgg cagtacagtt agtctctcca cttacatttg atactgatgg 27960
aaacataaag cttaccttag acagaggttt gcatgttaca acaggagatg caattgaaag 28020
caacataagc tgggctaaag gtttaaaatt tgaagatgga gccatagcaa ccaacattgg 28080
aaatgggtta gagtttggaa gcagtagtac agaaacaggt gttgatgatg cttacccaat 28140
ccaagttaaa cttggatctg gccttagctt tgacagtaca ggagccataa tggctggtaa 28200
caaagaagac gataaactca ctttgtggac aacacctgat ccatcaccaa actgtcaaat 28260

actcgcagaa aatgatgcaa aactaacact ttgcttgact aaatgtggta gtcaaatact 28320
ggccactgtg tcagtcttag ttgtaggaag tggaaaccta aaccccatta ctggcaccgt 28380
aagcagtgct caggtgtttc tacgttttga tgcaaacggt gttcttttaa cagaacattc 28440
tacactaaaa aaatactggg ggtataggca gggagatagc atagatggca ctccatatac 28500
caatgctgta ggattcatgc ccaatttaaa agcttatcca aagtcacaaa gttctactac 28560
taaaaataat atagtagggc aagtatacat gaatggagat gtttcaaaac ctatgcttct 28620
cactataacc ctcaatggta ctgatgacag caacagtaca tattcaatgt cattttcata 28680

cacctggact aatggaagct atgttggagc aacatttggg gctaactctt ataccttctc 28740

atacatcgcc caagaatgaa cactgtatcc caccctgcat gccaaccctt cccaccccac 28800

tctgtggaac aaactctgaa acacaaaata aaataaagtt caagtgtttt attgattcaa 28860

cagttttaca ggattcgagc agttatttttt cctccaccct cccaggacat ggaatacacc 28920

accctctccc cccgcacagc cttgaacatc tgaatgccat tggtgatgga catgcttttg 28980

gtctccacgt tccacacagt ttcagagcga gccagtctcg ggtcggtcag ggagatgaaa 29040

ccctccgggc actcccgcat ctgcacctca cagctcaaca gctgaggatt gtcctcggtg 29100

gtcgggatca cggttatctg gaagaagcag aagagcggcg gtgggaatca tagtccgcga 29160

acgggatcgg ccggtggtgt cgcatcaggc cccgcagcag tcgctgccgc cgccgctccg 29220

tcaagctgct gctcaggggg tccgggtcca gggactccct cagcatgatg cccacggccc 29280

tcagcatcag tcgtctggtg cggcgggcgc agcagcgcat gcggatctcg ctcaggtcgc 29340

tgcagtacgt gcaacacaga accaccaggt tgttcaacag tccatagttc aacacgctcc 29400

agccgaaact catcgcggga aggatgctac ccacgtggcc gtcgtaccag atcctcaggt 29460

aaatcaagtg gtgccccctc cagaacacgc tgcccacgta catgatctcc ttgggcatgt 29520

ggcggttcac cacctcccgg taccacatca ccctctggtt gaacatgcag ccccggatga 29580

tcctgcggaa ccacagggcc agcaccgccc cgcccgccat gcagcgaaga gaccccgggt 29640

cccggcaatg gcaatggagg acccaccgct cgtacccgtg gatcatctgg gagctgaaca 29700

agtctatgtt ggcacagcac aggcatatgc tcatgcatct cttcagcact ctcaactcct 29760

cgggggtcaa aaccatatcc cagggcacgg ggaactcttg caggacagcg aaccccgcag 29820

aacagggcaa tcctcgcaca gaacttacat tgtgcatgga cagggtatcg caatcaggca 29880

gcaccgggtg atcctccacc agagaagcgc gggtctcggt ctcctcacag cgtggtaagg 29940

gggccggccg atacgggtga tggcgggacg cggctgatcg tgttcgcgac cgtgtcatga 30000

tgcagttgct ttcggacatt ttcgtacttg ctgtagcaga acctggtccg ggcgctgcac 30060

accgatcgcc ggcggcggtc tcggcgcttg gaacgctcgg tgttgaaatt gtaaaacagc 30120

cactctctca gaccgtgcag cagatctagg gcctcaggag tgatgaagat cccatcatgc 30180

ctgatggctc tgatcacatc gaccaccgtg gaatgggcca gacccagcca gatgatgcaa 30240

ttttgttggg tttcggtgac ggcgggggag ggaagaacag gaagaaccat gattaacttt 30300

taatccaaac ggtctcggag tacttcaaaa tgaagatcgc ggagatggca cctctcgccc 30360

ccgctgtgtt ggtggaaaat aacagccagg tcaaaggtga tacggttctc gagatgttcc 30420

acggtggctt ccagcaaagc ctccacgcgc acatccagaa acaagacaat agcgaaagcg 30480

ggagggttct ctaattcctc aatcatcatg ttacactcct gcaccatccc cagataattt 30540

tcatttttcc agccttgaat gattcgaact agttcgtgag gtaaatccaa gccagccatg 30600

ataaagagct cgcgcagagc gccctccacc ggcattctta agcacaccct cataattcca 30660

agatattctg ctcctggttc acctgcagca gattgacaag cggaatatca aaatctctgc 30720

cgcgatccct gagctcctcc ctcagcaata actgtaagta ctctttcata tcctctccga 30780

aatttttagc cataggacca ccaggaataa gattagggca agccacagta cagataaacc 30840

gaagtcctcc ccagtgagca ttgccaaatg caagactgct ataagcatgc tggctagacc 30900

cggtgatatc ttccagataa ctggacagaa aatcgcccag gcaattttta agaaaatcaa 30960

caaaagaaaa atcctccagg tggacgttta gagcctcggg aacaacgatg aagtaaatgc 31020

aagcggtgcg ttccagcatg gttagttagc tgatctgtag aaaaaacaaa aatgaacatt 31080

aaaccatgct agcctggcga acaggtgggt aaatcgttct ctccagcacc aggcaggcca 31140

cggggtctcc ggcgcgaccc tcgtaaaaat tgtcgctatg attgaaaacc atcacagaga 31200

gacgttcccg gtggccggcg tgaatgattc gacaagatga atacaccccc ggaacattgg 31260

cgtccgcgag tgaaaaaaag cgcccgagga agcaataagg cactacaatg ctcagtctca 31320

agtccagcaa agcgatgcca tgcggatgaa gcacaaaatt ctcaggtgcg tacaaaatgt 31380

aattactccc ctcctgcaca ggcagcaaag cccccgatcc ctccaggtac acatacaaag 31440

cctcagcgtc catagcttac cgagcagcag cacacaacag gcgcaagagt cagagaaagg 31500

ctgagctcta acctgtccac ccgctctctg ctcaatatat agcccagatc tacactgacg 31560

taaaggccaa agtctaaaaa tacccgccaa ataatcacac acgcccagca cacgcccaga 31620

aaccggtgac acactcaaaa aaatacgcgc acttcctcaa acgcccaaaa ctgccgtcat 31680

ttccgggttc ccacgctacg tcatcaaaac acgactttca aattccgtcg accgttaaaa 31740

acgtcacccg ccccgcccct aacggtcgcc cgtctctcag ccaatcagcg ccccgcatcc 31800

ccaaattcaa acacctcatt tgcatattaa cgcgcacaaa aagtttgagg tatattattg 31860

atgatgg 31867

<210> 12

<211> 32788

<212> DNA

<213> Artificial Sequence

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

<400> 12

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

ccgggatgtt ccaggcactg tccgaaggct gcacacccta tgatattaac cagatgctga 1320

atgtcctggg agaccaccag gtctctggcc tggagcagct ggagagcatc atcaacttcg 1380

agaagctgac cgagtggaca agctccaatg tgatgcctat cctgtcccca ctgaccaagg 1440

gcatcctggg cttcgtgttt accctgacag tgccttctga gcggggcctg tcttgcatca 1500

gcgaggcaga cgcaaccaca ccagagtccg ccaatctggg cgaggagatc ctgtctcagc 1560

tgtacctgtg gccccgggtg acatatcact ccccttctta cgcctatcac cagttcgagc 1620

ggagagccaa gtacaagaga cacttcccag gctttggcca gtctctgctg ttcggctacc 1680

ccgtgtacgt gttcggcgat tgcgtgcagg gcgactggga tgccatccgg tttagatact 1740

gcgcaccacc tggatatgca ctgctgaggt gtaacgacac caattattcc gccctgctgg 1800

cagtgggcgc cctggagggc cctcgcaatc aggattggct gggcgtgcca aggcagctgg 1860

tgacacgcat gcaggccatc cagaacgcag gcctgtgcac cctggtggca atgctggagg 1920

agacaatctt ctggctgcag gcctttctga tggccctgac cgacagcggc cccaagacaa 1980

acatcatcgt ggattcccag tacgtgatgg gcatctccaa gccttctttc caggagtttg 2040

tggactggga gaacgtgagc ccagagctga attccaccga tcagccattc tggcaggcag 2100

gaatcctggc aaggaacctg gtgcctatgg tggccacagt gcagggccag aatctgaagt 2160

accagggcca gagcctggtc atcagcgcct ccatcatcgt gtttaacctg ctggagctgg 2220

agggcgacta tcgggacgat ggcaacgtgt gggtgcacac cccactgagc cccagaacac 2280

tgaacgcctg ggtgaaggcc gtggaggaga agaagggcat cccagtgcac ctggagctgg 2340

cctccatgac caatatggag ctgatgtcta gcatcgtgca ccagcaggtg aggacatacg 2400

gacccgtgtt catgtgcctg ggaggcctgc tgaccatggt ggcaggagcc gtgtggctga 2460

cagtgcgggt gctggagctg ttcagagccg cccagctggc caacgatgtg gtgctgcaga 2520

tcatggagct gtgcggagca gcctttcgcc aggtgtgcca caccacagtg ccatggccca 2580

atgcctccct gacccccaag tggaacaatg agacaacaca gcctcagatc gccaactgta 2640

gcgtgtacga cttcttcgtg tggctgcact actatagcgt gagggatacc ctgtggcccc 2700

gcgtgacata ccacatgaat aagtacgcct atcacatgct ggagaggcgc gccaagtata 2760

agagaggccc tggcccaggc gcaaagtttg tggcagcatg gaccctgaag gccgccgccg 2820

gccccggccc cggccagtat atcaaggcta acagtaagtt cattggaatc acagagctgg 2880

gacccggacc tggataatga gtttaaactc ccatttaaat gtgagggtta atgcttcgag 2940

cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 3000

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 3060

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 3120

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaataact ataacggtcc 3180

taaggtagcg agtgagtagt gttctggggc gggggaggac ctgcatgagg gccagaataa 3240

ctgaaatctg tgcttttctg tgtgttgcag cagcatgagc ggaagcggct cctttgaggg 3300

aggggtattc agcccttatc tgacggggcg tctcccctcc tgggcgggag tgcgtcagaa 3360

tgtgatggga tccacggtgg acggccggcc cgtgcagccc gcgaactctt caaccctgac 3420

ctatgcaacc ctgagctctt cgtcgttgga cgcagctgcc gccgcagctg ctgcatctgc 3480

cgccagcgcc gtgcgcggaa tggccatggg cgccggctac tacggcactc tggtggccaa 3540

ctcgagttcc accaataatc ccgccagcct gaacgaggag aagctgttgc tgctgatggc 3600

ccagctcgag gccttgaccc agcgcctggg cgagctgacc cagcaggtgg ctcagctgca 3660

ggagcagacg cgggccgcgg ttgccacggt gaaatccaaa taaaaaatga atcaataaat 3720

aaacggagac ggttgttgat tttaacacag agtctgaatc tttatttgat ttttcgcgcg 3780

cggtaggccc tggaccaccg gtctcgatca ttgagcaccc ggtggatctt ttccaggacc 3840
cggtagaggt gggcttggat gttgaggtac atgggcatga gcccgtcccg ggggtggagg 3900
tagctccatt gcagggcctc gtgctcgggg gtggtgttgt aaatcaccca gtcatagcag 3960
gggcgcaggg catggtgttg cacaatatct ttgaggagga gactgatggc cacgggcagc 4020
cctttggtgt aggtgtttac aaatctgttg agctgggagg gatgcatgcg gggggagatg 4080
aggtgcatct tggcctggat cttgagattg gcgatgttac cgcccagatc ccgcctgggg 4140
ttcatgttgt gcaggaccac cagcacggtg tatccggtgc acttggggaa tttatcatgc 4200

aacttggaag ggaaggcgtg aaagaatttg gcgacgcctt tgtgcccgcc caggttttcc 4260
atgcactcat ccatgatgat ggcgatgggc ccgtgggcgg cggcctgggc aaagacgttt 4320
cgggggtcgg acacatcata gttgtggtcc tgggtgaggt catcataggc cattttaatg 4380
aatttggggc ggagggtgcc ggactggggg acaaaggtac cctcgatccc gggggcgtag 4440
ttcccctcac agatctgcat ctcccaggct ttgagctcgg agggggggat catgtccacc 4500
tgcggggcga taaagaacac ggtttccggg gcgggggaga tgagctgggc cgaaagcaag 4560
ttccggagca gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc 4620

tgcaggtggt agttgaggga gagacagctg ccgtcctccc ggaggagggg ggccacctcg 4680
ttcatcatct cgcgcacgtg catgttctcg cgcaccagtt ccgccaggag gcgctctccc 4740
cccagggata ggagctcctg gagcgaggcg aagtttttca gcggcttgag tccgtcggcc 4800
atgggcattt tggagagggt ttgttgcaag agttccaggc ggtcccagag ctcggtgatg 4860
tgctctacgg catctcgatc cagcagacct cctcgtttcg cgggttggga cggctgcggg 4920
agtagggcac cagacgatgg gcgtccagcg cagccagggt ccggtccttc cagggtcgca 4980
gcgtccgcgt cagggtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg 5040

cgagggtgcg cttcaggctc atccggctgg tcgaaaaccg ctcccgatcg gcgccctgcg 5100
cgtcggccag gtagcaattg accatgagtt cgtagttgag cgcctcggcc gcgtggcctt 5160
tggcgcggag cttacctttg gaagtctgcc cgcaggcggg acagaggagg gacttgaggg 5220
cgtagagctt gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagtggg 5280
cgcagacggt ctcgcactcc acgagccagg tgaggtcggg ctggtcgggg tcaaaaacca 5340
gtttcccgcc gttctttttg atgcgtttct taccttggt ctccatgagc tcgtgtcccc 5400
gctgggtgac aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga 5460

gcggtgtgcc gcggtcctcc tcgtagagga accccgccca ctccgagacg aaagcccggg 5520

tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt 5580

ccacctttc cagggtatgc aaacacatgt ccccctcgtc cacatccagg aaggtgattg 5640

gcttgtaagt gtaggccacg tgaccggggg tcccggccgg gggggtataa aagggtgcgg 5700

gtccctgctc gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta 5760

ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg 5820

aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct 5880

ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt 5940

tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct 6000

tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg 6060

tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt 6120

ccacactggt ggccacctcg ccgcgcaggg gctcattagt ccagcagagg cgtccgccct 6180

tgcgcgagca gaaggggggc agggggtcca gcatgacctc gtcgggggggg tcggcatcga 6240

tggtgaagat gccgggcagg aggtcggggt caaagtagct gatggaagtg gccagatcgt 6300

ccagggcagc ttgccattcg cgcacggcca gcgcgctctc gtagggactg aggggcgtgc 6360

cccagggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga 6420

ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc 6480

gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cgggcccagg ttggtgcgac 6540

tgggcttttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg 6600

tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggatgaagt 6660

gggcgtagga gtcttgcagc ttggcgacga gctcggcggt gactaggacg tccagagcgc 6720

agtagtcgag ggtctcctgg atgatgtcat acttgagctg tccctttgt ttccacagct 6780

cgcggttgag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aacccgtcct 6840

gatctgcacg gtaagagcct agcatgtaga actggttgac ggccttgtag gcgcagcagc 6900

ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg 6960

cgaaagtgtc cctgaccatg accttgagga actggtgctt gaagtcgata tcgtcgcagc 7020

ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga 7080

aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga 7140

aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga 7200

agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt 7260

ggggcagttt cttgagctcc tcgtaggtga gctcgtcggg gtcgctgagc ccgtgctgct 7320

cgagcgccca gtcggcgaga tgggggttgg cgcggaggaa ggaagtccag agatccacgg 7380

ccagggcggt ttgcagacgg tcccggtact gacggaactg ctgcccgacg gccatttttt 7440

cgggggtgac gcagtagaag gtgcgggggt ccccgtgcca gcgatcccat ttgagctgga 7500

gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtttc atgaccagca 7560

tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg 7620

tgaggaagag cctttcggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc 7680

accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac 7740

actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt 7800

gctgcacgag ctgtacctga gttcctttga cgaggaattt cagtgggaag tggagtcgtg 7860

gcgcctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga 7920

tggtggtcat gctgacgagc ccgcgcggga ggcaggtcca gacctcggcg cgagcgggtc 7980

ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag 8040

tcaggtcagt gggcagcggc ggcgcgcggt tgacttgcag gagtttttcc agggcgcgcg 8100

ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca 8160

gggtcccgtg cccctggggt gtgaccaccg tcccccgttt cttcttgggc ggctggggcg 8220

acgggggcgg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag 8280

gggcggctcg gggcccggag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag 8340

gttctggtac tgcgcccgga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg 8400

gatctgacgc ctctgggtga aggccacggg acccgtgagt ttgaacctga aagagagttc 8460

gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc 8520

cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc 8580

tccgcggccg gcgcgctcca cggtggccgc gaggtcgttg gagatgcggc ccatgagctg 8640

cgagaaggcg ttcatgcccg cctcgttcca gacgcggctg tagaccacga cgccctcggg 8700

atcgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc 8760

gtagttgcag aggcgctggt agaggtagtt gagcgtggtg gcgatgtgct cggtgacgaa 8820

gaaatacatg atccagcggc ggagcggcat ctcgctgacg tcgcccagcg cctccaaacg 8880

ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac 8940

ggtcaactcc tcctccagaa gacggatgag ctcggcgatg gtggcgcgca cctcgcgctc 9000

gaaggccccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac 9060

ttcctcctca ggcggcagtg gtggcggggg aggggggcctg cgtcgccggc ggcgcacggg 9120

cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac 9180

ggcgcgcccg tcctcgcggg gccgcagcgt gaagacgccg ccgcgcatct ccaggtggcc 9240

gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt 9300
agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac 9360
gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc 9420
atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct 9480
gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag 9540
acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg 9600
catgagccgc tccacgggca cctcctcctc gcccgcgcgg ccgtgcatgc gcgtgagccc 9660

gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc 9720
ttgctggatc tgggtgaggg tggtctggaa gtcatcaaag tcgacgaagc ggtggtaggc 9780
tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc 9840
cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc 9900
gttgcaggtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggta 9960
gagcggccat cgctcggtgg cgggggcgcc gggcgcgagg tcctcgagca tggtgcggtg 10020
gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg 10080

gaactcgcgg acgcggttcc agatgttgcg cagcggcagg aagtagttca tggtgggcac 10140
ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg 10200
tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc 10260
cggttcgaat ctcgaatcag gctggagccg cagctaacgt ggtattggca ctcccgtctc 10320
gacccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt 10380
ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta 10440
gtctggagaa gaatcgccag ggttgcgttg cggtgtgccc cggttcgagg ccggccggat 10500

tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact 10560
tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcatcccgt 10620
actgcggcag atgcgccccc accaccctcc accgcaacaa cagccccctc cacagccggc 10680
gcttctgccc ccgccccagc agcaacttcc agccacgacc gccgcggccg ccgtgagcgg 10740
ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgcct 10800
gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctcgcgaggc 10860
ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg 10920

cgcggcccgg ttccacgcgg ggcgggagct gcggcgcggc ctggaccgaa agagggtgct 10980

gagggacgag gatttcgagg cggacgagct gacggggatc agccccgcgc gcgcgcacgt 11040

ggccgcggcc aacctggtca cggcgtacga gcagaccgtg aaggaggaga gcaacttcca 11100

aaaatccttc aacaaccacg tgcgcaccct gatcgcgcgc gaggaggtga ccctgggcct 11160

gatgcacctg tgggacctgc tggaggccat cgtgcagaac cccaccagca agccgctgac 11220

ggcgcagctg ttcctggtgg tgcagcatag tcgggacaac gaagcgttca gggaggcgct 11280

gctgaatatc accgagcccg agggccgctg gctcctggac ctggtgaaca ttctgcagag 11340

catcgtggtg caggagcgcg ggctgccgct gtccgagaag ctggcggcca tcaacttctc 11400

ggtgctgagt ttgggcaagt actacgctag gaagatctac aagaccccgt acgtgcccat 11460

agacaaggag gtgaagatcg acgggtttta catgcgcatg accctgaaag tgctgaccct 11520

gagcgacgat ctgggggtgt accgcaacga caggatgcac cgtgcggtga gcgccagcag 11580

gcggcgcgag ctgagcgacc aggagctgat gcatagtctg cagcgggccc tgaccggggc 11640

cgggaccgag ggggagagct actttgacat gggcgcggac ctgcactggc agcccagccg 11700

ccgggccttg gaggcggcgg caggacccta cgtagaagag gtggacgatg aggtggacga 11760

ggagggcgag tacctggaag actgatggcg cgaccgtatt tttgctagat gcaacaacaa 11820

cagccacctc ctgatcccgc gatgcgggcg gcgctgcaga gccagccgtc cggcattaac 11880

tcctcggacg attggaccca ggccatgcaa cgcatcatgg cgctgacgac ccgcaacccc 11940

gaagccttta gacagcagcc ccaggccaac cggctctcgg ccatcctgga ggccgtggtg 12000

ccctcgcgct ccaaccccac gcacgagaag gtcctggcca tcgtgaacgc gctggtggag 12060

aacaaggcca tccgcggcga cgaggccggc ctggtgtaca acgcgctgct ggagcgcgtg 12120

gcccgctaca acagcaccaa cgtgcagacc aacctggacc gcatggtgac cgacgtgcgc 12180

gaggccgtgg cccagcgcga gcggttccac cgcgagtcca acctgggatc catggtggcg 12240

ctgaacgcct tcctcagcac ccagcccgcc aacgtgcccc ggggccagga ggactacacc 12300

aacttcatca gcgccctgcg cctgatggtg accgaggtgc cccagagcga ggtgtaccag 12360

tccgggccgg actacttctt ccagaccagt cgccagggct tgcagaccgt gaacctgagc 12420

caggctttca agaacttgca gggcctgtgg ggcgtgcagg ccccggtcgg ggaccgcgcg 12480

acggtgtcga gcctgctgac gccgaactcg cgcctgctgc tgctgctggt ggcccccttc 12540

acggacagcg gcagcatcaa ccgcaactcg tacctgggct acctgattaa cctgtaccgc 12600

gaggccatcg gccaggcgca cgtggacgag cagacctacc aggagatcac ccacgtgagc 12660

cgcgccctgg gccaggacga cccgggcaac ctggaagcca ccctgaactt tttgctgacc 12720

aaccggtcgc agaagatccc gccccagtac gcgctcagca ccgaggagga gcgcatcctg 12780

cgttacgtgc agcagagcgt gggcctgttc ctgatgcagg agggggccac ccccagcgcc 12840

gcgctcgaca tgaccgcgcg caacatggag cccagcatgt acgccagcaa ccgcccgttc 12900

atcaataaac tgatggacta cttgcatcgg gcggccgcca tgaactctga ctatttcacc 12960

aacgccatcc tgaatcccca ctggctcccg ccgccggggt tctacacggg cgagtacgac 13020

atgcccgacc ccaatgacgg gttcctgtgg gacgatgtgg acagcagcgt gttctccccc 13080

cgaccgggtg ctaacgagcg ccccttgtgg aagaaggaag gcagcgaccg acgcccgtcc 13140

tcggcgctgt ccggccgcga gggtgctgcc gcggcggtgc ccgaggccgc cagtcctttc 13200

ccgagcttgc ccttctcgct gaacagtatc cgcagcagcg agctgggcag gatcacgcgc 13260

ccgcgcttgc tgggcgaaga ggagtacttg aatgactcgc tgttgagacc cgagcgggag 13320

aagaacttcc ccaataacgg gatagaaagc ctggtggaca agatgagccg ctggaagacg 13380

tatgcgcagg agcacaggga cgatccccgg gcgtcgcagg gggccacgag ccggggcagc 13440

gccgcccgta aacgccggtg gcacgacagg cagcggggac agatgtggga cgatgaggac 13500

tccgccgacg acagcagcgt gttggacttg ggtgggagtg gtaacccgtt cgctcacctg 13560

cgcccccgta tcgggcgcat gatgtaagag aaaccgaaaa taaatgatac tcaccaaggc 13620

catggcgacc agcgtgcgtt cgtttcttct ctgttgttgt tgtatctagt atgatgaggc 13680

gtgcgtaccc ggagggtcct cctccctcgt acgagagcgt gatgcagcag gcgatggcgg 13740

cggcggcgat gcagcccccg ctggaggctc cttacgtgcc cccgcggtac ctggcgccta 13800

cggaggggcg gaacagcatt cgttactcgg agctggcacc cttgtacgat accacccggt 13860

tgtacctggt ggacaacaag tcggcggaca tcgcctcgct gaactaccag aacgaccaca 13920

gcaacttcct gaccaccgtg gtgcagaaca atgacttcac ccccacggag gccagcaccc 13980

agaccatcaa ctttgacgag cgctcgcggt ggggcggcca gctgaaaacc atcatgcaca 14040

ccaacatgcc caacgtgaac gagttcatgt acagcaacaa gttcaaggcg cgggtgatgg 14100

tctcccgcaa gacccccaat ggggtgacag tgacagagga ttatgatggt agtcaggatg 14160

agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga 14220

ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg 14280

gggtgctgga gagcgacatc ggcgtgaagt tcgacactag gaacttcagg ctgggctggg 14340

accccgtgac cgagctggtc atgcccgggg tgtacaccaa cgaggctttc catcccgata 14400

ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg 14460

gcattcgcaa gaggcagccc ttccaggaag gcttccagat catgtacgag gatctggagg 14520

ggggcaacat ccccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag 14580

cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg 14640

caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcattc 14700

agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca 14760
ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtgcgct 14820
cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc 14880
tgcccgacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc 14940
cggtggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg 15000
tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg 15060
agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc 15120

ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg 15180
tgaccgttac tgacgccaga cgccgcacct gcccctacgt ctacaaggcc ctgggcatag 15240
tcgcgccgcg cgtcctctcg agccgcacct tctaaatgtc cattctcatc tcgcccagta 15300
ataacaccgg ttggggcctg cgcgcgccca gcaagatgta cggaggcgct cgccaacgct 15360
ccacgcaaca ccccgtgcgc gtgcgcgggc acttccgcgc tccctggggc gccctcaagg 15420
gccgcgtgcg gtcgcgcacc accgtcgacg acgtgatcga ccaggtggtg gccgacgcgc 15480
gcaactacac ccccgccgcc gcgcccgtct ccaccgtgga cgccgtcatc gacagcgtgg 15540

tggccgacgc gcgccggtac gcccgcgcca agagccggcg gcggcgcatc gcccggcggc 15600
accggagcac ccccgccatg cgcgcggcgc gagccttgct gcgcagggcc aggcgcacgg 15660
gacgcagggc catgctcagg gcggccagac gcgcggcttc aggcgccagc gccggcagga 15720
cccggagacg cgcggccacg gcggcggcag cggccatcgc cagcatgtcc cgcccgcggc 15780
gagggaacgt gtactgggtg cgcgacgccg ccaccggtgt gcgcgtgccc gtgcgcaccc 15840
gcccccctcg cacttgaaga tgttcacttc gcgatgttga tgtgtcccag cggcgaggag 15900
gatgtccaag cgcaaattca aggaagagat gctccaggtc atcgcgcctg agatctacgg 15960

ccctgcggtg gtgaaggagg aaagaaagcc ccgcaaaatc aagcgggtca aaaaggacaa 16020
aaaggaagaa gaaagtgatg tggacggatt ggtggagttt gtgcgcgagt tcgccccccg 16080
gcggcgcgtg cagtggcgcg ggcggaaggt gcaaccggtg ctgagacccg gcaccaccgt 16140
ggtcttcacg cccggcgagc gctccggcac cgcttccaag cgctcctacg acgaggtgta 16200
cggggatgat gatattctgg agcaggcggc cgagcgcctg ggcgagtttg cttacggcaa 16260
gcgcagccgt tccgcaccga aggaagaggc ggtgtccatc ccgctggacc acggcaaccc 16320
cacgccgagc ctcaagcccg tgaccttgca gcaggtgctg ccgaccgcgg cgccgcgccg 16380

ggggttcaag cgcgagggcg aggatctgta ccccaccatg cagctgatgg tgcccaagcg 16440

ccagaagctg gaagacgtgc tggagaccat gaaggtggac ccggacgtgc agcccgaggt 16500

caaggtgcgg cccatcaagc aggtggcccc gggcctgggc gtgcagaccg tggacatcaa 16560

gattcccacg gagcccatgg aaacgcagac cgagcccatg atcaagccca gcaccagcac 16620

catggaggtg cagacggatc cctggatgcc atcggctcct agtcgaagac cccggcgcaa 16680

gtacggcgcg gccagcctgc tgatgcccaa ctacgcgctg catccttcca tcatccccac 16740

gccgggctac cgcggcacgc gcttctaccg cggtcatacc agcagccgcc gccgcaagac 16800

caccactcgc cgccgccgtc gccgcaccgc cgctgcaacc acccctgccg ccctggtgcg 16860

gagagtgtac cgccgcggcc gcgcacctct gaccctgccg cgcgcgcgct accacccgag 16920

catcgccatt taaactttcg cctgctttgc agatcaatgg ccctcacatg ccgccttcgc 16980

gttcccatta cgggctaccg aggaagaaaa ccgcgccgta gaaggctggc ggggaacggg 17040

atgcgtcgcc accaccaccg gcggcggcgc gccatcagca agcggttggg gggaggcttc 17100

ctgcccgcgc tgatccccat catcgccgcg gcgatcgggg cgatccccgg cattgcttcc 17160

gtggcggtgc aggcctctca gcgccactga gacacacttg gaaacatctt gtaataaacc 17220

aatggactct gacgctcctg gtcctgtgat gtgttttcgt agacagatgg aagacatcaa 17280

tttttcgtcc ctggctccgc gacacggcac gcggccgttc atgggcacct ggagcgacat 17340

cggcaccagc caactgaacg ggggcgcctt caattggagc agtctctgga gcgggcttaa 17400

gaatttcggg tccacgctta aaacctatgg cagcaaggcg tggaacagca ccacagggca 17460

ggcgctgagg gataagctga aagagcagaa cttccagcag aaggtggtcg atgggctcgc 17520

ctcgggcatc aacggggtgg tggacctggc caaccaggcc gtgcagcggc agatcaacag 17580

ccgcctggac ccggtgccgc ccgccggctc cgtggagatg ccgcaggtgg aggaggagct 17640

gcctcccctg gacaagcggg gcgagaagcg accccgcccc gatgcggagg agacgctgct 17700

gacgcacacg gacgagccgc ccccgtacga ggaggcggtg aaactgggtc tgcccaccac 17760

gcggcccatc gcgcccctgg ccaccggggt gctgaaaccc gaaaagcccg cgaccctgga 17820

cttgcctcct ccccagcctt cccgcccctc tacagtggct aagcccctgc cgccggtggc 17880

cgtggcccgc gcgcgacccg ggggcaccgc ccgccctcat gcgaactggc agagcactct 17940

gaacagcatc gtgggtctgg gagtgcagag tgtgaagcgc cgccgctgct attaaaccta 18000

ccgtagcgct taacttgctt gtctgtgtgt gtatgtatta tgtcgccgcc gccgctgtcc 18060

accagaagga ggagtgaaga ggcgcgtcgc cgagttgcaa gatggccacc ccatcgatgc 18120

tgccccagtg ggcgtacatg cacatcgccg gacaggacgc ttcggagtac ctgagtccgg 18180

gtctggtgca gtttgcccgc gccacagaca cctacttcag tctggggaac aagtttagga 18240

accccacggt ggcgcccacg cacgatgtga ccaccgaccg cagccagcgg ctgacgctgc 18300

gcttcgtgcc cgtggaccgc gaggacaaca cctactcgta caaagtgcgc tacacgctgg 18360

ccgtgggcga caaccgcgtg ctggacatgg ccagcaccta ctttgacatc cgcggcgtgc 18420

tggatcgggg ccctagcttc aaaccctact ccggcaccgc ctacaacagt ctggccccca 18480

agggagcacc caacacttgt cagtggacat ataaagccga tggtgaaact gccacagaaa 18540

aaacctatac atatggaaat gcacccgtgc agggcattaa catcacaaaa gatggtattc 18600

aacttggaac tgacaccgat gatcagccaa tctacgcaga taaaacctat cagcctgaac 18660

ctcaagtggg tgatgctgaa tggcatgaca tcactggtac tgatgaaaag tatggaggca 18720

gagctcttaa gcctgatacc aaaatgaagc cttgttatgg ttcttttgcc aagcctacta 18780

ataaagaagg aggtcaggca aatgtgaaaa caggaacagg cactactaaa gaatatgaca 18840

tagacatggc tttctttgac aacagaagtg cggctgctgc tggcctagct ccagaaattg 18900

ttttgtatac tgaaaatgtg gatttggaaa ctccagatac ccatattgta tacaaagcag 18960

gcacagatga cagcagctct tctattaatt tgggtcagca agccatgccc aacagaccta 19020

actacattgg tttcagagac aactttatcg ggctcatgta ctacaacagc actggcaata 19080

tgggggtgct ggccggtcag gcttctcagc tgaatgctgt ggttgacttg caagacagaa 19140

acaccgagct gtcctaccag ctcttgcttg actctctggg tgacagaacc cggtatttca 19200

gtatgtggaa tcaggcggtg gacagctatg atcctgatgt gcgcattatt gaaaatcatg 19260

gtgtggagga tgaacttccc aactattgtt tccctctgga tgctgttggc agaacagata 19320

cttatcaggg aattaaggct aatggaactg atcaaaccac atggaccaaa gatgacagtg 19380

tcaatgatgc taatgagata ggcaagggta atccattcgc catggaaatc aacatccaag 19440

ccaacctgtg gaggaacttc ctctacgcca acgtggccct gtacctgccc gactcttaca 19500

agtacacgcc ggccaatgtt accctgccca ccaacaccaa cacctacgat tacatgaacg 19560

gccgggtggt ggcgccctcg ctggtggact cctacatcaa catcggggcg cgctggtcgc 19620

tggatcccat ggacaacgtg aaccccttca accaccaccg caatgcgggg ctgcgctacc 19680

gctccatgct cctgggcaac gggcgctacg tgcccttcca catccaggtg ccccagaaat 19740

ttttcgccat caagagcctc ctgctcctgc ccgggtccta cacctacgag tggaacttcc 19800

gcaaggacgt caacatgatc ctgcagagct ccctcggcaa cgacctgcgc acggacgggg 19860

cctccatctc cttcaccagc atcaacctct acgccacctt cttccccatg gcgcacaaca 19920

cggcctccac gctcgaggcc atgctgcgca acgacaccaa cgaccagtcc ttcaacgact 19980

acctctcggc ggccaacatg ctctacccca tcccggccaa cgccaccaac gtgcccatct 20040

ccatcccctc gcgcaactgg gccgccttcc gcggctggtc cttcacgcgt ctcaagacca 20100

aggagacgcc ctcgctgggc tccgggttcg acccctactt cgtctactcg ggctccatcc 20160

cctacctcga cggcaccttc tacctcaacc acaccttcaa gaaggtctcc atcaccttcg 20220
actcctccgt cagctggccc ggcaacgacc ggctcctgac gcccaacgag ttcgaaatca 20280
agcgcaccgt cgacggcgag ggctacaacg tggcccagtg caacatgacc aaggactggt 20340
tcctggtcca gatgctggcc cactacaaca tcggctacca gggcttctac gtgcccgagg 20400
gctacaagga ccgcatgtac tccttcttcc gcaacttcca gcccatgagc cgccaggtgg 20460
tggacgaggt caactacaag gactaccagg ccgtcaccct ggcctaccag cacaacaact 20520
cgggcttcgt cggctacctc gcgcccacca tgcgccaggg ccagccctac cccgccaact 20580

acccctaccc gctcatcggc aagagcgccg tcaccagcgt cacccagaaa aagttcctct 20640
gcgacagggt catgtggcgc atccccttct ccagcaactt catgtccatg ggcgcgctca 20700
ccgacctcgg ccagaacatg ctctatgcca actccgccca cgcgctagac atgaatttcg 20760
aagtcgaccc catggatgag tccacccttc tctatgttgt cttcgaagtc ttcgacgtcg 20820
tccgagtgca ccagccccac cgcggcgtca tcgaggccgt ctacctgcgc acccccttct 20880
cggccggtaa cgccaccacc taagctcttg cttcttgcaa gccatggccg cgggctccgg 20940
cgagcaggag ctcagggcca tcatccgcga cctgggctgc gggccctact tcctgggcac 21000

cttcgataag cgcttcccgg gattcatggc cccgcacaag ctggcctgcg ccatcgtcaa 21060
cacggccggc cgcgagaccg ggggcgagca ctggctggcc ttcgcctgga acccgcgctc 21120
gaacacctgc tacctcttcg acccccttcgg gttctcggac gagcgcctca agcagatcta 21180
ccagttcgag tacgagggcc tgctgcgccg cagcgccctg gccaccgagg accgctgcgt 21240
caccctggaa aagtccaccc agaccgtgca gggtccgcgc tcggccgcct gcgggctctt 21300
ctgctgcatg ttcctgcacg ccttcgtgca ctggcccgac cgccccatgg acaagaaccc 21360
caccatgaac ttgctgacgg gggtgcccaa cggcatgctc cagtcgcccc aggtggaacc 21420

caccctgcgc cgcaaccagg aggcgctcta ccgcttcctc aactcccact ccgcctactt 21480
tcgctcccac cgcgcgcgca tcgagaaggc caccgccttc gaccgcatga atcaagacat 21540
gtaaaccgtg tgtgtatgtt aaatgtcttt aataaacagc actttcatgt tacacatgca 21600
tctgagatga tttatttaga aatcgaaagg gttctgccgg gtctcggcat ggcccgcggg 21660
cagggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt 21720
gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cgcgtcagtt gcagggcgcc 21780
cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga 21840

gttgcggtac acggggttgc agcactggaa caccatcagg gccgggtgct tcacgctcgc 21900

cagcaccgtc gcgtcggtga tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa 21960

gggggtcatc ttgcaggtct gccttcccat ggtgggcacg cacccgggct tgtggttgca 22020

atcgcagtgc agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat 22080

ggccttcatg aaagcctcca attgcctgaa cgcctgctgg gccttggctc cctcggtgaa 22140

gaagaccccg caggacttgc tagagaactg gttggtggcg cacccggcgt cgtgcacgca 22200

gcagcgcgcg tcgttgttgg ccagctgcac cacgctgcgc ccccagcggt tctgggtgat 22260

cttggcccgg tcggggttct ccttcagcgc gcgctgcccg ttctcgctcg ccacatccat 22320

ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc aggcaccgca gcttgccctc 22380

ggcctcggtg cacccgtgca gccacagcgc gcacccggtg cactcccagt tcttgtgggc 22440

gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt 22500

cttgttgcta gtgaaggtca gcggaatgcc gcggtgctcc tcgttgatgt acaggtggca 22560

gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcggt 22620

ctccacgcgg tagcggtcca tcagcatagt catgatttcc ataccttct cccaggccga 22680

gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag 22740

ggggtcgctc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac 22800

cgggggggtag ctgaagccca cggccgccag ctcctcctcg gcctgtcttt cgtcctcgct 22860

gtcctggctg acgtcctgca ggaccacatg cttggtcttg cggggtttct tcttgggcgg 22920

cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac 22980

tatctcttcc tcttcttggt ccgaggccac gcggcggtag gtatgtctct tcgggggcag 23040

aggcggaggc gacgggctct cgccgccgcg acttggcgga tggctggcag agccccttcc 23100

gcgttcgggg gtgcgctccc ggcggcgctc tgactgactt cctccgcggc cggccattgt 23160

gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc 23220

ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgcccag 23280

ccccgccacc tccgacgcgg ccgtcccaga catgcaagag atggaggaat ccatcgagat 23340

tgacctgggc tatgtgacgc ccgcggagca cgaggaggag ctggcagtgc gcttttcaca 23400

agaagagata caccaagaac agccagagca ggaagcagag aatgagcaga gtcaggctgg 23460

gctcgagcat gacggcgact acctccacct gagcgggggg gaggacgcgc tcatcaagca 23520

tctggcccgg caggccacca tcgtcaagga tgcgctgctc gaccgcaccg aggtgccgct 23580

cagcgtggag gagctcagcc gcgcctacga gttgaacctc ttctcgccgc gcgtgccccc 23640

caagcgccag cccaatggca cctgcgagcc caacccgcgc ctcaacttct acccggtctt 23700

cgcggtgccc gaggccctgg ccacctacca catctttttc aagaaccaaa agatccccgt 23760

ctcctgccgc gccaaccgca cccgcgccga cgccctttttc aacctgggtc ccggcgcccg 23820

cctacctgat atcgcctcct tggaagaggt tcccaagatc ttcgagggtc tgggcagcga 23880

cgagactcgg gccgcgaacg ctctgcaagg agaaggagga gagcatgagc accacagcgc 23940

cctggtcgag ttggaaggcg acaacgcgcg gctggcggtg ctcaaacgca cggtcgagct 24000

gacccatttc gcctacccgg ctctgaacct gccccccaaa gtcatgagcg cggtcatgga 24060

ccaggtgctc atcaagcgcg cgtcgcccat ctccgaggac gagggcatgc aagactccga 24120

ggagggcaag cccgtggtca gcgacgagca gctggcccgg tggctgggtc ctaatgctag 24180

tccccagagt ttggaagagc ggcgcaaact catgatggcc gtggtcctgg tgaccgtgga 24240

gctggagtgc ctgcgccgct tcttcgccga cgcggagacc ctgcgcaagg tcgaggagaa 24300

cctgcactac ctcttcaggc acgggttcgt gcgccaggcc tgcaagatct ccaacgtgga 24360

gctgaccaac ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt 24420

gctgcacacc accctgcgcg gggaggcccg gcgcgactac atccgcgact gcgtctacct 24480

ctacctctgc cacacctggc agacgggcat gggcgtgtgg cagcagtgtc tggaggagca 24540

gaacctgaaa gagctctgca agctcctgca gaagaacctc aagggtctgt ggaccgggtt 24600

cgacgagcgc accaccgcct cggacctggc cgacctcatt ttccccgagc gcctcaggct 24660

gacgctgcgc aacggcctgc ccgactttat gagccaaagc atgttgcaaa actttcgctc 24720

tttcatcctc gaacgctccg gaatcctgcc cgccacctgc tccgcgctgc cctcggactt 24780

cgtgccgctg accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg 24840

cctggccaac tacctggcct accactcgga cgtgatcgag gacgtcagcg gcgagggcct 24900

gctcgagtgc cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc 24960

ccagctgctg agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccagcgaagg 25020

cgagggttca gccgccaagg ggggtctgaa actcaccccg gggctgtgga cctcggccta 25080

cttgcgcaag ttcgtgcccg aggactacca tcccttcgag atcaggttct acgaggacca 25140

atcccatccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg cgatcctggc 25200

ccaattgcaa gccatccaga aatcccgcca agaattcttg ctgaaaaagg gccgcggggt 25260

ctacctcgac ccccagaccg gtgaggagct caaccccggc ttcccccagg atgccccgag 25320

gaaacaagaa gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga 25380

acagcagtca ggcagaggag gaggagatgg aggaagactg ggacagcact caggcagagg 25440

aggacagcct gcaagacagt ctggaggaag acgaggagga ggcagaggag gaggtggaag 25500

aagcagccgc cgccagaccg tcgtcctcgg cgggggagaa agcaagcagc acggatacca 25560

tctccgctcc gggtcggggt cccgctcgac cacacagtag atgggacgag accggacgat 25620

tcccgaaccc caccacccag accggtaaga aggagcggca gggatacaag tcctggcggg 25680
ggcacaaaaa cgccatcgtc tcctgcttgc aggcctgcgg gggcaacatc tccttcaccc 25740
ggcgctacct gctcttccac cgcggggtga actttccccg caacatcttg cattactacc 25800
gtcacctcca cagcccctac tacttccaag aagaggcagc agcagcagaa aaagaccagc 25860
agaaaaccag cagctagaaa atccacagcg gcggcagcag gtggactgag gatcgcggcg 25920
aacgagccgg cgcaaacccg ggagctgagg aaccggatct ttcccaccct ctatgccatc 25980
ttccagcaga gtcgggggca ggagcaggaa ctgaaagtca agaaccgttc tctgcgctcg 26040

ctcacccgca gttgtctgta tcacaagagc gaagaccaac ttcagcgcac tctcgaggac 26100
gccgaggctc tcttcaacaa gtactgcgcg ctcactctta aagagtagcc cgcgcccgcc 26160
cagtcgcaga aaaaggcggg aattacgtca cctgtgccct tcgccctagc cgcctccacc 26220
catcatcatg agcaaagaga ttcccacgcc ttacatgtgg agctaccagc cccagatggg 26280
cctggccgcc ggtgccgccc aggactactc cacccgcatg aattggctca gcgccgggcc 26340
cgcgatgatc tcacgggtga atgacatccg cgcccaccga aaccagatac tcctagaaca 26400
gtcagcgctc accgccacgc cccgcaatca cctcaatccg cgtaattggc ccgccgccct 26460

ggtgtaccag gaaattcccc agcccacgac cgtactactt ccgcgagacg cccaggccga 26520
agtccagctg actaactcag gtgtccagct ggcgggcggc gccaccctgt gtcgtcaccg 26580
ccccgctcag ggtataaagc ggctggtgat ccggggcaga ggcacacagc tcaacgacga 26640
ggtggtgagc tcttcgctgg gtctgcgacc tgacggagtc ttccaactcg ccggatcggg 26700
gagatcttcc ttcacgcctc gtcaggccgt cctgactttg gagagttcgt cctcgcagcc 26760
ccgctcgggt ggcatcggca ctctccagtt cgtggaggag ttcactccct cggtctactt 26820
caaccccttc tccggctccc ccggccacta cccggacgag ttcatcccga acttcgacgc 26880

catcagcgag tcggtggacg gctacgattg aatgtcccat ggtggcgcag ctgacctagc 26940
tcggcttcga cacctggacc actgccgccg cttccgctgc ttcgctcggg atctcgccga 27000
gtttgcctac tttgagctgc ccgaggagca ccctcagggc ccggcccacg gagtgcggat 27060
cgtcgtcgaa gggggcctcg actcccacct gcttcggatc ttcagccagc gtccgatcct 27120
ggtcgagcgc gagcaaggac agacccttct gactctgtac tgcatctgca accaccccgg 27180
cctgcatgaa agtctttgtt gtctgctgtg tactgagtat aataaaagct gagatcagcg 27240
actactccgg acttccgtgt gttcctgaat ccatcaacca gtctttgttc ttcaccggga 27300

acgagaccga gctccagctc cagtgtaagc cccacaagaa gtacctcacc tggctgttcc 27360

agggctcccc gatcgccgtt gtcaaccact gcgacaacga cggagtcctg ctgagcggcc 27420

ctgccaacct tacttttcc acccgcagaa gcaagctcca gctcttccaa cccttcctcc 27480

ccgggaccta tcagtgcgtc tcgggaccct gccatcacac cttccacctg atcccgaata 27540

ccacagcgtc gctccccgct actaacaacc aaactaacct ccaccaacgc caccgtcgcg 27600

acggccacaa tacatgccca tattagacta tgaggccgag ccacagcgac ccatgctccc 27660

cgctattagt tacttcaatc taaccggcgg agatgactga cccactggcc aacaacaacg 27720

tcaacgacct tctcctggac atggacggcc gcgcctcgga gcagcgactc gcccaacttc 27780

gcattcgcca gcagcaggag agagccgtca aggagctgca ggatgcggtg gccatccacc 27840

agtgcaagag aggcatcttc tgcctggtga aacaggccaa gatctcctac gaggtcactc 27900

caaacgacca tcgcctctcc tacgagctcc tgcagcagcg ccagaagttc acctgcctgg 27960

tcggagtcaa ccccatcgtc atcacccagc agtctggcga taccaagggg tgcatccact 28020

gctcctgcga ctcccccgac tgcgtccaca ctctgatcaa gaccctctgc ggcctccgcg 28080

acctcctccc catgaactaa tcacccccct atccagtgaa ataaagatca tattgatgat 28140

gattttacag aaataaaaaa taatcatttg atttgaaata aagatacaat catattgatg 28200

atttgagttt aacaaaaaaa taaagaatca cttacttgaa atctgatacc aggtctctgt 28260

ccatgttttc tgccaacacc acttcactcc cctcttccca gctctggtac tgcaggcccc 28320

ggcgggctgc aaacttcctc cacacgctga aggggatgtc aaattcctcc tgtccctcaa 28380

tcttcatttt atcttctatc agatgtccaa aaagcgcgtc cgggtggatg atgacttcga 28440

ccccgtctac ccctacgatg cagacaacgc accgaccgtg cccttcatca acccccccct 28500

cgtctcttca gatggattcc aagagaagcc cctgggggtg ttgtccctgc gactggccga 28560

ccccgtcacc accaagaacg gggaaatcac cctcaagctg ggagaggggg tggacctcga 28620

ttcctcggga aaactcatct ccaacacggc caccaaggcc gccgcccctc tcagtttttc 28680

caacaacacc atttccctta acatggatca ccccttttac actaaagatg gaaaattatc 28740

cttacaagtt tctccaccat taaatatact gagaacaagc attctaaaca cactagcttt 28800

aggttttgga tcaggtttag gactccgtgg ctctgccttg gcagtacagt tagtctctcc 28860

acttacattt gatactgatg gaaacataaa gcttacctta gacagaggtt tgcatgttac 28920

aacaggagat gcaattgaaa gcaacataag ctgggctaaa ggtttaaaat ttgaagatgg 28980

agccatagca accaacattg gaaatgggtt agagtttgga agcagtagta cagaaacagg 29040

tgttgatgat gcttacccaa tccaagttaa acttggatct ggccttagct ttgacagtac 29100

aggagccata atggctggta acaaagaaga cgataaactc actttgtgga caacacctga 29160

tccatcacca aactgtcaaa tactcgcaga aaatgatgca aaactaacac tttgcttgac 29220

taaatgtggt agtcaaatac tggccactgt gtcagtctta gttgtaggaa gtggaaacct 29280

aaaccccatt actggcaccg taagcagtgc tcaggtgttt ctacgttttg atgcaaacgg 29340

tgttctttta acagaacatt ctacactaaa aaaatactgg gggtataggc agggagatag 29400

catagatggc actccatata ccaatgctgt aggattcatg cccaatttaa aagcttatcc 29460

aaagtcacaa agttctacta ctaaaaataa tatagtaggg caagtataca tgaatggaga 29520

tgtttcaaaa cctatgcttc tcactataac cctcaatggt actgatgaca gcaacagtac 29580

atattcaatg tcattttcat acacctggac taatggaagc tatgttggag caacatttgg 29640

ggctaactct tataccttct catacatcgc ccaagaatga acactgtatc ccaccctgca 29700

tgccaaccct tcccacccca ctctgtggaa caaactctga aacacaaaat aaaataaagt 29760

tcaagtgttt tattgattca acagttttac aggattcgag cagttatttt tcctccaccc 29820

tcccaggaca tggaatacac caccctctcc ccccgcacag ccttgaacat ctgaatgcca 29880

ttggtgatgg acatgctttt ggtctccacg ttccacacag tttcagagcg agccagtctc 29940

gggtcggtca gggagatgaa accctccggg cactcccgca tctgcacctc acagctcaac 30000

agctgaggat tgtcctcggt ggtcgggatc acggttatct ggaagaagca gaagagcggc 30060

ggtgggaatc atagtccgcg aacgggatcg gccggtggtg tcgcatcagg ccccgcagca 30120

gtcgctgccg ccgccgctcc gtcaagctgc tgctcagggg gtccgggtcc agggactccc 30180

tcagcatgat gcccacggcc ctcagcatca gtcgtctggt gcggcgggcg cagcagcgca 30240

tgcggatctc gctcaggtcg ctgcagtacg tgcaacacag aaccaccagg ttgttcaaca 30300

gtccatagtt caacacgctc cagccgaaac tcatcgcggg aaggatgcta cccacgtggc 30360

cgtcgtacca gatcctcagg taaatcaagt ggtgcccccc ccagaacacg ctgcccacgt 30420

acatgatctc cttgggcatg tggcggttca ccacctcccg gtaccacatc accctctggt 30480

tgaacatgca gccccggatg atcctgcgga accacagggc cagcaccgcc ccgcccgcca 30540

tgcagcgaag agaccccggg tcccggcaat ggcaatggag gacccaccgc tcgtacccgt 30600

ggatcatctg ggagctgaac aagtctatgt tggcacagca caggcatatg ctcatgcatc 30660

tcttcagcac tctcaactcc tcgggggtca aaaccatatc ccagggcacg gggaactctt 30720

gcaggacagc gaaccccgca gaacagggca atcctcgcac agaacttaca ttgtgcatgg 30780

acagggtatc gcaatcaggc agcaccgggt gatcctccac cagagaagcg cgggtctcgg 30840

tctcctcaca gcgtggtaag ggggccggcc gatacgggtg atggcgggac gcggctgatc 30900

gtgttcgcga ccgtgtcatg atgcagttgc tttcggacat tttcgtactt gctgtagcag 30960

aacctggtcc gggcgctgca caccgatcgc cggcggcggt ctcggcgctt ggaacgctcg 31020

gtgttgaaat tgtaaaacag ccactctctc agaccgtgca gcagatctag ggcctcagga 31080

gtgatgaaga tcccatcatg cctgatggct ctgatcacat cgaccaccgt ggaatgggcc 31140
agacccagcc agatgatgca attttgttgg gtttcggtga cggcgggggа gggaagaaca 31200
ggaagaacca tgattaactt ttaatccaaa cggtctcgga gtacttcaaa atgaagatcg 31260
cggagatggc acctctcgcc cccgctgtgt tggtggaaaa taacagccag gtcaaaggtg 31320
atacggttct cgagatgttc cacggtggct tccagcaaag cctccacgcg cacatccaga 31380
aacaagacaa tagcgaaagc gggagggttc tctaattcct caatcatcat gttacactcc 31440
tgcaccatcc ccagataatt ttcatttttc cagccttgaa tgattcgaac tagttcgtga 31500

ggtaaatcca agccagccat gataaagagc tcgcgcagag cgccctccac cggcattctt 31560
aagcacaccc tcataattcc aagatattct gctcctggtt cacctgcagc agattgacaa 31620
gcggaatatc aaaatctctg ccgcgatccc tgagctcctc cctcagcaat aactgtaagt 31680
actctttcat atcctctccg aaatttttag ccataggacc accaggaata agattagggc 31740
aagccacagt acagataaac cgaagtcctc cccagtgagc attgccaaat gcaagactgc 31800
tataagcatg ctggctagac ccggtgatat cttccagata actggacaga aaatcgccca 31860
ggcaattttt aagaaaatca acaaaagaaa aatcctccag gtggacgttt agagcctcgg 31920

gaacaacgat gaagtaaatg caagcggtgc gttccagcat ggttagttag ctgatctgta 31980
gaaaaaacaa aaatgaacat taaaccatgc tagcctggcg aacaggtggg taaatcgttc 32040
tctccagcac caggcaggcc acggggtctc cggcgcgacc ctcgtaaaaa ttgtcgctat 32100
gattgaaaac catcacagag agacgttccc ggtggccggc gtgaatgatt cgacaagatg 32160
aatacacccc cggaacattg gcgtccgcga gtgaaaaaaa gcgcccgagg aagcaataag 32220
gcactacaat gctcagtctc aagtccagca aagcgatgcc atgcggatga agcacaaaat 32280
tctcaggtgc gtacaaaatg taattactcc cctcctgcac aggcagcaaa gcccccgatc 32340

cctccaggta cacatacaaa gcctcagcgt ccatagctta ccgagcagca gcacacaaca 32400
ggcgcaagag tcagagaaag gctgagctct aacctgtcca cccgctctct gctcaatata 32460
tagcccagat ctacactgac gtaaaggcca aagtctaaaa atacccgcca aataatcaca 32520
cacgcccagc acacgcccag aaaccggtga cacactcaaa aaaatacgcg cacttcctca 32580
aacgcccaaa actgccgtca tttccgggtt cccacgctac gtcatcaaaa cacgactttc 32640
aaattccgtc gaccgttaaa aacgtcaccc gccccgcccc taacggtcgc ccgtctctca 32700
gccaatcagc gccccgcatc cccaaattca aacacctcat ttgcatatta acgcgcacaa 32760

aaagtttgag gtatattatt gatgatgg 32788

<210> 13

<211> 30684

<212> DNA

<213> Artificial Sequence

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

<400> 13

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg 1320

acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca 1380

agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg 1440

tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac atgaagcagc 1500

acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca 1560
aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga 1620
accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc 1680
tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca 1740
tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc 1800
actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac aaccactacc 1860
tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc 1920

tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag 1980
tttaaactcc catttaaatg tgagggttaa tgcttcgagc agacatgata agatacattg 2040
atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 2100
gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 2160
attgcattca ttttatgttt caggttcagg gggagatgtg ggaggttttt taaagcaagt 2220
aaaacctcta caaatgtggt aaaataacta taacggtcct aaggtagcga gtgagtagtg 2280
ttctggggcg ggggaggacc tgcatgaggg ccagaataac tgaaatctgt gcttttctgt 2340

gtgttgcagc agcatgagcg gaagcggctc ctttgaggga ggggtattca gcccttatct 2400
gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat ccacggtgga 2460
cggccggccc gtgcagcccg cgaactcttc aaccctgacc tatgcaaccc tgagctcttc 2520
gtcgttggac gcagctgccg ccgcagctgc tgcatctgcc gccagcgccg tgcgcggaat 2580
ggccatgggc gccggctact acggcactct ggtggccaac tcgagttcca ccaataatcc 2640
cgccagcctg aacgaggaga agctgttgct gctgatggcc cagctcgagg ccttgaccca 2700
gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gagcagacgc gggccgcggt 2760

tgccacggtg aaatccaaat aaaaaatgaa tcaataaata aacggagacg gttgttgatt 2820
ttaacacaga gtctgaatct ttatttgatt tttcgcgcgc ggtaggccct ggaccaccgg 2880
tctcgatcat tgagcacccg gtggatcttt tccaggaccc ggtagaggtg ggcttggatg 2940
ttgaggtaca tgggcatgag cccgtcccgg gggtggaggt agctccattg cagggcctcg 3000
tgctcggggg tggtgttgta aatcacccag tcatagcagg ggcgcagggc atggtgttgc 3060
acaatatctt tgaggaggag actgatggcc acgggcagcc ctttggtgta ggtgtttaca 3120
aatctgttga gctgggaggg atgcatgcgg ggggagatga ggtgcatctt ggcctggatc 3180

ttgagattgg cgatgttacc gcccagatcc cgcctggggt tcatgttgtg caggaccacc 3240

agcacggtgt atccggtgca cttggggaat ttatcatgca acttggaagg gaaggcgtga 3300

aagaatttgg cgacgccttt gtgcccgccc aggttttcca tgcactcatc catgatgatg 3360

gcgatgggcc cgtgggcggc ggcctgggca aagacgtttc gggggtcgga cacatcatag 3420

ttgtggtcct gggtgaggtc atcataggcc attttaatga atttggggcg gagggtgccg 3480

gactgggggа caaaggtacc ctcgatcccg ggggcgtagt tcccctcaca gatctgcatc 3540

tcccaggctt tgagctcgga ggggggggatc atgtccacct gcggggcgat aaagaacacg 3600

gtttccgggg cgggggagat gagctgggcc gaaagcaagt tccggagcag ctgggacttg 3660

ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgagggag 3720

agacagctgc cgtcctcccg gaggaggggg gccacctcgt tcatcatctc gcgcacgtgc 3780

atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg 3840

agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt 3900

tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc 3960

agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg 4020

cgtccagcgc agccagggtc cggtccttcc agggtcgcag cgtccgcgtc agggtggtct 4080

ccgtcacggt gaaggggtgc gcgccgggct gggcgcttgc gagggtgcgc ttcaggctca 4140

tccggctggt cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga 4200

ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcggagc ttacctttgg 4260

aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga 4320

agacggactc gggggcgtag gcgtccgcgc cgcagtgggc gcagacggtc tcgcactcca 4380

cgagccaggt gaggtcgggc tggtcggggt caaaaaccag tttcccgccg ttctttttga 4440

tgcgtttctt acctttggtc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt 4500

ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cggtgtgccg cggtcctcct 4560

cgtagaggaa ccccgcccac tccgagacga aagcccgggt ccaggccagc acgaaggagg 4620

ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc cacctttcc agggtatgca 4680

aacacatgtc cccctcgtcc acatccagga aggtgattgg cttgtaagtg taggccacgt 4740

gaccgggggt cccggccggg ggggtataaa agggtgcggg tccctgctcg tcctcactgt 4800

cttccggatc gctgtccagg agcgccagct gttggggtag gtattccctc tcgaaggcgg 4860

gcatgacctc ggcactcagg ttgtcagttt ctagaaacga ggaggatttg atattgacgg 4920

tgccggcgga gatgcctttc aagagcccct cgtccatctg gtcagaaaag acgatctttt 4980

tgttgtcgag cttggtggcg aaggagccgt agagggcgtt ggagaggagc ttggcgatgg 5040

agcgcatggt ctggtttttt tccttgtcgg cgcgctcctt ggcggcgatg ttgagctgca 5100

cgtactcgcg cgccacgcac ttccattcgg ggaagacggt ggtcagctcg tcgggcacga 5160

ttctgacctg ccagccccga ttatgcaggg tgatgaggtc cacactggtg gccacctcgc 5220

cgcgcagggg ctcattagtc cagcagaggc gtccgccctt gcgcgagcag aaggggggca 5280

gggggtccag catgacctcg tcgggggggt cggcatcgat ggtgaagatg ccgggcagga 5340

ggtcggggtc aaagtagctg atggaagtgg ccagatcgtc cagggcagct tgccattcgc 5400

gcacggccag cgcgcgctcg tagggactga ggggcgtgcc ccagggcatg ggatgggtaa 5460

gcgcggaggc gtacatgccg cagatgtcgt agacgtagag gggctcctcg aggatgccga 5520

tgtaggtggg gtagcagcgc cccccgcgga tgctggcgcg cacgtagtca tacagctcgt 5580

gcgaggggggc gaggagcccc gggcccaggt tggtgcgact gggcttttcg gcgcggtaga 5640

cgatctggcg gaaaatggca tgcgagttgg aggagatggt gggcctttgg aagatgttga 5700

agtgggcgtg gggcagtccg accgagtcgc ggatgaagtg ggcgtaggag tcttgcagct 5760

tggcgacgag ctcggcggtg actaggacgt ccagagcgca gtagtcgagg gtctcctgga 5820

tgatgtcata cttgagctgt ccctttttgtt tccacagctc gcggttgaga aggaactctt 5880

cgcggtcctt ccagtactct tcgaggggga acccgtcctg atctgcacgg taagagccta 5940

gcatgtagaa ctggttgacg gccttgtagg cgcagcagcc cttctccacg gggagggcgt 6000

aggcctgggc ggccttgcgc agggaggtgt gcgtgagggc gaaagtgtcc ctgaccatga 6060

ccttgaggaa ctggtgcttg aagtcgatat cgtcgcagcc cccctgctcc cagagctgga 6120

agtccgtgcg cttcttgtag gcggggttgg gcaaagcgaa agtaacatcg ttgaagagga 6180

tcttgcccgc gcggggcata aagttgcgag tgatgcggaa aggttggggc acctcggccc 6240

ggttgttgat gacctgggcg gcgagcacga tctcgtcgaa gccgttgatg ttgtggccca 6300

cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct 6360

cgtaggtgag ctcgtcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat 6420

gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcggtt tgcagacggt 6480

cccggtactg acggaactgc tgcccgacgg ccatttttttc gggggtgacg cagtagaagg 6540

tgcgggggtc cccgtgccag cgatcccatt tgagctggag ggcgagatcg agggcgagct 6600

cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc 6660

cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcggtgc 6720

gaggatgcga gccgatgggg aagaactgga tctcctgcca ccaattggag gaatggctgt 6780

tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca 6840

agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag 6900

ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta 6960

ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcatg ctgacgagcc 7020

cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc 7080

gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggcagcggcg 7140

gcgcgcggtt gacttgcagg agtttttcca gggcgcgcgg gaggtccaga tggtacttga 7200

tctccaccgc gccattggtg gcgacgtcga tggcttgcag ggtcccgtgc ccctggggtg 7260

tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcggt gcctcttcca 7320

tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggcggctcgg ggcccggagg 7380

caggggcggc aggggcacgt cggcgccgcg cgcgggtagg ttctggtact gcgcccggag 7440

aagactggcg tgagcgacga cgcgacggtt gacgtcctgg atctgacgcc tctgggtgaa 7500

ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc 7560

gttgacggcg gcctgccgca ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat 7620

ctcggtcatg aactgctcga tctcctcctc ttgaaggtct ccgcggccgg cgcgctccac 7680

ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc 7740

ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac 7800

ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta 7860

gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg 7920

gagcggcatc tcgctgacgt cgcccagcgc ctccaaacgt tccatggcct cgtaaaagtc 7980

cacggcgaag ttgaaaaact gggagttgcg cgccgagacg gtcaactcct cctccagaag 8040

acggatgagc tcggcgatgg tggcgcgcac ctcgcgctcg aaggcccccg ggagttcctc 8100

cacttcctct tcttcctcct ccactaacat ctcttctact tcctcctcag gcggcagtgg 8160

tggcgggggа gggggcctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgctc 8220

gatggtctcg ccgcgccggc gtcgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg 8280

ccgcagcgtg aagacgccgc cgcgcatctc caggtggccg ggggggtccc cgttgggcag 8340

ggagagggcg ctgacgatgc atcttatcaa ttgccccgta gggactccgc gcaaggacct 8400

gagcgtctcg agatccacgg gatctgaaaa ccgctgaacg aaggcttcga gccagtcgca 8460

gtcgcaaggt aggctgagca cggtttcttc tggcgggtca tgttggttgg gagcggggcg 8520

ggcgatgctg ctggtgatga agttgaaata ggcggttctg agacggcgga tggtggcgag 8580

gagcaccagg tctttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc 8640

gtggtcctga cacctggcca ggtccttgta gtagtcctgc atgagccgct ccacgggcac 8700

ctcctcctcg cccgcgcggc cgtgcatgcg cgtgagcccg aagccgcgct ggggctggac 8760

gagcgccagg tcggcgacga cgcgctcggc gaggatggct tgctggatct gggtgagggt 8820

ggtctggaag tcatcaaagt cgacgaagcg gtggtaggct ccggtgttga tggtgtagga 8880

gcagttggcc atgacggacc agttgacggt ctggtggccc ggacgcacga gctcgtggta 8940

cttgaggcgc gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggtgc gcaccaggta 9000

ctggtagccg atgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc 9060

gggggcgccg ggcgcgaggt cctcgagcat ggtgcggtgg tagccgtaga tgtacctgga 9120

catccaggtg atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca 9180

gatgttgcgc agcggcagga agtagttcat ggtgggcacg gtctggcccg tgaggcgcgc 9240

gcagtcgtgg atgctctata cgggcaaaaa cgaaagcggt cagcggctcg actccgtggc 9300

ctggaggcta agcgaacggg ttgggctgcg cgtgtacccc ggttcgaatc tcgaatcagg 9360

ctggagccgc agctaacgtg gtattggcac tcccgtctcg acccaagcct gcaccaaccc 9420

tccaggatac ggaggcgggt cgttttgcaa ctttttttg gaggccggat gagactagta 9480

agcgcggaaa gcggccgacc gcgatggctc gctgccgtag tctggagaag aatcgccagg 9540

gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg 9600

gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc 9660

cctcttttgt tttgtttgtt tttgccagat gcatcccgta ctgcggcaga tgcgccccca 9720

ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca 9780

gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca 9840

ccagctggcc ttggaagagg gcgaggggct ggcgcgcctg ggggcgtcgt cgccggagcg 9900

gcacccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgccca agcagaacct 9960

gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggcccggt tccacgcggg 10020

gcgggagctg cggcgcggcc tggaccgaaa gagggtgctg agggacgagg atttcgaggc 10080

ggacgagctg acggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac 10140

ggcgtacgag cagaccgtga aggaggagag caacttccaa aaatccttca acaaccacgt 10200

gcgcaccctg atcgcgcgcg aggaggtgac cctgggcctg atgcacctgt gggacctgct 10260

ggaggccatc gtgcagaacc ccaccagcaa gccgctgacg gcgcagctgt tcctggtggt 10320

gcagcatagt cgggacaacg aagcgttcag ggaggcgctg ctgaatatca ccgagcccga 10380

gggccgctgg ctcctggacc tggtgaacat tctgcagagc atcgtggtgc aggagcgcgg 10440

gctgccgctg tccgagaagc tggcggccat caacttctcg gtgctgagtt tgggcaagta 10500

ctacgctagg aagatctaca agaccccgta cgtgcccata gacaaggagg tgaagatcga 10560

cgggttttac atgcgcatga ccctgaaagt gctgaccctg agcgacgatc tgggggtgta 10620

ccgcaacgac aggatgcacc gtgcggtgag cgccagcagg cggcgcgagc tgagcgacca 10680

ggagctgatg catagtctgc agcgggccct gaccggggcc gggaccgagg gggagagcta 10740

ctttgacatg ggcgcggacc tgcactggca gcccagccgc cgggccttgg aggcggcggc 10800

aggaccctac gtagaagagg tggacgatga ggtggacgag gagggcgagt acctggaaga 10860

ctgatggcgc gaccgtattt ttgctagatg caacaacaac agccacctcc tgatcccgcg 10920

atgcgggcgg cgctgcagag ccagccgtcc ggcattaact cctcggacga ttggacccag 10980

gccatgcaac gcatcatggc gctgacgacc cgcaaccccg aagcctttag acagcagccc 11040

caggccaacc ggctctcggc catcctggag gccgtggtgc cctcgcgctc caaccccacg 11100

cacgagaagg tcctggccat cgtgaacgcg ctggtggaga acaaggccat ccgcggcgac 11160

gaggccggcc tggtgtacaa cgcgctgctg gagcgcgtgg cccgctacaa cagcaccaac 11220

gtgcagacca acctggaccg catggtgacc gacgtgcgcg aggccgtggc ccagcgcgag 11280

cggttccacc gcgagtccaa cctgggatcc atggtggcgc tgaacgcctt cctcagcacc 11340

cagcccgcca acgtgccccg gggccaggag gactacacca acttcatcag cgccctgcgc 11400

ctgatggtga ccgaggtgcc ccagagcgag gtgtaccagt ccgggccgga ctacttcttc 11460

cagaccagtc gccagggctt gcagaccgtg aacctgagcc aggctttcaa gaacttgcag 11520

ggcctgtggg gcgtgcaggc cccggtcggg gaccgcgcga cggtgtcgag cctgctgacg 11580

ccgaactcgc gcctgctgct gctgctggtg gcccccttca cggacagcgg cagcatcaac 11640

cgcaactcgt acctgggcta cctgattaac ctgtaccgcg aggccatcgg ccaggcgcac 11700

gtggacgagc agacctacca ggagatcacc cacgtgagcc gcgccctggg ccaggacgac 11760

ccgggcaacc tggaagccac cctgaacttt ttgctgacca accggtcgca gaagatcccg 11820

ccccagtacg cgctcagcac cgaggaggag cgcatcctgc gttacgtgca gcagagcgtg 11880

ggcctgttcc tgatgcagga gggggccacc cccagcgccg cgctcgacat gaccgcgcgc 11940

aacatggagc ccagcatgta cgccagcaac cgcccgttca tcaataaact gatggactac 12000

ttgcatcggg cggccgccat gaactctgac tatttcacca acgccatcct gaatccccac 12060

tggctcccgc cgccggggtt ctacacgggc gagtacgaca tgcccgaccc caatgacggg 12120

ttcctgtggg acgatgtgga cagcagcgtg ttctcccccc gaccgggtgc taacgagcgc 12180

cccttgtgga agaaggaagg cagcgaccga cgcccgtcct cggcgctgtc cggccgcgag 12240

ggtgctgccg cggcggtgcc cgaggccgcc agtcctttcc cgagcttgcc cttctcgctg 12300

aacagtatcc gcagcagcga gctgggcagg atcacgcgcc cgcgcttgct gggcgaagag 12360

gagtacttga atgactcgct gttgagaccc gagcgggaga agaacttccc caataacggg 12420

atagaaagcc tggtggacaa gatgagccgc tggaagacgt atgcgcagga gcacagggac 12480
gatccccggg cgtcgcaggg ggccacgagc cggggcagcg ccgcccgtaa acgccggtgg 12540
cacgacaggc agcggggaca gatgtgggac gatgaggact ccgccgacga cagcagcgtg 12600
ttggacttgg gtgggagtgg taacccgttc gctcacctgc gcccccgtat cgggcgcatg 12660
atgtaagaga aaccgaaaat aaatgatact caccaaggcc atggcgacca gcgtgcgttc 12720
gtttcttctc tgttgttgtt gtatctagta tgatgaggcg tgcgtacccg gagggtcctc 12780
ctccctcgta cgagagcgtg atgcagcagg cgatggcggc ggcggcgatg cagcccccgc 12840

tggaggctcc ttacgtgccc ccgcggtacc tggcgcctac ggaggggcgg aacagcattc 12900
gttactcgga gctggcaccc ttgtacgata ccacccggtt gtacctggtg gacaacaagt 12960
cggcggacat cgcctcgctg aactaccaga acgaccacag caacttcctg accaccgtgg 13020
tgcagaacaa tgacttcacc cccacggagg ccagcaccca gaccatcaac tttgacgagc 13080
gctcgcggtg gggcggccag ctgaaaacca tcatgcacac caacatgccc aacgtgaacg 13140
agttcatgta cagcaacaag ttcaaggcgc gggtgatggt ctcccgcaag acccccaatg 13200
gggtgacagt gacagaggat tatgatggta gtcaggatga gctgaagtat gaatgggtgg 13260

aatttgagct gcccgaaggc aacttctcgg tgaccatgac catcgacctg atgaacaacg 13320
ccatcatcga caattacttg gcggtggggc ggcagaacgg ggtgctggag agcgacatcg 13380
gcgtgaagtt cgacactagg aacttcaggc tgggctggga ccccgtgacc gagctggtca 13440
tgcccggggt gtacaccaac gaggctttcc atcccgatat tgtcttgctg cccggctgcg 13500
gggtggactt caccgagagc cgcctcagca acctgctggg cattcgcaag aggcagccct 13560
tccaggaagg cttccagatc atgtacgagg atctggaggg gggcaacatc cccgcgctcc 13620
tggatgtcga cgcctatgag aaaagcaagg aggatgcagc agctgaagca actgcagccg 13680

tagctaccgc ctctaccgag gtcaggggcg ataattttgc aagcgccgca gcagtggcag 13740
cggccgaggc ggctgaaacc gaaagtaaga tagtcattca gccggtggag aaggatagca 13800
agaacaggag ctacaacgta ctaccggaca agataaacac cgcctaccgc agctggtacc 13860
tagcctacaa ctatggcgac cccgagaagg gcgtgcgctc ctggacgctg ctcaccacct 13920
cggacgtcac ctgcggcgtg gagcaagtct actggtcgct gcccgacatg atgcaagacc 13980
cggtcacctt ccgctccacg cgtcaagtta gcaactaccc ggtggtgggc gccgagctcc 14040
tgcccgtcta ctccaagagc ttcttcaacg agcaggccgt ctactcgcag cagctgcgcg 14100

ccttcacctc gcttacgcac gtcttcaacc gcttccccga gaaccagatc ctcgtccgcc 14160

cgcccgcgcc caccattacc accgtcagtg aaaacgttcc tgctctcaca gatcacggga 14220

ccctgccgct gcgcagcagt atccggggag tccagcgcgt gaccgttact gacgccagac 14280

gccgcacctg cccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtcctctcga 14340

gccgcacctt ctaaatgtcc attctcatct cgcccagtaa taacaccggt tggggcctgc 14400

gcgcgcccag caagatgtac ggaggcgctc gccaacgctc cacgcaacac cccgtgcgcg 14460

tgcgcgggca cttccgcgct ccctggggcg ccctcaaggg ccgcgtgcgg tcgcgcacca 14520

ccgtcgacga cgtgatcgac caggtggtgg ccgacgcgcg caactacacc cccgccgccg 14580

cgcccgtctc caccgtggac gccgtcatcg acagcgtggt ggccgacgcg cgccggtacg 14640

cccgcgccaa gagccggcgg cggcgcatcg cccggcggca ccggagcacc cccgccatgc 14700

gcgcggcgcg agccttgctg cgcagggcca ggcgcacggg acgcagggcc atgctcaggg 14760

cggccagacg cgcggcttca ggcgccagcg ccggcaggac ccggagacgc gcggccacgg 14820

cggcggcagc ggccatcgcc agcatgtccc gcccgcggcg agggaacgtg tactgggtgc 14880

gcgacgccgc caccggtgtg cgcgtgcccg tgcgcacccg ccccccctcgc acttgaagat 14940

gttcacttcg cgatgttgat gtgtcccagc ggcgaggagg atgtccaagc gcaaattcaa 15000

ggaagagatg ctccaggtca tcgcgcctga gatctacggc cctgcggtgg tgaaggagga 15060

aagaaagccc cgcaaaatca agcgggtcaa aaaggacaaa aaggaagaag aaagtgatgt 15120

ggacggattg gtggagtttg tgcgcgagtt cgcccccccgg cggcgcgtgc agtggcgcgg 15180

gcggaaggtg caaccggtgc tgagacccgg caccaccgtg gtcttcacgc ccggcgagcg 15240

ctccggcacc gcttccaagc gctcctacga cgaggtgtac ggggatgatg atattctgga 15300

gcaggcggcc gagcgcctgg gcgagtttgc ttacggcaag cgcagccgtt ccgcaccgaa 15360

ggaagaggcg gtgtccatcc cgctggacca cggcaacccc acgccgagcc tcaagcccgt 15420

gaccttgcag caggtgctgc cgaccgcggc gccgcgccgg gggttcaagc gcgagggcga 15480

ggatctgtac cccaccatgc agctgatggt gcccaagcgc cagaagctgg aagacgtgct 15540

ggagaccatg aaggtggacc cggacgtgca gcccgaggtc aaggtgcggc ccatcaagca 15600

ggtggccccg ggcctgggcg tgcagaccgt ggacatcaag attcccacgg agcccatgga 15660

aacgcagacc gagcccatga tcaagcccag caccagcacc atggaggtgc agacggatcc 15720

ctggatgcca tcggctccta gtcgaagacc ccggcgcaag tacggcgcgg ccagcctgct 15780

gatgcccaac tacgcgctgc atccttccat catccccacg ccgggctacc gcggcacgcg 15840

cttctaccgc ggtcatacca gcagccgccg ccgcaagacc accactcgcc gccgccgtcg 15900

ccgcaccgcc gctgcaacca cccctgccgc cctggtgcgg agagtgtacc gccgcggccg 15960

cgcacctctg accctgccgc gcgcgcgcta ccacccgagc atcgccattt aaactttcgc 16020

ctgctttgca gatcaatggc cctcacatgc cgccttcgcg ttcccattac gggctaccga 16080

ggaagaaaac cgcgccgtag aaggctggcg gggaacggga tgcgtcgcca ccaccaccgg 16140

cggcggcgcg ccatcagcaa gcggttgggg ggaggcttcc tgcccgcgct gatccccatc 16200

atcgccgcgg cgatcggggc gatccccggc attgcttccg tggcggtgca ggcctctcag 16260

cgccactgag acacacttgg aaacatcttg taataaacca atggactctg acgctcctgg 16320

tcctgtgatg tgttttcgta gacagatgga agacatcaat ttttcgtccc tggctccgcg 16380

acacggcacg cggccgttca tgggcacctg gagcgacatc ggcaccagcc aactgaacgg 16440

gggcgccttc aattggagca gtctctggag cgggcttaag aatttcgggt ccacgcttaa 16500

aacctatggc agcaaggcgt ggaacagcac cacagggcag gcgctgaggg ataagctgaa 16560

agagcagaac ttccagcaga aggtggtcga tgggctcgcc tcgggcatca acggggtggt 16620

ggacctggcc aaccaggccg tgcagcggca gatcaacagc cgcctggacc cggtgccgcc 16680

cgccggctcc gtggagatgc cgcaggtgga ggaggagctg cctcccctgg acaagcgggg 16740

cgagaagcga ccccgccccg atgcggagga gacgctgctg acgcacacgg acgagccgcc 16800

cccgtacgag gaggcggtga aactgggtct gcccaccacg cggcccatcg cgcccctggc 16860

caccggggtg ctgaaacccg aaaagcccgc gaccctggac ttgcctcctc cccagccttc 16920

ccgcccctct acagtggcta agcccctgcc gccggtggcc gtggcccgcg cgcgacccgg 16980

gggcaccgcc cgccctcatg cgaactggca gagcactctg aacagcatcg tgggtctggg 17040

agtgcagagt gtgaagcgcc gccgctgcta ttaaacctac cgtagcgctt aacttgcttg 17100

tctgtgtgtg tatgtattat gtcgccgccg ccgctgtcca ccagaaggag gagtgaagag 17160

gcgcgtcgcc gagttgcaag atggccaccc catcgatgct gccccagtgg gcgtacatgc 17220

acatcgccgg acaggacgct tcggagtacc tgagtccggg tctggtgcag tttgcccgcg 17280

ccacagacac ctacttcagt ctggggaaca agtttaggaa ccccacggtg gcgcccacgc 17340

acgatgtgac caccgaccgc agccagcggc tgacgctgcg cttcgtgccc gtggaccgcg 17400

aggacaacac ctactcgtac aaagtgcgct acacgctggc cgtgggcgac aaccgcgtgc 17460

tggacatggc cagcacctac tttgacatcc gcggcgtgct ggatcggggc cctagcttca 17520

aaccctactc cggcaccgcc tacaacagtc tggcccccaa gggagcaccc aacacttgtc 17580

agtggacata taaagccgat ggtgaaactg ccacagaaaa aacctataca tatggaaatg 17640

cacccgtgca gggcattaac atcacaaaag atggtattca acttggaact gacaccgatg 17700

atcagccaat ctacgcagat aaaacctatc agcctgaacc tcaagtgggt gatgctgaat 17760

ggcatgacat cactggtact gatgaaaagt atggaggcag agctcttaag cctgatacca 17820

aaatgaagcc ttgttatggt tcttttgcca agcctactaa taaagaagga ggtcaggcaa 17880

atgtgaaaac aggaacaggc actactaaag aatatgacat agacatggct ttctttgaca 17940
acagaagtgc ggctgctgct ggcctagctc cagaaattgt tttgtatact gaaaatgtgg 18000
atttggaaac tccagatacc catattgtat acaaagcagg cacagatgac agcagctctt 18060
ctattaattt gggtcagcaa gccatgccca acagacctaa ctacattggt ttcagagaca 18120
actttatcgg gctcatgtac tacaacagca ctggcaatat gggggtgctg gccggtcagg 18180
cttctcagct gaatgctgtg gttgacttgc aagacagaaa caccgagctg tcctaccagc 18240
tcttgcttga ctctctgggt gacagaaccc ggtatttcag tatgtggaat caggcggtgg 18300

acagctatga tcctgatgtg cgcattattg aaaatcatgg tgtggaggat gaacttccca 18360
actattgttt ccctctggat gctgttggca gaacagatac ttatcaggga attaaggcta 18420
atggaactga tcaaaccaca tggaccaaag atgacagtgt caatgatgct aatgagatag 18480
gcaagggtaa tccattcgcc atggaaatca acatccaagc caacctgtgg aggaacttcc 18540
tctacgccaa cgtggccctg tacctgcccg actcttacaa gtacacgccg gccaatgtta 18600
ccctgcccac caacaccaac acctacgatt acatgaacgg ccgggtggtg gcgccctcgc 18660
tggtggactc ctacatcaac atcggggcgc gctggtcgct ggatcccatg gacaacgtga 18720

acccccttcaa ccaccaccgc aatgcggggc tgcgctaccg ctccatgctc ctgggcaacg 18780
ggcgctacgt gcccttccac atccaggtgc cccagaaatt tttcgccatc aagagcctcc 18840
tgctcctgcc cgggtcctac acctacgagt ggaacttccg caaggacgtc aacatgatcc 18900
tgcagagctc cctcggcaac gacctgcgca cggacggggc ctccatctcc ttcaccagca 18960
tcaacctcta cgccaccttc ttccccatgg cgcacaacac ggcctccacg ctcgaggcca 19020
tgctgcgcaa cgacaccaac gaccagtcct tcaacgacta cctctcggcg gccaacatgc 19080
tctaccccat cccggccaac gccaccaacg tgcccatctc catcccctcg cgcaactggg 19140

ccgccttccg cggctggtcc ttcacgcgtc tcaagaccaa ggagacgccc tcgctgggct 19200
ccgggttcga cccctacttc gtctactcgg gctccatccc ctacctcgac ggcaccttct 19260
acctcaacca caccttcaag aaggtctcca tcaccttcga ctcctccgtc agctggcccg 19320
gcaacgaccg gctcctgacg cccaacgagt tcgaaatcaa gcgcaccgtc gacggcgagg 19380
gctacaacgt ggcccagtgc aacatgacca aggactggtt cctggtccag atgctggccc 19440
actacaacat cggctaccag ggcttctacg tgcccgaggg ctacaaggac cgcatgtact 19500
ccttcttccg caacttccag cccatgagcc gccaggtggt ggacgaggtc aactacaagg 19560

actaccaggc cgtcaccctg gcctaccagc acaacaactc gggcttcgtc ggctacctcg 19620

cgcccaccat gcgccagggc cagccctacc ccgccaacta cccctacccg ctcatcggca 19680

agagcgccgt caccagcgtc acccagaaaa agttcctctg cgacagggtc atgtggcgca 19740

tccccttctc cagcaacttc atgtccatgg gcgcgctcac cgacctcggc cagaacatgc 19800

tctatgccaa ctccgcccac gcgctagaca tgaatttcga agtcgacccc atggatgagt 19860

ccacccttct ctatgttgtc ttcgaagtct tcgacgtcgt ccgagtgcac cagccccacc 19920

gcggcgtcat cgaggccgtc tacctgcgca ccccccttctc ggccggtaac gccaccacct 19980

aagctcttgc ttcttgcaag ccatggccgc gggctccggc gagcaggagc tcagggccat 20040

catccgcgac ctgggctgcg ggccctactt cctgggcacc ttcgataagc gcttcccggg 20100

attcatggcc ccgcacaagc tggcctgcgc catcgtcaac acggccggcc gcgagaccgg 20160

gggcgagcac tggctggcct tcgcctggaa cccgcgctcg aacacctgct acctcttcga 20220

ccccttcggg ttctcggacg agcgcctcaa gcagatctac cagttcgagt acgagggcct 20280

gctgcgccgc agcgccctgg ccaccgagga ccgctgcgtc accctggaaa agtccaccca 20340

gaccgtgcag ggtccgcgct cggccgcctg cgggctcttc tgctgcatgt tcctgcacgc 20400

cttcgtgcac tggcccgacc gccccatgga caagaacccc accatgaact tgctgacggg 20460

ggtgcccaac ggcatgctcc agtcgcccca ggtggaaccc accctgcgcc gcaaccagga 20520

ggcgctctac cgcttcctca actcccactc cgcctacttt cgctcccacc gcgcgcgcat 20580

cgagaaggcc accgccttcg accgcatgaa tcaagacatg taaaccgtgt gtgtatgtta 20640

aatgtcttta ataaacagca ctttcatgtt acacatgcat ctgagatgat ttatttagaa 20700

atcgaaaggg ttctgccggg tctcggcatg gcccgcgggc agggacacgt tgcggaactg 20760

gtacttggcc agccacttga actcggggat cagcagtttg ggcagcgggg tgtcggggaa 20820

ggagtcggtc cacagcttcc gcgtcagttg cagggcgccc agcaggtcgg gcgcggagat 20880

cttgaaatcg cagttgggac ccgcgttctg cgcgcgggag ttgcggtaca cggggttgca 20940

gcactggaac accatcaggg ccgggtgctt cacgctcgcc agcaccgtcg cgtcggtgat 21000

gctctccacg tcgaggtcct cggcgttggc catcccgaag ggggtcatct tgcaggtctg 21060

ccttcccatg gtgggcacgc acccgggctt gtggttgcaa tcgcagtgca gggggatcag 21120

catcatctgg gcctggtcgg cgttcatccc cgggtacatg gccttcatga aagcctccaa 21180

ttgcctgaac gcctgctggg ccttggctcc ctcggtgaag aagaccccgc aggacttgct 21240

agagaactgg ttggtggcgc acccggcgtc gtgcacgcag cagcgcgcgt cgttgttggc 21300

cagctgcacc acgctgcgcc cccagcggtt ctgggtgatc ttggcccggt cggggttctc 21360

cttcagcgcg cgctgcccgt tctcgctcgc cacatccatc tcgatcatgt gctccttctg 21420

gatcatggtg gtcccgtgca ggcaccgcag cttgccctcg gcctcggtgc acccgtgcag 21480

ccacagcgcg cacccggtgc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac 21540

gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggtcag 21600

cggaatgccg cggtgctcct cgttgatgta caggtggcag atgcggcggt acacctcgcc 21660

ctgctcgggc atcagctgga agttggcttt caggtcggtc tccacgcggt agcggtccat 21720

cagcatagtc atgatttcca tacccttctc ccaggccgag acgatgggca ggctcatagg 21780

gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtcgctct cgtccagggt 21840

ctcaaagctc cgcttgccgt ccttctcggt gatccgcacc ggggggtagc tgaagcccac 21900

ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag 21960

gaccacatgc ttggtcttgc ggggtttctt cttgggcggc agcggcggcg gagatgttgg 22020

agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc 22080

cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcggaggcg acgggctctc 22140

gccgccgcga cttggcggat ggctggcaga gcccccttccg cgttcgggggg tgcgctcccg 22200

gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac 22260

aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgccg acgagaagca 22320

gcagcagcag aatgaaagct taaccgcccc gccgcccagc cccgccacct ccgacgcggc 22380

cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc 22440

cgcggagcac gaggaggagc tggcagtgcg cttttcacaa gaagagatac accaagaaca 22500

gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta 22560

cctccacctg agcgggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat 22620

cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg 22680

cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac 22740

ctgcgagccc aacccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc 22800

cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac 22860

ccgcgccgac gccctttca acctgggtcc cggcgcccgc ctacctgata tcgcctcctt 22920

ggaagaggtt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc 22980

tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga 23040

caacgcgcgg ctggcggtgc tcaaacgcac ggtcgagctg acccatttcg cctacccggc 23100

tctgaacctg ccccccaaag tcatgagcgc ggtcatggac caggtgctca tcaagcgcgc 23160

gtcgcccatc tccgaggacg agggcatgca agactccgag gagggcaagc ccgtggtcag 23220

cgacgagcag ctggcccggt ggctgggtcc taatgctagt ccccagagtt tggaagagcg 23280

gcgcaaactc atgatggccg tggtcctggt gaccgtggag ctggagtgcc tgcgccgctt 23340

cttcgccgac gcggagaccc tgcgcaaggt cgaggagaac ctgcactacc tcttcaggca 23400
cgggttcgtg cgccaggcct gcaagatctc caacgtggag ctgaccaacc tggtctccta 23460
catgggcatc ttgcacgaga accgcctggg gcagaacgtg ctgcacacca ccctgcgcgg 23520
ggaggcccgg cgcgactaca tccgcgactg cgtctacctc tacctctgcc acacctggca 23580
gacgggcatg ggcgtgtggc agcagtgtct ggaggagcag aacctgaaag agctctgcaa 23640
gctcctgcag aagaacctca agggtctgtg gaccgggttc gacgagcgca ccaccgcctc 23700
ggacctggcc gacctcattt tccccgagcg cctcaggctg acgctgcgca acggcctgcc 23760

cgactttatg agccaaagca tgttgcaaaa ctttcgctct ttcatcctcg aacgctccgg 23820
aatcctgccc gccacctgct ccgcgctgcc ctcggacttc gtgccgctga ccttccgcga 23880
gtgccccccg ccgctgtgga gccactgcta cctgctgcgc ctggccaact acctggccta 23940
ccactcggac gtgatcgagg acgtcagcgg cgagggcctg ctcgagtgcc actgccgctg 24000
caacctctgc acgccgcacc gctccctggc ctgcaacccc cagctgctga gcgagaccca 24060
gatcatcggc accttcgagt tgcaagggcc cagcgaaggc gagggttcag ccgccaaggg 24120
gggtctgaaa ctcaccccgg ggctgtggac ctcggcctac ttgcgcaagt tcgtgcccga 24180

ggactaccat cccttcgaga tcaggttcta cgaggaccaa tcccatccgc ccaaggccga 24240
gctgtcggcc tgcgtcatca cccaggggggc gatcctggcc caattgcaag ccatccagaa 24300
atcccgccaa gaattcttgc tgaaaaaggg ccgcggggtc tacctcgacc cccagaccgg 24360
tgaggagctc aaccccggct tcccccagga tgccccgagg aaacaagaag ctgaaagtgg 24420
agctgccgcc cgtggaggat ttggaggaag actgggagaa cagcagtcag gcagaggagg 24480
aggagatgga ggaagactgg gacagcactc aggcagagga ggacagcctg caagacagtc 24540
tggaggaaga cgaggaggag gcagaggagg aggtggaaga agcagccgcc gccagaccgt 24600

cgtcctcggc gggggagaaa gcaagcagca cggataccat ctccgctccg ggtcggggtc 24660
ccgctcgacc acacagtaga tgggacgaga ccggacgatt cccgaacccc accacccaga 24720
ccggtaagaa ggagcggcag ggatacaagt cctggcgggg gcacaaaaac gccatcgtct 24780
cctgcttgca ggcctgcggg ggcaacatct ccttcacccg gcgctacctg ctcttccacc 24840
gcggggtgaa ctttccccgc aacatcttgc attactaccg tcacctccac agcccctact 24900
acttccaaga agaggcagca gcagcagaaa aagaccagca gaaaaccagc agctagaaaa 24960
tccacagcgg cggcagcagg tggactgagg atcgcggcga acgagccggc gcaaacccgg 25020

gagctgagga accggatctt tcccaccctc tatgccatct tccagcagag tcgggggcag 25080

gagcaggaac tgaaagtcaa gaaccgttct ctgcgctcgc tcacccgcag ttgtctgtat 25140

cacaagagcg aagaccaact tcagcgcact ctcgaggacg ccgaggctct cttcaacaag 25200

tactgcgcgc tcactcttaa agagtagccc gcgcccgccc agtcgcagaa aaaggcggga 25260

attacgtcac ctgtgccctt cgccctagcc gcctccaccc atcatcatga gcaaagagat 25320

tcccacgcct tacatgtgga gctaccagcc ccagatgggc ctggccgccg gtgccgccca 25380

ggactactcc acccgcatga attggctcag cgccgggccc gcgatgatct cacgggtgaa 25440

tgacatccgc gcccaccgaa accagatact cctagaacag tcagcgctca ccgccacgcc 25500

ccgcaatcac ctcaatccgc gtaattggcc cgccgccctg gtgtaccagg aaattcccca 25560

gcccacgacc gtactacttc cgcgagacgc ccaggccgaa gtccagctga ctaactcagg 25620

tgtccagctg gcgggcggcg ccaccctgtg tcgtcaccgc cccgctcagg gtataaagcg 25680

gctggtgatc cggggcagag gcacacagct caacgacgag gtggtgagct cttcgctggg 25740

tctgcgacct gacggagtct tccaactcgc cggatcgggg agatcttcct tcacgcctcg 25800

tcaggccgtc ctgactttgg agagttcgtc ctcgcagccc cgctcgggtg gcatcggcac 25860

tctccagttc gtggaggagt tcactccctc ggtctacttc aacccccttct ccggctcccc 25920

cggccactac ccggacgagt tcatcccgaa cttcgacgcc atcagcgagt cggtggacgg 25980

ctacgattga aactaatcac ccccttatcc agtgaaataa agatcatatt gatgatgatt 26040

ttacagaaat aaaaaataat catttgattt gaaataaaga tacaatcata ttgatgattt 26100

gagtttaaca aaaaaataaa gaatcactta cttgaaatct gataccaggt ctctgtccat 26160

gttttctgcc aacaccactt cactcccctc ttcccagctc tggtactgca ggccccggcg 26220

ggctgcaaac ttcctccaca cgctgaaggg gatgtcaaat tcctcctgtc cctcaatctt 26280

cattttatct tctatcagat gtccaaaaag cgcgtccggg tggatgatga cttcgacccc 26340

gtctacccct acgatgcaga caacgcaccg accgtgccct tcatcaaccc ccccttcgtc 26400

tcttcagatg gattccaaga gaagcccctg ggggtgttgt ccctgcgact ggccgacccc 26460

gtcaccacca agaacgggga aatcaccctc aagctgggag agggggtgga cctcgattcc 26520

tcgggaaaac tcatctccaa cacggccacc aaggccgccg cccctctcag tttttccaac 26580

aacaccattt cccttaacat ggatcacccc ttttacacta aagatggaaa attatcctta 26640

caagtttctc caccattaaa tatactgaga acaagcattc taaacacact agctttaggt 26700

tttggatcag gtttaggact ccgtggctct gccttggcag tacagttagt ctctccactt 26760

acatttgata ctgatggaaa cataaagctt accttagaca gaggtttgca tgttacaaca 26820

ggagatgcaa ttgaaagcaa cataagctgg gctaaaggtt taaaatttga agatggagcc 26880

atagcaacca acattggaaa tgggttagag tttggaagca gtagtacaga aacaggtgtt 26940

gatgatgctt acccaatcca agttaaactt ggatctggcc ttagctttga cagtacagga 27000

gccataatgg ctggtaacaa agaagacgat aaactcactt tgtggacaac acctgatcca 27060

tcaccaaact gtcaaatact cgcagaaaat gatgcaaaac taacactttg cttgactaaa 27120

tgtggtagtc aaatactggc cactgtgtca gtcttagttg taggaagtgg aaacctaaac 27180

cccattactg gcaccgtaag cagtgctcag gtgtttctac gttttgatgc aaacggtgtt 27240

cttttaacag aacattctac actaaaaaaa tactgggggt ataggcaggg agatagcata 27300

gatggcactc catataccaa tgctgtagga ttcatgccca atttaaaagc ttatccaaag 27360

tcacaaagtt ctactactaa aaataatata gtagggcaag tatacatgaa tggagatgtt 27420

tcaaaaccta tgcttctcac tataaccctc aatggtactg atgacagcaa cagtacatat 27480

tcaatgtcat tttcatacac ctggactaat ggaagctatg ttggagcaac atttggggct 27540

aactcttata ccttctcata catcgcccaa gaatgaacac tgtatcccac cctgcatgcc 27600

aacccttccc accccactct gtggaacaaa ctctgaaaca caaaataaaa taaagttcaa 27660

gtgttttatt gattcaacag ttttacagga ttcgagcagt tatttttcct ccaccctccc 27720

aggacatgga atacaccacc ctctcccccc gcacagcctt gaacatctga atgccattgg 27780

tgatggacat gcttttggtc tccacgttcc acacagtttc agagcgagcc agtctcgggt 27840

cggtcaggga gatgaaaccc tccgggcact cccgcatctg cacctcacag ctcaacagct 27900

gaggattgtc ctcggtggtc gggatcacgg ttatctggaa gaagcagaag agcggcggtg 27960

ggaatcatag tccgcgaacg ggatcggccg gtggtgtcgc atcaggcccc gcagcagtcg 28020

ctgccgccgc cgctccgtca agctgctgct cagggggtcc gggtccaggg actccctcag 28080

catgatgccc acggccctca gcatcagtcg tctggtgcgg cgggcgcagc agcgcatgcg 28140

gatctcgctc aggtcgctgc agtacgtgca acacagaacc accaggttgt tcaacagtcc 28200

atagttcaac acgctccagc cgaaactcat cgcgggaagg atgctaccca cgtggccgtc 28260

gtaccagatc ctcaggtaaa tcaagtggtg ccccctccag aacacgctgc ccacgtacat 28320

gatctccttg ggcatgtggc ggttcaccac ctcccggtac cacatcaccc tctggttgaa 28380

catgcagccc cggatgatcc tgcggaacca cagggccagc accgccccgc ccgccatgca 28440

gcgaagagac cccgggtccc ggcaatggca atggaggacc caccgctcgt acccgtggat 28500

catctgggag ctgaacaagt ctatgttggc acagcacagg catatgctca tgcatctctt 28560

cagcactctc aactcctcgg gggtcaaaac catatcccag ggcacgggga actcttgcag 28620

gacagcgaac cccgcagaac agggcaatcc tcgcacagaa cttacattgt gcatggacag 28680

ggtatcgcaa tcaggcagca ccgggtgatc ctccaccaga gaagcgcggg tctcggtctc 28740

ctcacagcgt ggtaaggggg ccggccgata cgggtgatgg cgggacgcgg ctgatcgtgt 28800

tcgcgaccgt gtcatgatgc agttgctttc ggacattttc gtacttgctg tagcagaacc 28860
tggtccgggc gctgcacacc gatcgccggc ggcggtctcg gcgcttggaa cgctcggtgt 28920
tgaaattgta aaacagccac tctctcagac cgtgcagcag atctagggcc tcaggagtga 28980
tgaagatccc atcatgcctg atggctctga tcacatcgac caccgtggaa tgggccagac 29040
ccagccagat gatgcaattt tgttgggttt cggtgacggc gggggaggga agaacaggaa 29100
gaaccatgat taacttttaa tccaaacggt ctcggagtac ttcaaaatga agatcgcgga 29160
gatggcacct ctcgcccccg ctgtgttggt ggaaaataac agccaggtca aaggtgatac 29220

ggttctcgag atgttccacg gtggcttcca gcaaagcctc cacgcgcaca tccagaaaca 29280
agacaatagc gaaagcggga gggttctcta attcctcaat catcatgtta cactcctgca 29340
ccatccccag ataattttca tttttccagc cttgaatgat tcgaactagt tcctgaggta 29400
aatccaagcc agccatgata aagagctcgc gcagagcgcc ctccaccggc attcttaagc 29460
acaccctcat aattccaaga tattctgctc ctggttcacc tgcagcagat tgacaagcgg 29520
aatatcaaaa tctctgccgc gatccctgag ctcctccctc agcaataact gtaagtactc 29580
tttcatatcc tctccgaaat ttttagccat aggaccacca ggaataagat tagggcaagc 29640

cacagtacag ataaaccgaa gtcctcccca gtgagcattg ccaaatgcaa gactgctata 29700
agcatgctgg ctagacccgg tgatatcttc cagataactg gacagaaaat cgcccaggca 29760
atttttaaga aaatcaacaa aagaaaaatc ctccaggtgg acgtttagag cctcgggaac 29820
aacgatgaag taaatgcaag cggtgcgttc cagcatggtt agttagctga tctgtagaaa 29880
aaacaaaaat gaacattaaa ccatgctagc ctggcgaaca ggtgggtaaa tcgttctctc 29940
cagcaccagg caggccacgg ggtctccggc gcgaccctcg taaaaattgt cgctatgatt 30000
gaaaaccatc acagagagac gttcccggtg gccggcgtga atgattcgac aagatgaata 30060

caccccggga acattggcgt ccgcgagtga aaaaaagcgc ccgaggaagc aataaggcac 30120
tacaatgctc agtctcaagt ccagcaaagc gatgccatgc ggatgaagca caaaattctc 30180
aggtgcgtac aaaatgtaat tactcccctc ctgcacaggc agcaaagccc ccgatccctc 30240
caggtacaca tacaaagcct cagcgtccat agcttaccga gcagcagcac acaacaggcg 30300
caagagtcag agaaaggctg agctctaacc tgtccacccg ctctctgctc aatatatagc 30360
ccagatctac actgacgtaa aggccaaagt ctaaaaatac ccgccaaata atcacacacg 30420
cccagcacac gcccagaaac cggtgacaca ctcaaaaaaa tacgcgcact tcctcaaacg 30480

cccaaaactg ccgtcatttc cgggttccca cgctacgtca tcaaaacacg actttcaaat 30540

tccgtcgacc gttaaaaacg tcacccgccc cgcccctaac ggtcgcccgt ctctcagcca 30600

atcagcgccc cgcatcccca aattcaaaca cctcatttgc atattaacgc gcacaaaaag 30660

tttgaggtat attattgatg atgg 30684

<210> 14

<211> 8602

<212> DNA

<213> Artificial Sequence

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

<400> 14

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140
acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260
gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500
gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560
ttaaagtccg ccatatgagg ccaccatgca gatcttcgtg aagaccctga ccggcaagac 7620

catcacccta gaggtggagc ccagtgacac catcgagaac gtgaaggcca agatccagga 7680
taaagagggc atccccccctg accagcagag gctgatcttt gccggcaagc agctggaaga 7740
tggccgcacc ctctctgatt acaacatcca gaaggagtca accctgcacc tggtccttcg 7800
cctgagaggt ggcgctgctt acagtataat caactttgaa aaactggctg cttacggcat 7860
cctgggcttt gtgtttacac tggctgccta cctgctgttt ggctatcctg tgtacgtggc 7920
cgcttatgga ctgtgtaccc tggtggccat gctggctgct tacaatctgg tgcctatggt 7980
ggccacagtg gccgcctatt gtcttggcgg actgctgaca atggtggcag cctacagccc 8040

gagctatgcg tatcatcagt ttgcagccta cggcccagga ccaggcgcta aatttgtggc 8100
tgcctggaca ctgaaagccg ccgctggacc aggtcctgga cagtacatca aggccaacag 8160
caagttcatc ggcatcaccg aactcggccc aggaccaggc tatccctacg atgtgcctga 8220
ttacgcctga tagtgatgat tcgaacggcc gtatcacgcc caaacattta cagccgcggt 8280
gtcaaaaacc gcgtggacgt ggttaacatc cctgctggga ggatcagccg taattattat 8340
aattggcttg gtgctggcta ctattgtggc catgtacgtg ctgaccaacc agaaacataa 8400
ttgaatacag cagcaattgg caagctgctt acatagaact cgcggcgatt ggcatgccgc 8460

cttaaaattt ttattttatt ttttcttttc ttttccgaat cggattttgt ttttaatatt 8520

tcaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 8580

aaaaaaaaaa aaaaaaaaaa aa 8602

<210> 15

<211> 9595

<212> DNA

<213> Artificial Sequence

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

<400> 15

atgggcggcg catgagagaa gcccagacca ttacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900
tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960
aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020
tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080
cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140
acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200
aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320
aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380
gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440
tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500
gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560
ttaaagtccg ccatatgaga tggaagatgc caaaaacatt aagaagggcc cagcgccatt 7620
ctacccactc gaagacggga ccgccggcga gcagctgcac aaagccatga agcgctacgc 7680

cctggtgccc ggcaccatcg cctttaccga cgcacatatc gaggtggaca ttacctacgc 7740
cgagtacttc gagatgagcg ttcggctggc agaagctatg aagcgctatg ggctgaatac 7800
aaaccatcgg atcgtggtgt gcagcgagaa tagcttgcag ttcttcatgc ccgtgttggg 7860
tgccctgttc atcggtgtgg ctgtggcccc agctaacgac atctacaacg agcgcgagct 7920
gctgaacagc atgggcatca gccagcccac cgtcgtattc gtgagcaaga aagggctgca 7980
aaagatcctc aacgtgcaaa agaagctacc gatcatacaa aagatcatca tcatggatag 8040
caagaccgac taccagggct tccaaagcat gtacaccttc gtgacttccc atttgccacc 8100

cggcttcaac gagtacgact tcgtgcccga gagcttcgac cgggacaaaa ccatcgccct 8160
gatcatgaac agtagtggca gtaccggatt gcccaagggc gtagccctac cgcaccgcac 8220
cgcttgtgtc cgattcagtc atgcccgcga ccccatcttc ggcaaccaga tcatccccga 8280
caccgctatc ctcagcgtgg tgccatttca ccacggcttc ggcatgttca ccacgctggg 8340
ctacttgatc tgcggctttc gggtcgtgct catgtaccgc ttcgaggagg agctattctt 8400
gcgcagcttg caagactata agattcaatc tgccctgctg gtgcccacac tatttagctt 8460
cttcgctaag agcactctca tcgacaagta cgacctaagc aacttgcacg agatcgccag 8520

cggcggggcg ccgctcagca aggaggtagg tgaggccgtg gccaaacgct tccacctacc 8580

aggcatccgc cagggctacg gcctgacaga aacaaccagc gccattctga tcacccccga 8640

aggggacgac aagcctggcg cagtaggcaa ggtggtgccc ttcttcgagg ctaaggtggt 8700

ggacttggac accggtaaga cactgggtgt gaaccagcgc ggcgagctgt gcgtccgtgg 8760

ccccatgatc atgagcggct acgttaacaa ccccgaggct acaaacgctc tcatcgacaa 8820

ggacggctgg ctgcacagcg gcgacatcgc ctactgggac gaggacgagc acttcttcat 8880

cgtggaccgg ctgaagagcc tgatcaaata caagggctac caggtagccc cagccgaact 8940

ggagagcatc ctgctgcaac accccaacat cttcgacgcc ggggtcgccg gcctgcccga 9000

cgacgatgcc ggcgagctgc ccgccgcagt cgtcgtgctg gaacacggta aaaccatgac 9060

cgagaaggag atcgtggact atgtggccag ccaggttaca accgccaaga agctgcgcgg 9120

tggtgttgtg ttcgtggacg aggtgcctaa aggactgacc ggcaagttgg acgcccgcaa 9180

gatccgcgag attctcatta aggccaagaa gggcggcaag atcgccgtgt aattcgaacg 9240

gccgtatcac gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgga cgtggttaac 9300

atccctgctg ggaggatcag ccgtaattat tataattggc ttggtgctgg ctactattgt 9360

ggccatgtac gtgctgacca accagaaaca taattgaata cagcagcaat tggcaagctg 9420

cttacataga actcgcggcg attggcatgc cgccttaaaa tttttatttt atttttttctt 9480

ttcttttccg aatcggattt tgtttttaat atttcaaaaa aaaaaaaaaa aaaaaaaaaa 9540

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa 9595

<210> 16

<211> 139

<212> PRT

<213> Artificial Sequence

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

<400> 16

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

1 5 10 15

Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp Trp Tyr Gln Gln Lys

20 25 30

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

35 40 45

Ser Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe

50 55 60

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

65 70 75 80

Cys Gln Gln Tyr Tyr Ser Thr Pro Phe Thr Phe Gly Pro Gly Thr Lys

85 90 95

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

100 105 110

Pro Ser Asp Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu

115 120 125

Leu Asn Asn Phe Tyr Pro Arg Glu Ala Lys Val

130 135

<210> 17

<211> 167

<212> PRT

<213> Artificial Sequence

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

<400> 17

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

1 5 10 15

Gly Phe Thr Phe Ser Ser Tyr Gly Met His Trp Val Arg Gln Ala Pro

20 25 30

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

35 40 45

Lys Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp

50 55 60

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

65 70 75 80

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

85 90 95

Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln Gly Thr Thr Val

100 105 110

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

<210> 18
<211> 10
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 18
Gly Phe Thr Phe Ser Ser Tyr Gly Met His
1 5 10

<210> 19
<211> 15
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 19
Val Ile Trp Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
1 5 10 15

<210> 20
<211> 16
<212> PRT
<213> Artificial Sequence
<220><223
> Description of Artificial Sequence: Synthetic

peptide

<400> 20

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

1 5 10 15

<210> 21

<211> 11

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 21

Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp

1 5 10

<210> 22

<211> 7

<212> PRT

<213> Artificial Sequence

<

220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 22

Ala Ala Ser Ser Leu Gln Ser

1 5

<210> 23

<211> 9

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 23

Gln Gln Tyr Tyr Ser Thr Pro Phe Thr

1 5

<210> 24

<211> 108

<212> PRT

<213> Artificial Sequence

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

<

400> 24

Glu Ile Val Leu Thr Gln Ser Pro Gly Thr Leu Ser Leu Ser Pro Gly
1 5 10 15

Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Arg Val Ser Ser Ser
20 25 30

Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu Leu
35 40 45

Ile Tyr Asp Ala Ser Ser Arg Ala Thr Gly Ile Pro Asp Arg Phe Ser
50 55 60

Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Arg Leu Glu
65 70 75 80

Pro Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Tyr Gly Ser Leu Pro
85 90 95

Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys
100 105

<210> 25

<211> 121

<212> PRT

<213> Artificial Sequence

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

<400

> 25

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Arg Tyr
20 25 30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

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

50 55 60

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

<210> 26
<211> 5
<212> PRT
<213
> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 26
Arg Tyr Trp Met Ser
1 5
<210> 27
<211> 17
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 27
Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val Lys
1 5 10 15
Gly

<210> 28
<211> 12
<212> PRT
<213> Artificial Sequence

<220

><223> Description of Artificial Sequence: Synthetic

peptide

<400> 28

Glu Gly Gly Trp Phe Gly Glu Leu Ala Phe Asp Tyr

1 5 10

<210> 29

<211> 12

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 29

Arg Ala Ser Gln Arg Val Ser Ser Ser Tyr Leu Ala

1 5 10

<210> 30

<211> 7

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 30

Asp Ala Ser Ser Arg Ala Thr

1 5

<210> 31

<211> 9

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 31

Gln Gln Tyr Gly Ser Leu Pro Trp Thr

1 5

<210> 32

<211> 2019

<212> DNA

<213> Artificial Sequence

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

<400> 32

gcccgggcat ttaaatgcga tcgcatcgat tacgactcta gaatagtcta gtccgcaggc 60

caccatgcag atcttcgtga agaccctgac cggcaagacc atcaccctag aggtggagcc 120

cagtgacacc atcgagaacg tgaaggccaa gatccaggat aaagagggca tccccccctga 180

ccagcagagg ctgatctttg ccggcaagca gctggaagat ggccgcaccc tctctgatta 240

caacatccag aaggagtcaa ccctgcacct ggtccttcgc ctgagaggtg ccatgtttca 300

ggcgctgagc gaaggctgca ccccgtatga tattaaccag atgctgaacg tgctgggcga 360

tcatcaggtc tcaggccttg agcagcttga gagtataatc aactttgaaa aactgactga 420

atggaccagt tctaatgtta tgcctatcct gtctcctctg acaaagggca tcctgggctt 480

cgtgtttacc ctgaccgtgc cttctgagag aggacttagc tgcattagcg aagcggatgc 540

gaccaccccg gaaagcgcga acctgggcga agaaattctg agccagctgt atctttggcc 600

aagggtgacc taccattccc ctagttatgc ttaccaccaa tttgaaagac gagccaaata 660

taaaagacac ttccccggct ttggccagag cctgctgttt ggctaccctg tgtacgtgtt 720

cggcgattgc gtgcagggcg attgggatgc gattcgcttt cgctattgcg cgccgccggg 780

ctatgcgctg ctgcgctgca acgataccaa ctatagcgct ctgctggctg tgggggccct 840

agaaggaccc aggaatcagg actggcttgg tgtcccaaga caacttgtaa ctcggatgca 900

ggctattcag aatgccggcc tgtgtaccct ggtggccatg ctggaagaga caatcttctg 960

gctgcaagcg tttctgatgg cgctgaccga tagcggcccg aaaaccaaca ttattgtgga 1020

tagccagtat gtgatgggca ttagcaaacc gagctttcag gaatttgtgg attgggaaaa 1080

cgtgagcccg gaactgaaca gcaccgatca gccgttttgg caagccggaa tcctggccag 1140

aaatctggtg cctatggtgg ccacagtgca gggccagaac ctgaagtacc agggtcagtc 1200

actagtcatc tctgcttcta tcattgtctt caacctgctg gaactggaag gtgattatcg 1260

agatgatggc aacgtgtggg tgcatacccc gctgagcccg cgcaccctga acgcgtgggt 1320

gaaagcggtg gaagaaaaaa aaggtattcc agttcaccta gagctggcca gtatgaccaa 1380

catggagctc atgagcagta ttgtgcatca gcaggtcaga acatacggcc ccgtgttcat 1440

gtgtctcggc ggactgctta caatggtggc tggtgctgtg tggctgacag tgcgagtgct 1500

cgagctgttc cgggccgcgc agctggccaa cgacgtggtc ctccagatca tggagctttg 1560

tggtgcagcg tttcgccagg tgtgccatac caccgtgccg tggccgaacg cgagcctgac 1620

cccgaaatgg aacaacgaaa ccacccagcc ccagatcgcc aactgcagcg tgtatgactt 1680

ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca 1740

tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg 1800

acctggcgct aaatttgtgg ccgcctggac actgaaagcc gctgctggtc ctggacctgg 1860

ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg 1920

ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac 1980

gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgg 2019

<210> 33

<211> 619

<212> PRT

<213> Artificial Sequence

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

<400> 33

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65 70 75 80

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
85 90 95

Leu Gly Asp His Gln Val Ser Gly Leu Glu Gln Leu Glu Ser Ile Ile
100 105 110

Asn Phe Glu Lys Leu Thr Glu Trp Thr Ser Ser Asn Val Met Pro Ile

115 120 125

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

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205
Ser Leu Leu Phe Gly Tyr Pro Val Tyr Val Phe Gly Asp Cys Val Gln
210 215 220
Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr
225 230 235 240
Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val
245 250 255

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

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

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala
370 375 380

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

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu
450 455 460
Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480
Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495
Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro
500 505 510

Trp Pro Asn Ala Ser Leu Thr Pro Lys Trp Asn Asn Glu Thr Thr Gln
515 520 525
Pro Gln Ile Ala Asn Cys Ser Val Tyr Asp Phe Phe Val Trp Leu His
530 535 540
Tyr Tyr Ser Val Arg Asp Thr Leu Trp Pro Arg Val Thr Tyr His Met
545 550 555 560
Asn Lys Tyr Ala Tyr His Met Leu Glu Arg Arg Ala Lys Tyr Lys Arg
565 570 575

Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590
Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605
Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly

610 615

<210> 34

<211> 1638

<212> DNA

<213> Artificial Sequence

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

<400> 34

atggccggga tgttccaggc actgtccgaa ggctgcacac cctatgatat taaccagatg 60

ctgaatgtcc tgggagacca ccaggtctct ggcctggagc agctggagag catcatcaac 120

ttcgagaagc tgaccgagtg gacaagctcc aatgtgatgc ctatcctgtc cccactgacc 180

aagggcatcc tgggcttcgt gtttaccctg acagtgcctt ctgagcgggg cctgtcttgc 240

atcagcgagg cagacgcaac cacaccagag tccgccaatc tgggcgagga gatcctgtct 300

cagctgtacc tgtggccccg ggtgacatat cactcccctt cttacgccta tcaccagttc 360

gagcggagag ccaagtacaa gagacacttc ccaggctttg gccagtctct gctgttcggc 420

taccccgtgt acgtgttcgg cgattgcgtg cagggcgact gggatgccat ccggtttaga 480

tactgcgcac cacctggata tgcactgctg aggtgtaacg acaccaatta ttccgccctg 540

ctggcagtgg gcgccctgga gggccctcgc aatcaggatt ggctgggcgt gccaaggcag 600

ctggtgacac gcatgcaggc catccagaac gcaggcctgt gcaccctggt ggcaatgctg 660

gaggagacaa tcttctggct gcaggccttt ctgatggccc tgaccgacag cggccccaag 720

acaaacatca tcgtggattc ccagtacgtg atgggcatct ccaagccttc tttccaggag 780

tttgtggact gggagaacgt gagcccagag ctgaattcca ccgatcagcc attctggcag 840

gcaggaatcc tggcaaggaa cctggtgcct atggtggcca cagtgcaggg ccagaatctg 900

aagtaccagg gccagagcct ggtcatcagc gcctccatca tcgtgtttaa cctgctggag 960

ctggagggcg actatcggga cgatggcaac gtgtgggtgc acaccccact gagccccaga 1020

acactgaacg cctgggtgaa ggccgtggag gagaagaagg gcatcccagt gcacctggag 1080

ctggcctcca tgaccaatat ggagctgatg tctagcatcg tgcaccagca ggtgaggaca 1140

tacggacccg tgttcatgtg cctgggaggc ctgctgacca tggtggcagg agccgtgtgg 1200

ctgacagtgc gggtgctgga gctgttcaga gccgcccagc tggccaacga tgtggtgctg 1260

cagatcatgg agctgtgcgg agcagccttt cgccaggtgt gccacaccac agtgccatgg 1320

cccaatgcct ccctgacccc caagtggaac aatgagacaa cacagcctca gatcgccaac 1380

tgtagcgtgt acgacttctt cgtgtggctg cactactata gcgtgaggga taccctgtgg 1440

ccccgcgtga cataccacat gaataagtac gcctatcaca tgctggagag gcgcgccaag 1500

tataagagag gccctggccc aggcgcaaag tttgtggcag catggaccct gaaggccgcc 1560

gccggccccg gccccggcca gtatatcaag gctaacagta agttcattgg aatcacagag 1620

ctgggacccg gacctgga 1638

<210> 35

<211> 546

<212> PRT

<213> Artificial Sequence

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

<400> 35

Met Ala Gly Met Phe Gln Ala Leu Ser Glu Gly Cys Thr Pro Tyr Asp
1 5 10 15

Ile Asn Gln Met Leu Asn Val Leu Gly Asp His Gln Val Ser Gly Leu
20 25 30

Glu Gln Leu Glu Ser Ile Ile Asn Phe Glu Lys Leu Thr Glu Trp Thr
35 40 45

Ser Ser Asn Val Met Pro Ile Leu Ser Pro Leu Thr Lys Gly Ile Leu
50 55 60

Gly Phe Val Phe Thr Leu Thr Val Pro Ser Glu Arg Gly Leu Ser Cys
65 70 75 80

Ile Ser Glu Ala Asp Ala Thr Thr Pro Glu Ser Ala Asn Leu Gly Glu
85 90 95

Glu Ile Leu Ser Gln Leu Tyr Leu Trp Pro Arg Val Thr Tyr His Ser
100 105 110

Pro Ser Tyr Ala Tyr His Gln Phe Glu Arg Arg Ala Lys Tyr Lys Arg
115 120 125

His Phe Pro Gly Phe Gly Gln Ser Leu Leu Phe Gly Tyr Pro Val Tyr
130 135 140

Val Phe Gly Asp Cys Val Gln Gly Asp Trp Asp Ala Ile Arg Phe Arg
145 150 155 160

Tyr Cys Ala Pro Pro Gly Tyr Ala Leu Leu Arg Cys Asn Asp Thr Asn
165 170 175

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

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

Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe Asn Leu Leu Glu
305 310 315 320
Leu Glu Gly Asp Tyr Arg Asp Asp Gly Asn Val Trp Val His Thr Pro
325 330 335
Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Ala Val Glu Glu Lys
340 345 350
Lys Gly Ile Pro Val His Leu Glu Leu Ala Ser Met Thr Asn Met Glu
355 360 365

Leu Met Ser Ser Ile Val His Gln Gln Val Arg Thr Tyr Gly Pro Val
370 375 380
Phe Met Cys Leu Gly Gly Leu Leu Thr Met Val Ala Gly Ala Val Trp
385 390 395 400
Leu Thr Val Arg Val Leu Glu Leu Phe Arg Ala Ala Gln Leu Ala Asn

405 410 415

Asp Val Val Leu Gln Ile Met Glu Leu Cys Gly Ala Ala Phe Arg Gln

420 425 430

Val Cys His Thr Thr Val Pro Trp Pro Asn Ala Ser Leu Thr Pro Lys

435 440 445

Trp Asn Asn Glu Thr Thr Gln Pro Gln Ile Ala Asn Cys Ser Val Tyr

450 455 460

Asp Phe Phe Val Trp Leu His Tyr Tyr Ser Val Arg Asp Thr Leu Trp

465 470 475 480

Pro Arg Val Thr Tyr His Met Asn Lys Tyr Ala Tyr His Met Leu Glu

485 490 495

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

500 505 510

Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Pro Gly Pro Gly Gln Tyr

515 520 525

Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu Gly Pro Gly

530 535 540

Pro Gly

545

<210> 36

<211> 2019

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 36

gcccgggcat ttaaatgcga tcgcatcgat tacgactcta gaatagtcta gtccgcaggc 60

caccatgcag atcttcgtga agaccctgac cggcaagacc atcaccctag aggtggagcc 120

cagtgacacc atcgagaacg tgaaggccaa gatccaggat aaagagggca tccccctga 180

ccagcagagg ctgatctttg ccggcaagca gctggaagat ggccgcaccc tctctgatta 240

caacatccag aaggagtcaa ccctgcacct ggtccttcgc ctgagaggtg ccatgtttca 300

ggcgctgagc gaaggctgca ccccgtatga tattaaccag atgctgaacg tgctgggcga 360

tcatcagttt aagcacatca aagcctttga ccggacattt gctaacaacc caggtcccat 420

ggttgtgttt gccacacctg ggcctatcct gtctcctctg acaaagggca tcctgggctt 480
cgtgtttacc ctgaccgtgc cttctgagag aggacttagc tgcattagcg aagcggatgc 540
gaccaccccg gaaagcgcga acctgggcga agaaattctg agccagctgt atctttggcc 600
aagggtgacc taccattccc ctagttatgc ttaccaccaa tttgaaagac gagccaaata 660
taaaagacac ttccccggct ttggccagag cctgctgttt ggctaccctg tgtacgtgtt 720
cggcgattgc gtgcagggcg attgggatgc gattcgcttt cgctattgcg cgccgccggg 780
ctatgcgctg ctgcgctgca acgataccaa ctatagcgct ctgctggctg tgggggccct 840

agaaggaccc aggaatcagg actggcttgg tgtcccaaga caacttgtaa ctcggatgca 900
ggctattcag aatgccggcc tgtgtaccct ggtggccatg ctggaagaga caatcttctg 960
gctgcaagcg tttctgatgg cgctgaccga tagcggcccg aaaaccaaca ttattgtgga 1020
tagccagtat gtgatgggca ttagcaaacc gagctttcag gaatttgtgg attgggaaaa 1080
cgtgagcccg gaactgaaca gcaccgatca gccgttttgg caagccggaa tcctggccag 1140
aaatctggtg cctatggtgg ccacagtgca gggccagaac ctgaagtacc agggtcagtc 1200
actagtcatc tctgcttcta tcattgtctt caacctgctg gaactggaag gtgattatcg 1260

agatgatggc aacgtgtggg tgcatacccc gctgagcccg cgcaccctga acgcgtgggt 1320
gaaagcggtg gaagaaaaaa aaggtattcc agttcaccta gagctggcca gtatgaccaa 1380
catggagctc atgagcagta ttgtgcatca gcaggtcaga acatacggcc ccgtgttcat 1440
gtgtctcggc ggactgctta caatggtggc tggtgctgtg tggctgacag tgcgagtgct 1500
cgagctgttc cgggccgcgc agctggccaa cgacgtggtc ctccagatca tggagctttg 1560
tggtgcagcg tttcgccagg tgtgccatac caccgtgccg tggccgaacg cgagcctgac 1620
cccgaaatgg aacaacgaaa ccacccagcc ccagatcgcc aactgcagcg tgtatgactt 1680

ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca 1740
tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg 1800
acctggcgct aaatttgtgg ccgcctggac actgaaagcc gctgctggtc ctggacctgg 1860
ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg 1920
ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac 1980
gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgg 2019

<210> 37

<211> 619

<212> PRT

<213> Artificial Sequence

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

<400> 37

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65 70 75 80

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
85 90 95

Leu Gly Asp His Gln Phe Lys His Ile Lys Ala Phe Asp Arg Thr Phe
100 105 110

Ala Asn Asn Pro Gly Pro Met Val Val Phe Ala Thr Pro Gly Pro Ile
115 120 125

Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr
130 135 140

Val Pro Ser Glu Arg Gly Leu Ser Cys Ile Ser Glu Ala Asp Ala Thr
145 150 155 160

Thr Pro Glu Ser Ala Asn Leu Gly Glu Glu Ile Leu Ser Gln Leu Tyr
165 170 175

Leu Trp Pro Arg Val Thr Tyr His Ser Pro Ser Tyr Ala Tyr His Gln
180 185 190

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205

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

210 215 220

Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr

225 230 235 240

Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val

245 250 255

Gly Ala Leu Glu Gly Pro Arg Asn Gln Asp Trp Leu Gly Val Pro Arg

260 265 270

Gln Leu Val Thr Arg Met Gln Ala Ile Gln Asn Ala Gly Leu Cys Thr

275 280 285

Leu Val Ala Met Leu Glu Glu Thr Ile Phe Trp Leu Gln Ala Phe Leu

290 295 300

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

305 310 315 320

Gln Tyr Val Met Gly Ile Ser Lys Pro Ser Phe Gln Glu Phe Val Asp

325 330 335

Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Trp

340 345 350

Gln Ala Gly Ile Leu Ala Arg Asn Leu Val Pro Met Val Ala Thr Val

355 360 365

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala

370 375 380

Ser Ile Ile Val Phe Asn Leu Leu Glu Leu Glu Gly Asp Tyr Arg Asp

385 390 395 400

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

405 410 415

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

420 425 430

Glu Leu Ala Ser Met Thr Asn Met Glu Leu Met Ser Ser Ile Val His

435 440 445

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu

450 455 460

Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480
Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495
Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro

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

565 570 575
Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590
Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605
Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly
610 615

<210> 38
<211> 228
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 38
atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60
gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120
cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180
atccagaagg agtcaaccct gcacctggtc cttcgcctga gaggtggc 228
<210> 39

<211> 228

<212> DNA

<213> Artificial Sequence

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

<400> 39

atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60

gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120

cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180

atccagaagg agtcaaccct gcacctggtc cttcgcctga gaggtgcc 228

<210> 40

<211> 78

<212> DNA

<213> Homo sapiens

<400> 40

atggccgtca tggcgccccg aaccctcgtc ctgctactct cgggggctct ggccctgacc 60

cagacctggg cgggctct 78

<210> 41

<211> 201

<212> DNA

<213> Homo sapiens

<400> 41

ccgtcttccc agcccaccat ccccatcgtg ggcatcattg ctggcctggt tctctttgga 60

gctgtgatca ctggagctgt ggtcgctgct gtgatgtgga ggaggaagag ctcagataga 120

aaaggaggga gctactctca ggctgcaagc agtgacagtg cccagggctc tgatgtgtct 180

ctcacagctt gtaaagtgtg a 201

<210> 42

<211> 60

<212> DNA

<213> Artificial Sequence

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

<400> 42

atggagaccg atacactgct gctgtgggtg ctgctcctgt gggtgccagg aagcacaggc 60

<210> 43

<211> 3178

<212> DNA

<213> Homo sapiens

<400> 43

ggcaccgatt cggggcctgc ccggacttcg ccgcacgctg cagaacctcg cccagcgccc 60

accatgcccc ggcagctcag cgcggcggcc gcgctcttcg cgtccctggc cgtaattttg 120

cacgatggca gtcaaatgag agcaaaagca tttccagaaa ccagagatta ttctcaacct 180

actgcagcag caacagtaca ggacataaaa aaacctgtcc agcaaccagc taagcaagca 240

cctcaccaaa ctttagcagc aagattcatg gatggtcata tcacctttca aacagcggcc 300

acagtaaaaa ttccaacaac taccccagca actacaaaaa acactgcaac caccagccca 360

attacctaca ccctggtcac aacccaggcc acacccaaca actcacacac agctcctcca 420

gttactgaag ttacagtcgg ccctagctta gccccttatt cactgccacc caccatcacc 480

ccaccagctc atacagctgg aaccagttca tcaaccgtca gccacacaac tgggaacacc 540

actcaaccca gtaaccagac cacccttcca gcaactttat cgatagcact gcacaaaagc 600

acaaccggtc agaagcctga tcaacccacc catgccccag gaacaacggc agctgcccac 660

aataccaccc gcacagctgc acctgcctcc acggttcctg ggcccaccct tgcacctcag 720

ccatcgtcag tcaagactgg aatttatcag gttctaaacg gaagcagact ctgtataaaa 780

gcagagatgg ggatacagct gattgttcaa gacaaggagt cggtttttc acctcggaga 840

tacttcaaca tcgaccccaa cgcaacgcaa gcctctggga actgtggcac ccgaaaatcc 900

aaccttctgt tgaattttca gggcggattt gtgaatctca catttaccaa ggatgaagaa 960

tcatattata tcagtgaagt gggagcctat ttgaccgtct cagatccaga gacagtttac 1020

caaggaatca aacatgcggt ggtgatgttc cagacagcag tcgggcattc cttcaagtgc 1080

gtgagtgaac agagcctcca gttgtcagcc cacctgcagg tgaaaacaac cgatgtccaa 1140

cttcaagcct ttgattttga agatgaccac tttggaaatg tggatgagtg ctcgtctgac 1200

tacacaattg tgcttcctgt gattggggcc atcgtggttg gtctctgcct tatgggtatg 1260

ggtgtctata aaatccgcct aaggtgtcaa tcatctggat accagagaat ctaattgttg 1320

cccgggggga atgaaaataa tggaatttag agaactcttt catcccttcc aggatggatg 1380

ttgggaaatt ccctcagagt gtgggtcctt caaacaatgt aaaccaccat cttctattca 1440

aatgaagtga gtcatgtgtg atttaagttc aggcagcaca tcaatttcta aatacttttt 1500

gtttatttta tgaaagatat agtgagctgt ttattttcta gtttccttta gaatatttta 1560

gccactcaaa gtcaacattt gagatatgtt gaattaacat aatatatgta aagtagaata 1620

agccttcaaa ttataaacca agggtcaatt gtaactaata ctactgtgtg tgcattgaag 1680

attttatttt acccttgatc ttaacaaagc ctttgctttg ttatcaaatg gactttcagt 1740

gcttttacta tctgtgtttt atggtttcat gtaacataca tattcctggt gtagcactta 1800

actccttttc cactttaaat ttgtttttgt tttttgagac ggagtttcac tcttgtcacc 1860

caggctggag tacagtggca cgatctcggc ttatggcaac ctccgcctcc cgggttcaag 1920

tgattctcct gcttcagctt cccgagtagc tgggattaca ggcacacact accacgcctg 1980

gctaattttt gtatttttat tatagacggg tttcaccatg ttggccagac tggtcttgaa 2040

ctcttgacct caggtgatcc acccacctca gcctcccaaa gtgctgggat tacaggcatg 2100

agccattgcg cccggcctta aatgtttttt ttaatcatca aaaagaacaa catatctcag 2160

gttgtctaag tgtttttatg taaaaccaac aaaaagaaca aatcagctta tatttttttat 2220

cttgatgact cctgctccag aattgctaga ctaagaatta ggtggctaca gatggtagaa 2280

ctaaacaata agcaagagac aataataatg gcccttaatt attaacaaag tgccagagtc 2340

taggctaagc actttatcta tatctcattt cattctcaca acttataagt gaatgagtaa 2400

actgagactt aagggaactg aatcacttaa atgtcacctg gctaactgat ggcagagcca 2460

gagcttgaat tcatgttggt ctgacatcaa ggtctttggt cttctcccta caccaagtta 2520

cctacaagaa caatgacacc acactctgcc tgaaggctca cacctcatac cagcatacgc 2580

tcaccttaca gggaaatggg tttatccagg atcatgagac attagggtag atgaaaggag 2640

agctttgcag ataacaaaat agcctatcct taataaatcc tccactctct ggaaggagac 2700

tgaggggctt tgtaaaacat tagtcagttg ctcattttta tgggattgct tagctgggct 2760

gtaaagatga aggcatcaaa taaactcaaa gtatttttaa atttttttga taatagagaa 2820

acttcgctaa ccaactgttc tttcttgagt gtatagcccc atcttgtggt aacttgctgc 2880

ttctgcactt catatccata tttcctattg ttcactttat tctgtagagc agcctgccaa 2940

gaattttatt tctgctgttt tttttgctgc taaagaaagg aactaagtca ggatgttaac 3000

agaaaagtcc acataaccct agaattctta gtcaaggaat aattcaagtc agcctagaga 3060

ccatgttgac tttcctcatg tgtttcctta tgactcagta agttggcaag gtcctgactt 3120

tagtcttaat aaaacattga attgtagtaa aggttttgc aataaaaact tactttgg 3178

<210> 44

<211> 1858

<212> DNA

<213> Mus sp.

<400> 44

attccggagg tgaaaaacaa tggcacaacg tgtataatgg ccagcttctc tgcctccttt 60

ctgaccacct acgagactgc gaatggttct cagatcgtga acatttccct gccagcctct 120

gcagaagtac tgaaaaatgg cagttcttgt ggtaaagaaa atgtttctga ccccagcctc 180

acaattactt ttggaagagg atatttactg acactcaact tcacaaaaaa tacaacacgt 240

tacagtgtcc agcatatgta ttttacatat aacttgtcag atacagaaca ttttcccaat 300

gccatcagca aagagatcta caccatggat tccacaactg acatcaaggc agacatcaac 360

aaagcatacc ggtgtgtcag tgatatccgg gtctacatga agaatgtgac cgttgtgctc 420

cgggatgcca ctatccaggc ctacctgtcg agtggcaact tcagcaagga agagacacac 480

tgcacacagg atggaccttc cccaaccact gggccaccca gcccctcacc accacttgtg 540

cccacaaacc ccactgtatc caagtacaat gttactggta acaacggaac ctgcctgctg 600

gcctctatgg cactgcaact gaatatcacc tacctgaaaa aggacaacaa gacggtgacc 660

agagcgttca acatcagccc aaatgacaca tctagtggga gttgcggtat caacttggtg 720

accctgaaag tggagaacaa gaacagagcc ctggaattgc agtttgggat gaatgccagc 780

tctagcctgt ttttcttgca aggagtgcgc ttgaatatga ctcttcctga tgccctagtg 840

cccacattca gcatctccaa ccattcactg aaagctcttc aggccactgt gggaaactca 900

tacaagtgca acactgagga acacatcttt gtcagcaaga tgctctccct caatgtcttc 960

agtgtgcagg tccaggcttt caaggtggac agtgacaggt ttgggtctgt ggaagagtgt 1020

gttcaggatg gtaacaacat gttgatcccc attgctgtgg gcggtgccct ggcagggctg 1080

atcctcatcg tcctcattgc ctacctcatt ggcaggaaga ggagtcacgc cggctatcag 1140

accatctagc ctggtgggca ggtgcaccag agatgcacag gggcctgttc tcacatcccc 1200

aagcttagat aggtgtggaa gggaggcaca ctttctggca aactgtttta aaatctgctt 1260

tatcaaatgt gaagttcatc ttgcaacatt tactatgcac aaaggaataa ctattgaaat 1320

gacggtgtta attttgctaa ctgggttaaa tattgatgag aaggctccac tgatttgact 1380

tttaagactt ggtgtttggt tcttcattct tttactcaga tttaagccta tcaaagggat 1440

actctggtcc agaccttggc ctggcaaggg tggctgatgg ttaggctgca cacacttaag 1500

aagcaacggg agcagggaag gcttgcacac aggcacgcac agggtcaacc tctggacact 1560

tggcttgggc tacctggcct tgggggggct gaactctggc atctggctgg gtacacaccc 1620

ccccaatttc tgtgctctgc cacccgtgag ctgccacttt cctaaataga aaatggcatt 1680

attttattt actttttgt aaagtgattt ccagtcttgt gttggcgttc agggtggccc 1740

tgtctctgca ctgtgtacaa taatagattc acactgctga cgtgtcttgc agcgtaggtg 1800

ggttgtacac tgggcatcag ctcacgtaat gcattgcctg taacgatgct aataaaaa 1858

<210> 45

<211> 2339

<212> DNA

<213> Homo sapiens

<400> 45

ggcccaaccg ccgcccgcgc ccccgctctc cgcaccgtac ccggccgcct cgcgccatgg 60

cggccccccgg cagcgcccgg cgacccctgc tgctgctact gctgttgctg ctgctcggcc 120

tcatgcattg tgcgtcagca gcaatgttta tggtgaaaaa tggcaacggg accgcgtgca 180

taatggccaa cttctctgct gccttctcag tgaactacga caccaagagt ggccctaaga 240

acatgacctt tgacctgcca tcagatgcca cagtggtgct caaccgcagc tcctgtggaa 300

aagagaacac ttctgacccc agtctcgtga ttgcttttgg aagaggacat acactcactc 360

tcaatttcac gagaaatgca acacgttaca gcgtccagct catgagtttt gtttataact 420

tgtcagacac acacctttc cccaatgcga gctccaaaga aatcaagact gtggaatcta 480

taactgacat cagggcagat atagataaaa aatacagatg tgttagtggc acccaggtcc 540

acatgaacaa cgtgaccgta acgctccatg atgccaccat ccaggcgtac ctttccaaca 600

gcagcttcag caggggagag acacgctgtg aacaagacag gccttcccca accacagcgc 660

cccctgcgcc acccagcccc tcgccctcac ccgtgcccaa gagcccctct gtggacaagt 720

acaacgtgag cggcaccaac gggacctgcc tgctggccag catggggctg cagctgaacc 780

tcacctatga gaggaaggac aacacgacgg tgacaaggct tctcaacatc aaccccaaca 840

agacctcggc cagcgggagc tgcggcgccc acctggtgac tctggagctg cacagcgagg 900

gcaccaccgt cctgctcttc cagttcggga tgaatgcaag ttctagccgg tttttcctac 960

aaggaatcca gttgaataca attcttcctg acgccagaga ccctgccttt aaagctgcca 1020

acggctccct gcgagcgctg caggccacag tcggcaattc ctacaagtgc aacgcggagg 1080

agcacgtccg tgtcacgaag gcgttttcag tcaatatatt caaagtgtgg gtccaggctt 1140

tcaaggtgga aggtggccag tttggctctg tggaggagtg tctgctggac gagaacagca 1200

tgctgatccc catcgctgtg ggtggtgccc tggcggggct ggtcctcatc gtcctcatcg 1260

cctacctcgt cggcaggaag aggagtcacg caggctacca gactatctag cctggtgcac 1320

gcaggcacag cagctgcagg ggcctctgtt cctttctctg ggcttagggt cctgtcgaag 1380

gggaggcaca ctttctggca aacgtttctc aaatctgctt catccaatgt gaagttcatc 1440

ttgcagcatt tactatgcac aacagagtaa ctatcgaaat gacggtgtta attttgctaa 1500

ctgggttaaa tattttgcta actggttaaa cattaatatt taccaaagta ggattttgag 1560

ggtgggggtg ctctctctga gggggtgggg gtgccgctgt ctctgagggg tgggggtgcc 1620

gctgtctctg aggggtgggg gtgccgctct ctctgagggg gtgggggtgc cgctttctct 1680

gaggggggtgg gggtgccgct ctctctgagg gggtgggggt gctgctctct ccgaggggtg 1740

gaatgccgct gtctctgagg ggtgggggtg ccgctctaaa ttggctccat atcatttgag 1800

tttagggttc tggtgtttgg tttcttcatt ctttactgca ctcagattta agccttacaa 1860

agggaaagcc tctggccgtc acacgtagga cgcatgaagg tcactcgtgg tgaggctgac 1920

atgctcacac attacaacag tagagaggga aaatcctaag acagaggaac tccagagatg 1980

agtgtctgga gcgcttcagt tcagctttaa aggccaggac gggccacacg tggctggcgg 2040

cctcgttcca gtggcggcac gtccttgggc gtctctaatg tctgcagctc aagggctggc 2100

actttttaa atataaaaat gggtgttatt tttatttttt tttgtaaagt gatttttggt 2160

cttctgttga cattcggggt gatcctgttc tgcgctgtgt acaatgtgag atcggtgcgt 2220

tctcctgatg ttttgccgtg gcttggggat tgtacacggg accagctcac gtaatgcatt 2280

gcctgtaaca atgtaataaa aagcctcttt cttttaaaaa aaaaaaaaaa aaaaaaaaa 2339

<210> 46

<211> 45

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 46

cagtacatca aggccaacag caagttcatc ggcatcaccg aactc 45

<210> 47

<211> 15

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 47

Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu

1 5 10 15

<210> 48

<211> 39

<212> DNA

<213> Artificial Sequence

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

<400> 48

gctaaatttg tggctgcctg gacactgaaa gccgccgct 39

<210> 49

<211> 13

<212> PRT

<213> Artificial Sequence

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

<400> 49

Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala

1 5 10

<210> 50

<211> 593

<212> DNA

<213> Woodchuck hepatitis virus

<400> 50

aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgct 60

ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcttcccgt 120

atggctttca ttttctcctc cttgtataaa tcctggttgc tgtctcttta tgaggagttg 180

tggcccgttg tcaggcaacg tggcgtggtg tgcactgtgt ttgctgacgc aacccccact 240

ggttggggca ttgccaccac ctgtcagctc ctttccggga ctttcgcttt cccctccct 300

attgccacgg cggaactcat cgccgcctgc cttgcccgct gctggacagg ggctcggctg 360

ttgggcactg acaattccgt ggtgttgtcg gggaagctga cgtcctttcc atggctgctc 420

gcctgtgttg ccacctggat tctgcgcggg acgtccttct gctacgtccc ttcggccctc 480

aatccagcgg accttccttc ccgcggcctg ctgccggctc tgcggcctct tccgcgtctt 540

cgccttcgcc ctcagacgag tcggatctcc ctttgggccg cctccccgcc tgt 593

<210> 51

<211> 589

<212> DNA

<213> Artificial Sequence

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

<400> 51

tctccccccc ccccctctcc ctcccccccc cctaacgtta ctggccgaag ccgcttggaa 60

taaggccggt gtgcgtttgt ctatatgtta ttttccacca tattgccgtc ttttggcaat 120

gtgagggccc ggaaacctgg ccctgtcttc ttgacgagca ttcctagggg tctttccct 180

ctcgccaaag gaatgcaagg tctgttgaat gtcgtgaagg aagcagttcc tctggaagct 240

tcttgaagac aaacaacgtc tgtagcgacc ctttgcaggc agcggaaccc cccacctggc 300

gacaggtgcc tctgcggcca aaagccacgt gtataagata cacctgcaaa ggcggcacaa 360

ccccagtgcc acgttgtgag ttggatagtt gtggaaagag tcaaatggct ctcctcaagc 420

gtattcaaca aggggctgaa ggatgcccag aaggtacccc attgtatggg atctgatctg 480

gggcctcggt gcacatgctt tacatgtgtt tagtcgaggt taaaaaaacg tctaggcccc 540

ccgaaccacg gggacgtggt tttcctttga aaaacacgat gataatatg 589

<210> 52

<211> 720

<212> DNA

<213> Artificial Sequence

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

<400> 52

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60

ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc 180

ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag 240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300

ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg 360

gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420

aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480

ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc 540

gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600

tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc 660

ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtag 720

<210> 53

<211> 1563

<212> DNA

<213> Artificial Sequence

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

<400> 53

atgctgctgc tgctgctgct gctgggcctg aggctacagc tctccctggg catcatccca 60

gttgaggagg agaacccgga cttctggaac cgcgaggcag ccgaggccct gggtgccgcc 120

aagaagctgc agcctgcaca gacagccgcc aagaacctca tcatcttcct gggcgatggg 180

atgggggtgt ctacggtgac agctgccagg atcctaaaag ggcagaagaa ggacaaactg 240

gggcctgaga taccctggc catggaccgc ttcccatatg tggctctgtc caagacatac 300

aatgtagaca aacatgtgcc agacagtgga gccacagcca cggcctacct gtgcggggtc 360

aagggcaact tccagaccat tggcttgagt gcagccgccc gctttaacca gtgcaacacg 420

acacgcggca acgaggtcat ctccgtgatg aatcgggcca agaaagcagg gaagtcagtg 480

ggagtggtaa ccaccacacg agtgcagcac gcctcgccag ccggcaccta cgcccacacg 540

gtgaaccgca actggtactc ggacgccgac gtgcctgcct cggcccgcca ggaggggtgc 600

caggacatcg ctacgcagct catctccaac atggacattg acgtgatcct aggtggaggc 660

cgaaagtaca tgtttcgcat gggaacccca gaccctgagt acccagatga ctacagccaa 720

ggtgggacca ggctggacgg gaagaatctg gtgcaggaat ggctggcgaa gcgccagggt 780

gcccggtatg tgtggaaccg cactgagctc atgcaggctt ccctggaccc gtctgtgacc 840

catctcatgg gtctctttga gcctggagac atgaaatacg agatccaccg agactccaca 900

ctggacccct ccctgatgga gatgacagag gctgccctgc gcctgctgag caggaacccc 960

cgcggcttct tcctcttcgt ggagggtggt cgcatcgacc atggtcatca tgaaagcagg 1020

gcttaccggg cactgactga gacgatcatg ttcgacgacg ccattgagag ggcgggccag 1080

ctcaccagcg aggaggacac gctgagcctc gtcactgccg accactccca cgtcttctcc 1140

ttcggaggct acccccctgcg agggagctcc atcttcgggc tggcccctgg caaggcccgg 1200

gacaggaagg cctacacggt cctcctatac ggaaacggtc caggctatgt gctcaaggac 1260

ggcgcccggc cggatgttac cgagagcgag agcgggagcc ccgagtatcg gcagcagtca 1320

gcagtgcccc tggacgaaga gacccacgca ggcgaggacg tggcggtgtt cgcgcgcggc 1380

ccgcaggcgc acctggttca cggcgtgcag gagcagacct tcatagcgca cgtcatggcc 1440

ttcgccgcct gcctggagcc ctacaccgcc tgcgacctgg cgcccccccgc cggcaccacc 1500

gacgccgcgc acccgggtta ctctagagtc ggggcggccg gccgcttcga gcagacatga 1560

taa 1563

<210> 54

<211> 1653

<212> DNA

<213> Artificial Sequence

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

<400> 54

atggaagatg ccaaaaacat taagaagggc ccagcgccat tctacccact cgaagacggg 60

accgccggcg agcagctgca caaagccatg aagcgctacg ccctggtgcc cggcaccatc 120

gcctttaccg acgcacatat cgaggtggac attacctacg ccgagtactt cgagatgagc 180

gttcggctgg cagaagctat gaagcgctat gggctgaata caaaccatcg gatcgtggtg 240

tgcagcgaga atagcttgca gttcttcatg cccgtgttgg gtgccctgtt catcggtgtg 300

gctgtggccc cagctaacga catctacaac gagcgcgagc tgctgaacag catgggcatc 360

agccagccca ccgtcgtatt cgtgagcaag aaagggctgc aaaagatcct caacgtgcaa 420

aagaagctac cgatcataca aaagatcatc atcatggata gcaagaccga ctaccagggc 480

ttccaaagca tgtacacctt cgtgacttcc catttgccac ccggcttcaa cgagtacgac 540

ttcgtgcccg agagcttcga ccgggacaaa accatcgccc tgatcatgaa cagtagtggc 600

agtaccggat tgcccaaggg cgtagcccta ccgcaccgca ccgcttgtgt ccgattcagt 660

catgcccgcg accccatctt cggcaaccag atcatccccg acaccgctat cctcagcgtg 720

gtgccatttc accacggctt cggcatgttc accacgctgg gctacttgat ctgcggcttt 780

cgggtcgtgc tcatgtaccg cttcgaggag gagctattct tgcgcagctt gcaagactat 840

aagattcaat ctgccctgct ggtgcccaca ctatttagct tcttcgctaa gagcactctc 900

atcgacaagt acgacctaag caacttgcac gagatcgcca gcggcggggc gccgctcagc 960

aaggaggtag gtgaggccgt ggccaaacgc ttccacctac caggcatccg ccagggctac 1020

ggcctgacag aaacaaccag cgccattctg atcacccccg aaggggacga caagcctggc 1080

gcagtaggca aggtggtgcc cttcttcgag gctaaggtgg tggacttgga caccggtaag 1140

acactgggtg tgaaccagcg cggcgagctg tgcgtccgtg gccccatgat catgagcggc 1200

tacgttaaca accccgaggc tacaaacgct ctcatcgaca aggacggctg gctgcacagc 1260

ggcgacatcg cctactggga cgaggacgag cacttcttca tcgtggaccg gctgaagagc 1320

ctgatcaaat acaagggcta ccaggtagcc ccagccgaac tggagagcat cctgctgcaa 1380

caccccaaca tcttcgacgc cggggtcgcc ggcctgcccg acgacgatgc cggcgagctg 1440

cccgccgcag tcgtcgtgct ggaacacggt aaaaccatga ccgagaagga gatcgtggac 1500

tatgtggcca gccaggttac aaccgccaag aagctgcgcg gtggtgttgt gttcgtggac 1560

gaggtgccta aaggactgac cggcaagttg gacgcccgca agatccgcga gattctcatt 1620

aaggccaaga agggcggcaa gatcgccgtg taa 1653

<210> 55

<211> 66

<212> DNA

<213> Foot-and-mouth disease virus

<400> 55

gtaaagcaaa cactgaactt tgaccttctc aagttggctg gagacgttga gtccaatcct 60

gggccc 66

<210> 56

<211> 5

<212> PRT

<213> Artificial Sequence

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

<400> 56

Gly Pro Gly Pro Gly

1 5