(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년07월23일
(11) 등록번호 10-1870056
(24) 등록일자 2018년06월15일

(51) 국제특허분류(Int. Cl.)
C12N 7/00 (2006.01) *C12N 15/113* (2010.01)
C12N 15/86 (2006.01) *C12N 9/10* (2006.01)
C12N 9/78 (2006.01)
(52) CPC특허분류
C12N 7/00 (2013.01)
C12N 15/113 (2013.01)
(21) 출원번호 10-2016-7016947(분할)
(22) 출원일자(국제) 2009년09월26일
심사청구일자 2016년06월24일
(85) 번역문제출일자 2016년06월24일
(65) 공개번호 10-2016-0079149
(43) 공개일자 2016년07월05일
(62) 원출원 특허 10-2011-7008744
원출원일자(국제) 2009년09월26일
심사청구일자 2014년07월03일
(86) 국제출원번호 PCT/US2009/058510
(87) 국제공개번호 WO 2010/045002
국제공개일자 2010년04월22일
(30) 우선권주장
61/100,666 2008년09월26일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
GenBank Accession Number 1YSB_A (2007.10.01.)
US05770428 A

(73) 특허권자
토카겐 인크.
미국, 캘리포니아 92109, 샌 디에고, 스위트 230, 번커 힐 스트리트 3030
(72) 발명자
그루베르, 해리, 이.
미국, 92067, 란초 산타 페, 우편번호 675272
졸리, 더글라스
미국, 캘리포니아 92024, 엔시니타스, 선셋 드라이브 382
(뒷면에 계속)
(74) 대리인
강명구, 이경민

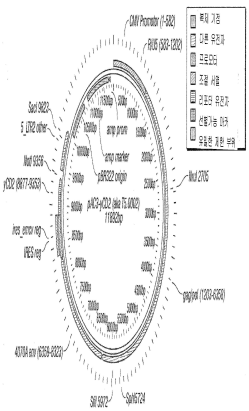
전체 청구항 수 : 총 14 항 심사관 : 최준호

(54) 발명의 명칭 **유전자 요법 벡터와 시토신 디아미나아제**

(57) 요 약

본 발명에서는 변형된 시토신 디아미나아제(CD)를 제시한다. 본 발명은 또한, 이런 변형된 CD를 발현하거나 포함하는 세포와 벡터, 그리고 질환과 질병의 치료에서 이런 변형된 CD를 이용하는 방법에 관계한다.

대 표 도



(52) CPC특허분류
C12N 15/86 (2013.01)
C12N 9/1077 (2013.01)
C12N 9/78 (2013.01)
C12Y 305/04001 (2013.01)
C12N 2310/141 (2013.01)
C12N 2740/13043 (2013.01)
C12N 2840/102 (2013.01)
C12N 2840/203 (2013.01)

(72) 발명자
페레즈, 오마르
미국, 캘리포니아 94107, 샌 프란시스코, 브란난 스트리트 #2씨 219
로그, 크리스토퍼, 알.
미국, 캘리포니아 91030, 사우스 파사데나, 웨인 애버뉴 1504

(30) 우선권주장
61/120,618 2008년12월08일 미국(US)
61/186,823 2009년06월13일 미국(US)

명 세 서

청구범위

청구항 1

다음을 포함하는 재조합 복제 적격성 감마레트로바이러스 (RCR):

a). 레트로바이러스 GAG 단백질;

b). 레트로바이러스 POL 단백질;

c). 레트로바이러스 외피;

d). 레트로바이러스 RNA 폴리뉴클레오티드, 여기서 상기 레트로바이러스 RNA 폴리뉴클레오티드는 레트로바이러스 RNA 폴리뉴클레오티드의 3' 말단에서 U3(번역되지 않은 3' 영역)-R 서열, 레트로바이러스 RNA 폴리뉴클레오티드의 5' 말단에서 R-U5(번역되지 않은 5') 서열, gag 핵산 도메인, pol 핵산 도메인, 그리고 env 핵산 도메인을 포함하고;

e). 시토신 디아미나아제 활성을 갖는 SEQ ID NO: 4의 열 안정화된 폴리펩티드를 인코드하고, 그리고 SEQ ID NO: 19에서 열거된 8877번 뉴클레오티드에서부터 9353번 뉴클레오티드까지의 서열을 포함하는 RNA 서열을 포함하는 이종성 핵산에 작동가능하게 연결된 내부 리보솜 결합 부위 (IRES)를 포함하는 카세트, 여기서 상기 카세트는 U3 영역의 5', 그리고 상기 레트로바이러스 외피를 인코딩하는 env 핵산 도메인의 3'에 배치되고; 그리고

f). 표적 세포 내에서 역전사, 포장 및 통합에 필요한 cis-작용 서열.

청구항 2

청구항 1에 있어서, 레트로바이러스 RNA 폴리뉴클레오티드는 뮤린 백혈병 바이러스 (MLV), 몰로니 뮤린 백혈병 바이러스 (MoMLV), 고양이 백혈병 바이러스 또는 Gibbon 원숭이 백혈병 바이러스 (GALV)로부터 유래되는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 3

청구항 2에 있어서, MLV는 양쪽성 (amphotropic) MLV인 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 4

삭제

청구항 5

청구항 1에 있어서, 표적 세포는 세포 증식성 질환을 앓는 세포인 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 6

청구항 1에 있어서, 표적 세포는 암 세포인 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 7

청구항 5에 있어서, 세포 증식성 질환은 폐암, 결장-직장암, 유방암, 전립선암, 요도관암, 자궁암, 뇌암, 두경부암, 췌장암, 흑색종, 위암 및 난소암으로 구성된 군에서 선택되는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 8

청구항 1에 있어서, 폴리뉴클레오티드의 gag 핵산 도메인은 감마레트로바이러스로부터 유래되는 것을 특징으로

하는 재조합 복제 적격성 감마레트로바이러스.

청구항 9

청구항 8에 있어서, 감염 이후에 gag 핵산 도메인은 서열 번호: 19의 1203번 뉴클레오티드에서부터 2819번 뉴클레오티드까지의 프로바이러스 DNA 서열을 포함하는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 10

청구항 1에 있어서, 폴리뉴클레오티드의 pol 핵산 도메인은 감마레트로바이러스로부터 유래되는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 11

청구항 10에 있어서, 감염 이후에 pol 도메인은 서열 번호:19의 2820번 뉴클레오티드에서부터 6358번 뉴클레오티드까지의 프로바이러스 DNA 서열을 포함하는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 12

청구항 1에 있어서, 감염 이후에 env 도메인은 서열 번호:19의 6359번 뉴클레오티드에서부터 8323번 뉴클레오티드까지의 프로바이러스 DNA 서열을 포함하는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 13

청구항 1에 있어서, IRES는 뇌심근염 바이러스로부터 유래되는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 14

청구항 13에 있어서, 감염 이후에 IRES는 서열 번호:19의 8327번 뉴클레오티드에서부터 8876번 뉴클레오티드까지의 프로바이러스 DNA 서열을 포함하는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

청구항 15

청구항 1에 있어서, 감염 이후에 이종성 핵산은 8877번 뉴클레오티드에서부터 9353번 뉴클레오티드까지, 서열 번호: 19에서 진술된 서열을 포함하는 것을 특징으로 하는 재조합 복제 적격성 감마레트로바이러스.

발명의 설명

기 술 분 야

[0001] **관련된 출원에 대한 교차 참조**

[0002] 본 출원은 2008년 9월 26일 제출된 U.S. Provisional Application Serial No. 61/100,666, 2008년 12월 8일 제출된 U.S. Provisional Application Serial No. 61/120,618, 그리고 2009년 6월 13일 제출된 U.S. Provisional Application Serial No. 61/186,823에 우선권을 주장하고, 이들의 내용은 본 발명에 참조로서 편입된다.

[0003] **기술 분야**

[0004] 본 발명은 변형된 시토신 디아미나아제 (CD)에 관계한다. 본 발명은 또한, 이런 변형된 CD를 발현하는 세포, 그리고 질환과 질병의 치료에서 이런 변형된 CD를 이용하는 방법에 관계한다.

배 경 기 술

[0005] **배경기술**

[0006] 효모, 또는 세균 시토신 디아미나아제는 무해한 항생성 프로-드러그 5-FC를 세포독성 화학치료제 5-플루오르우라실 (5-FU)로 전환시킨다. 인간 (일반적으로, 포유동물)은 유의미한 시토신 디아미나아제 활성을 갖는 효소를 인코딩하는 자연 발생 유전자를 보유하지 못하는 것으로 알려져 있다. 효모와 세균 시토신 디아미나아제는 상기 효소의 전달을 위한 유전자 전달과 바이러스 벡터, 그 이후에 5-FC로 치료를 이용한 암의 치료에서 주목을 받고

있으며, 5-FC는 이후, 상기 효소에 의해 세포독성 약물로 전환된다 (Miller *et al.*, Can Res 62:773-780 2002; Kievit *et al.*, Can Res 59:1417-1421 1999).

발명의 내용

해결하려는 과제

[0007] **요약**

[0008] 본 발명에서는 5-FC를 5-FU로 전환시키는 폴리펩티드를 제시한다. 또한, 이런 폴리펩티드를 인코딩하는 핵산 분자, 이들 폴리펩티드를 발현하는 세포, 이들 폴리뉴클레오티드와 폴리펩티드를 포함하는 벡터, 그리고 이들 폴리펩티드를 이용하여 적절한 세포로부터 5-FU 또는 이의 유도체를 합성하는 방법이 제시된다. 따라서 다양한 구체예에서, SEQ ID NO: 2에서 열거된 바와 같은 서열을 갖는 시토신 디아미나아제를 포함하는 분리된 또는 재조합 폴리펩티드가 제시된다. 다른 구체예에서, 폴리펩티드는 A23L, V108I, I140L과 이들의 임의의 조합으로 구성된 군에서 선택되는 돌연변이를 보유하는 SEQ ID NO: 2를 포함한다. 또 다른 구체예에서, 폴리펩티드는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함하고 잔기: (a) 23, 108과 140을 제외하고 최대 50, 25, 10, 또는 5개의 보존성 아미노산 치환을 보유한다. 일반적으로, 본 발명에 제시된 폴리펩티드는 5-FC를 5-FU로 전환하는데 유용한 시토신 디아미나아제 활성을 나타낸다.

[0009] 본 발명에서는 또한, SEQ ID NO: 4에 최소한 80%, 90%, 95%, 98% 또는 99% 동일한 서열을 포함하는 폴리펩티드를 제시하고, 여기서 상기 폴리펩티드는 23번 위치에서 류신, 108번 위치에서 이소류신, 그리고 140번 위치에서 류신을 보유하고, 그리고 상기 폴리펩티드는 시토신 디아미나아제 활성을 갖는다. 또 다른 구체예에서, 상기 폴리펩티드는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함한다.

[0010] 본 발명에서는 또한, 우라실 포스포리보실전이효소 (uracil phosphoribosyltransferase, UPRT) 또는 오로테이트 포스포리보실전이효소 (orotate phosphoribosyltransferase, OPRT)에 작동가능하게 연결된 상기한 폴리펩티드 중에서 임의의 한 가지를 포함하는 융합 구조체를 제시한다. 한 구체예에서, 융합 구조체는 UPRT 또는 OPRT 활성을 갖는 두 번째 폴리펩티드에 연결된, A23L, V108I, I140L과 이들의 임의의 조합으로 구성된 군에서 선택되는 돌연변이를 보유하는 SEQ ID NO: 2를 포함하는 첫 번째 폴리펩티드를 포함한다. 또 다른 구체예에서, 융합 구조체는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 갖고 잔기: 23, 108과 140을 제외하고 최대 50, 25, 10, 또는 5개의 보존성 아미노산 치환을 보유하는 첫 번째 폴리펩티드를 포함하고, 여기서 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 갖는 두 번째 폴리펩티드에 연결된다. 본 발명에서는 또한, 시토신 디아미나아제 활성을 갖고 SEQ ID NO: 4에 최소한 80%, 90%, 95%, 98% 또는 99% 동일한 서열을 포함하는 첫 번째 폴리펩티드를 포함하는 융합 구조체를 제시하고, 여기서 상기 폴리펩티드는 23번 위치에서 류신, 108번 위치에서 이소류신, 그리고 140번 위치에서 류신을 보유하고, 그리고 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 포함하는 두 번째 폴리펩티드에 연결된다. 한 구체예에서, UPRT 또는 OPRT 활성을 갖는 폴리펩티드는 각각, SEQ ID NO: 8과 10에서 열거된 바와 같은 서열, 또는 이들의 변이체를 포함한다. 또 다른 구체예에서, 시토신 디아미나아제 활성을 포함하는 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 갖는 폴리펩티드에 연결되고, 여기서 이들 폴리펩티드는 펩티드 링커에 의해 분리된다. 다른 구체예에서, 융합 구조체는 SEQ ID NO: 12, 14, 16 또는 18로 구성된 군에서 선택되는 서열을 포함한다.

[0011] 본 발명에서는 또한, 상기한 폴리펩티드 중에서 임의의 한 가지를 인코딩하는 폴리뉴클레오티드를 제시한다. 가령, 본 발명에서는 A23L, V108I, I140L과 이들의 임의의 조합으로 구성된 군에서 선택되는 돌연변이를 보유하는 SEQ ID NO: 2를 포함하는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 제시한다. 또 다른 구체예에서, 폴리뉴클레오티드는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함하고 잔기: (a) 23, 108과 140을 제외하고 최대 50, 25, 10, 또는 5개의 보존성 아미노산 치환을 보유하는 폴리펩티드를 인코딩한다. 진전된 구체예에서, 본 발명에서는 SEQ ID NO: 4에 최소한 80%, 90%, 95%, 98% 또는 99% 동일한 서열을 포함하는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 제시하고, 여기서 상기 폴리펩티드는 23번 위치에서 류신, 108번 위치에서 이소류신, 그리고 140번 위치에서 류신을 보유하고, 그리고 상기 폴리펩티드는 시토신 디아미나아제 활성을 갖는다. 또 다른 구체예에서, 폴리뉴클레오티드는 SEQ ID NO: 4를 포함하는 폴리펩티드를 인코딩한다. 진전된 구체예에서, 폴리뉴클레오티드는 SEQ ID NO: 4, 6, 8, 10, 11, 12, 또는 13에서 열거된 바와 같은 서열을 포함하는 폴리펩티드를 인코딩한다.

[0012] 본 발명에서는 시토신 디아미나아제를 인코딩하는 인간 코돈 최적화된 폴리뉴클레오티드를 제시한다. 한 구체예에서, 인간 코돈 최적화된 폴리뉴클레오티드는 SEQ ID NO: 3에서 열거된 바와 같은 서열을 포함한다. 또 다른

구체예에서, 폴리뉴클레오티드는 SEQ ID NO: 5, 7, 또는 9에서 열거된 바와 같은 서열을 포함한다.

[0013] 다른 구체예에서, 본 발명의 시토신 디아미나아제 또는 융합 구조체는 유전자 전달 시스템 (GDS)을 이용하여 전달된다. 다른 측면에서, 시토신 디아미나아제를 인코딩하는 폴리뉴클레오티드는 바이러스 또는 바이러스 유래된 벡터인 GDS로 전달된다. 바이러스 벡터는 복제성 또는 비-복제성일 수 있고, 그리고 아데노바이러스 벡터, 홍역 벡터, 포진 벡터, 레트로바이러스 벡터 (렌티바이러스 벡터 포함), 랍도바이러스 벡터, 예를 들면, 수포성 구내염 바이러스 벡터, 레오바이러스 벡터, 세네카 밸리 바이러스 벡터, 폭스바이러스 벡터 (수두 또는 우두 유래된 벡터 포함), 파보바이러스 벡터 (AAV 벡터 포함), 알파바이러스 벡터 또는 당업자에게 공지된 기타 바이러스 벡터일 수 있다. 한 구체예에서, 바이러스 벡터는 복제성 포유동물 세포를 감염시킬 수 있는 복제 적격성 레트로바이러스 벡터일 수 있다. 복제 적격성 레트로바이러스 벡터는 오르토레트로바이러스 또는 더욱 전형적으로, 감마 레트로바이러스 벡터를 포함할 수 있다. 한 측면에서, 복제 적격성 레트로바이러스 벡터는 시토신 디아미나아제를 인코딩하는 본 발명의 폴리뉴클레오티드에 대하여 5'에 내부 리보솜 결합 부위 (IRES)를 포함한다. 한 구체예에서, 시토신 디아미나아제를 인코딩하는 폴리뉴클레오티드는 레트로바이러스 벡터의 ENV 폴리뉴클레오티드에 대하여 3'에 위치한다.

[0014] 다른 구체예에서, 본 발명의 시토신 디아미나아제 또는 융합 구조체, 또는 본 발명의 폴리뉴클레오티드 또는 융합 구조체를 포함하는 벡터로 형질감염되는 숙주 세포가 제시된다. 숙주 세포에는 진핵 세포, 예를 들면, 효모 세포, 곤충 세포, 또는 동물 세포가 포함된다. 숙주 세포에는 또한, 원핵 세포, 예를 들면, 세균 세포가 포함된다.

[0015] 본 발명에서는 또한, 암을 비롯한 세포 증식성 질환을 치료하는 방법을 제시하고, 상기 방법은 본 발명의 폴리뉴클레오티드 또는 폴리펩티드를 개체에 투여하는 단계, 그리고 상기 개체를 5-플루시토신 (5-FC)을 포함하는 세포독성 약물과 접촉시키는 단계를 포함한다.

[0016] 본 발명에서는 레트로바이러스 GAG 단백질; 레트로바이러스 POL 단백질; 레트로바이러스 외피; 레트로바이러스 폴리뉴클레오티드, 여기서 상기 레트로바이러스 폴리뉴클레오티드는 3' 말단에서 긴 말단 반복 (Long-Terminal Repeat, LTR) 서열, 5' 말단에서 프로모터 서열, gag 핵산 도메인, pol 핵산 도메인, 그리고 env 핵산 도메인을 포함하고, 상기 프로모터는 포유동물 세포에서 발현에 적합하고; 이종성 폴리뉴클레오티드에 작동가능하게 연결된 내부 리보솜 결합 부위 (IRES)를 포함하는 카세트, 여기서 상기 카세트는 3' LTR에 대하여 5'에, 그리고 레트로바이러스 외피를 인코딩하는 env 핵산 도메인에 대하여 3'에 배치되고; 그리고 표적 세포 내에서 역전사, 포장과 통합에 필요한 cis-작용 서열을 포함하는 재조합 복제 적격성 레트로바이러스 (RCR)를 제시하고, 여기서 상기 RCR은 pACE 벡터 (SEQ ID NO: 21)와 비교하여 6회 계대 (passage)후 더욱 높은 복제 적격을 유지한다. 한 구체예에서, 레트로바이러스 폴리뉴클레오티드 서열은 뮤린 백혈병 바이러스 (murine leukemia virus, MLV), Moloney 뮤린 백혈병 바이러스 (MoMLV), 고양이 백혈병 바이러스 (Feline leukemia Virus) 또는 Gibbon 원숭이 백혈병 바이러스 (Gibbon ape leukemia virus, GALV)로부터 유래된다. 다른 구체예에서, MLV는 양쪽성 MLV이다. 또 다른 구체예에서, 레트로바이러스는 온코레트로바이러스 (oncoretrovirus)이다. 또 다른 구체예에서, 표적 세포는 세포 증식성 질환을 앓는 세포이다. 세포 증식성 질환은 폐 암, 결장-직장 암, 유방 암, 전립선 암, 요도관 암, 자궁 암, 뇌 암, 두경부암, 췌장 암, 흑색종, 위 암과 난소 암, 류머티스성 관절염 또는 기타 자가-면역 질환으로 구성된 군에서 선택될 수 있다. 한 구체예에서, 프로모터는 1번 뉴클레오티드에서부터 대략 582번 뉴클레오티드까지 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열을 갖는 CMV 프로모터를 포함하고, 하나 이상의 핵산 염기에 변형을 보유할 수 있고, 그리고 전사를 주동하고 개시할 수 있다. 더욱 진전된 구체예에서, 프로모터는 1번 뉴클레오티드에서부터 대략 582번 뉴클레오티드까지 SEQ ID NO: 19, 20 또는 22에서 열거된 바와 같은 서열을 포함한다. 진전된 구체예에서, 프로모터는 CMV-R-U5 도메인 폴리뉴클레오티드를 포함한다. 한 구체예에서, CMV-R-U5 도메인은 MLV R-U5 영역에 연결된 인간 사이토메갈로바이러스로부터 극초기 프로모터를 포함한다. 또 다른 구체예에서, CMV-R-U5 도메인 폴리뉴클레오티드는 대략 1번 뉴클레오티드에서부터 대략 1202번 뉴클레오티드까지 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열, 또는 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열에 최소한 95% 동일한 서열을 포함하고, 여기서 상기 폴리뉴클레오티드는 그것에 작동가능하게 연결된 핵산 분자의 전사를 촉진한다. 다른 구체예에서, 상기 폴리뉴클레오티드의 gag와 pol은 온코레트로바이러스로부터 유래된다. gag 핵산 도메인은 SEQ ID NO: 19의 대략 1203번 뉴클레오티드에서부터 대략 2819번 뉴클레오티드까지의 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.8% 동일성을 갖는 서열을 포함할 수 있다. pol 도메인은 SEQ ID NO: 19의 대략 2820번 뉴클레오티드에서부터 대략 6358번 뉴클레오티드까지의 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.9% 동일성을 갖는 서열을 포함할 수 있다. 한 구체예에서, env 도메인은 양쪽성 env 단백질을 인코딩한다. env 도메인은 SEQ ID NO: 19의 대략 6359번 뉴

클레오티드에서부터 대략 8323번 뉴클레오티드까지의 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.8% 동일성을 갖는 서열을 포함할 수 있다. 상기 벡터의 IRES 도메인은 임의의 IRES일 수 있지만, 한 구체예에서, IRES는 뇌심근염 바이러스로부터 유래된다. 진전된 구체예에서, IRES는 SEQ ID NO: 19의 대략 8327번 뉴클레오티드에서부터 대략 8876번 뉴클레오티드까지의 서열, 또는 그것에 최소한 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함한다. 더욱 진전된 구체예에서, 이종성 폴리뉴클레오티드는 SEQ ID NO: 3, 5, 11, 13, 15 또는 17에서 열거된 바와 같은 서열을 갖는 폴리뉴클레오티드를 포함한다. 진전된 구체예에서, 이종성 서열은 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함하는 폴리펩티드를 인코딩한다. 이종성 핵산은 인간 코돈 최적화되고 SEQ ID NO: 4에서 열거된 바와 같은 폴리펩티드를 인코딩한다. 진전된 구체예에서, 이종성 핵산은 대략 8877번 뉴클레오티드에서부터 대략 9353번 뉴클레오티드까지 SEQ ID NO: 19에서 열거된 바와 같은 서열을 포함한다. 한 구체예에서, 3' LTR은 온코레트로바이러스로부터 유래된다. 진전된 구체예에서, 3' LTR은 U3-R-U5 도메인을 포함한다. 더욱 진전된 구체예에서, 3' LTR은 대략 9405번 뉴클레오티드에서부터 대략 9998번 뉴클레오티드까지 SEQ ID NO: 19에서 열거된 바와 같은 서열, 또는 그것에 최소한 95%, 98% 또는 99.5% 동일한 서열을 포함한다.

[0017] 본 발명에서는 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열을 포함하는 폴리뉴클레오티드를 제시한다.

[0018] 본 발명에서는 5'에서 3'방향으로, MLV R-U5 영역에 인간 사이토메갈로바이러스로부터 극초기 프로모터의 CMV-R-U5 융합체; 역전사효소에 대한 PBS, 프라이머 결합 부위; 5' 접합 부위; Ψ 포장 신호; MLV 군 특정 항원에 대한 gag 코딩 서열; MLV 중합효소 다중단백질에 대한 pol 코딩 서열; 3' 접합 부위; MLV 균주 4070A의 외피 단백질에 대한 4070A env 코딩 서열; 뇌심근염 바이러스로부터 내부 리보솜 결합 부위 (IRES); 변형된 시토신 디아미나아제 코딩 서열; 폴리퓨린 다발 (polypurine tract); 그리고 U3-R-U5 MLV 긴 말단 반복을 포함하는 분리된 폴리뉴클레오티드를 제시한다.

[0019] 본 발명에서는 세포 증식성 질환을 앓는 개체를 치료하는 방법을 제시하고, 상기 방법은 시토신 디아미나아제 활성을 갖는 본 발명의 폴리뉴클레오티드가 발현되도록 하는 조건 하에 상기 폴리뉴클레오티드와 개체를 접촉시키는 단계, 그리고 상기 개체를 5-플루시토신과 접촉시키는 단계를 포함한다.

[0020] 본 발명에서는 또한, 개체에서 세포 증식성 질환을 치료하는 방법을 제시하고, 상기 방법은 본 발명의 레트로바이러스와 개체를 접촉시키는 단계를 포함하고, 여기서 이종성 핵산 서열은 종양 세포의 증식을 저해하는 치료 단백질을 인코딩한다. 한 구체예에서, 레트로바이러스는 SEQ ID NO: 4, 12, 14, 16, 또는 18에서 열거된 바와 같은 서열을 포함하는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 포함한다.

[0021] 본 발명의 하나 이상의 구체예의 상세는 첨부된 도면 및 하기 상세한 설명에서 설명된다. 다른 특징, 목적, 그리고 이점은 상세한 설명과 도면, 그리고 특허청구범위로부터 명백할 것이다.

과제의 해결 수단

[0022] **상세한 설명**

[0023] 본 명세서에서, 그리고 첨부된 특허청구범위에서, 단수 형태는 달리 명시되지 않으면, 복수 지시물을 포함한다. 따라서 예로써, "세포"에 대한 지시는 복수의 세포를 포함하고, 그리고 "작용제"에 대한 지시는 당업자에게 공지된 한 가지 이상의 작용제에 대한 지시를 포함한다.

[0024] 또한, "또는"의 이용은 달리 명시되지 않으면 "및/또는"을 의미한다. 유사하게, "포함한다," "포함하는," "보유한다," 그리고 "보유하는"은 동의어로서 이용되고 한정하는 것으로 의도되지 않는다.

[0025] 또한, 다양한 구체예의 설명이 용어 "포함하는"을 이용하는 경우에, 당업자는 일부 특정한 사례에서, 구체예가 용어 "본질적으로 구성되는" 또는 "구성되는"을 이용하여 대안적으로 기술될 수 있음을 이해할 것이다.

[0026] 달리 정의되지 않으면, 본 명세서에 이용된 모든 기술 용어와 과학 용어는 본 발명이 속하는 분야에 통상의 지식을 가진 자에게 일반적으로 이해되는 바와 동일한 의미를 갖는다. 본 명세서에 기술된 것들과 유사하거나 동등한 방법과 물질이 이들 개시된 방법과 조성물의 실시에 이용될 수 있긴 하지만, 전형적인 방법, 장치와 물질이 본 명세서에 기술된다.

[0027] 본 명세서 전반에서 언급된 간행물은 오로지, 본 출원의 출원일 이전에 이들의 공개 목적으로만 제공된다. 여기서 어떤 것도 본 발명자들이 보다 앞선 공개에 의해 이런 공개를 선행할 권리가 없다는 용인으로서 간주되지 않아야 한다.

[0028] 시토신 디아미나아제 (EC 3.5.4.1)는 하기 화학 반응을 촉진하는 효소이다:

[0029] 시토신 + H_2O → 우라실 + NH_3

[0030] 따라서 상기 효소의 2가지 기질은 시토신과 H_2O이고, 반면 이의 2가지 산물은 우라실과 NH_3이다. 상기 효소는 가수분해효소 (hydrolase)에 집단에 속하는데, 이들은 특히, 환상 아미딘 (cyclic amidine) 내에서 펩티드 결합이 아닌 탄소-질소 결합에 작용한다. 상기 효소 군의 체계적 명칭 (systematic name)은 시토신 아미노가수분해효소이다. 상기 효소는 또한, 이소시토신 디아미나아제로 불린다. 상기 효소는 피리미딘 물질대사에 참여한다.

[0031] 더욱 구체적으로, 시토신 디아미나아제는 피리미딘에 대한 물질대사 경로에 관여하는 효소이고, 상기 경로를 통하여 외인성 시토신이 가수분해성 탈아미노화 (hydrolytic deamination)를 통해 우라실로 변형된다. 시토신 디아미나아제 (CDase 또는 CD) 활성은 원핵생물과 하등 진핵생물에서 증명되었지만, 이들은 포유동물에서 부재한다 (Koechlin *et al.*, 1966, Biochem. Pharmacol. 15, 435-446; Polak *et al.*, 1976, Chemotherapy 22, 137-153). 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*, 또는 *S. cerevisiae*)의 *FCY1* 유전자, 그리고 대장균 (*E. coli*)의 *coda* 유전자는 각각, 이들 두 생물체의 CDase를 인코딩하는 것으로 알려져 있고, 그들의 서열은 공개되어 있다 (EP 402 108; Erbs *et al.*, 1997, Curr. Genet. 31, 1-6; WO 93/01281). CDase는 또한, 시토신 유사체, 5-플루시토신 (5-FC)을 5-플루오르우라실 (5-FU)로 탈아미노화시키고, 이는 특히, 5-플루오르-UMP (5-FUMP)로 전환될 때 고도 세포독성 화합물이다. 상기 효소를 인코딩하는 유전자의 불활성화 돌연변이, 또는 상기 효소에 대한 자연적인 결핍 (가령, 포유동물 세포)으로 인하여, CDase 활성이 부재하는 세포는 5-FC에 저항한다 (Jund and Lacroute, 1970, J. Bacteriol. 102, 607-615; Kilstrup *et al.*, 1989, J. Bacteriol., 171, 2124-2127). 다른 한편으로, CDase 활성을 인코딩하는 서열이 이전된 포유동물 세포에 5-FC 민감성을 전달하는 것이 가능한 것으로 증명되었다 (Huber *et al.*, 1993, Cancer Res. 53, 4619-4626; Mullen *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89, 33-37; WO 93/01281). 따라서 CD의 이용은 유전자 요법, 특히 항암 유전자 요법의 배경에서 유리하다. 하지만, 5-FC 민감성은 세포주에 따라, 상당히 변화한다. 낮은 민감성은 예로써, 대장균 (*E. coli*)의 *coda* 유전자를 발현하는 레트로바이러스로 형질도입된 PANC-1 (췌장의 암종)과 SK-BR-3 (유방선암종) 인간 종양 세포주에서 관찰된다 (Harris *et al.*, 1994, Gene Therapy 1, 170-175). 이러한 현상은 CDase의 효소 작용에 의해 형성된 5-FU의 세포독성 5-FUMP로의 내인성 전환 (endogenous conversion)의 부재 또는 부실에 의해 설명될 수 있다. 포유동물 세포 내에서 오로테이트 포스포리보실전이효소 (OPRTase)에 의해 정상적으로 수행되는 이러한 단계는 일정한 종양에서 부재하고, 따라서 CDase에 기초된 유전자 요법을 무력화시킬 수 있다. 원핵동물과 하등 진핵동물에서, 우라실은 우라실 포스포리보실전이효소 (UPRTase 활성)의 작용을 통해 UMP로 변형된다. 상기 효소는 또한, 5-FU를 5-FUMP로 전환시킨다. 중요하게는, 세균 우라실 포스포리보실전이효소 (UPRT)는 포유동물 세포의 오로테이트 포스포리보실전이효소 (OPRT) 또는 우리딘-5'-모노포스페이트 신타아제와 기능적으로 동등하다. 이들 효소는 5-플루오르우라실 (5-FU) (우라실의 플루오르화된 유사체)의 5-플루오르우리딘 5' 모노포스페이트 (5-FUMP)로의 전환을 매개한다. 5-플루오르우리딘 5' 모노포스페이트는 차후에, 포유동물 de novo 피리미딘 경로를 통해 5-FdUDP와 FdUTP로 전환된다. 각 5-FdUTP는 티미딜레이트 신타아제 (Thy-A)의 비가역적 저해물질이고 dTTP 고갈을 유발한다. 이러한 전환은 5-플루오르우라실의 세포독성 효과를 달성하기 위한 필수 경로 중의 하나이고, 그리고 세균 기원의 세균 우라실 포스포리보실전이효소는 포유동물 오로테이트 포스포리보실전이효소가 하는 것처럼, 5-플루오르우라실을 동일한 활성 대사산물로 전환시킬 수 있는 것으로 폭넓게 인정되고 있다. UPRTase 활성 또는 OPRTase 활성의 부재에서, CDase에 의한 5-FC의 탈아미노화로부터 기원하는 5-FU는 세포독성 5-FUMP로 변형되지 않는다 (Jund and Lacroute, 1970, J. Bacteriol. 102, 607-615).

[0032] 본 명세서에 기술된 바와 같이, 본 발명에서는 시토신 디아미나아제 활성을 갖는 첫 번째 폴리펩티드와 UPRT 또는 OPRT 활성을 포함하는 두 번째 폴리펩티드의 융합체를 포함하는 폴리펩티드 및 이들 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 제시한다. 이들 융합 구조체는 우라실, 또는 이의 유도체의 모노포스페이트 유사체로의 전환, 특히 5-FU의 5-FUMP로의 전환에 유용하다.

[0033] 하기에 더욱 상세하게 기술된 바와 같이, 본 발명은 5-FC의 5-FU로의 전환을 촉진하는 신규한 폴리펩티드의 산출과 발현에 최소한 부분적으로 기초한다. 한 구체예에서, 5-FC의 5-FU로의 전환을 위한 향상된 촉진 특성을 갖도록 조작된 신규한 폴리펩티드가 제시된다. 이들 폴리펩티드에는 특정 잔기에서 아미노산 치환을 보유하도록 변경된 변이체가 포함된다. 이들 변이체가 하기에 더욱 상세하게 기술되긴 하지만, 본 발명의 폴리펩티드는 하나 이상의 변형된 아미노산을 포함할 수 있는 것으로 이해된다. 변형된 아미노산의 존재는 예로써, (a) 폴리펩티드의 생체내 반감기를 증가시키고, (b) 폴리펩티드 항원성을 감소시키거나 증가시키고, 그리고 (c) 폴리펩티

드 보관 안정성을 증가시키는데 유리할 수 있다. 아미노산(들)은 재조합 생산 동안 예로써, 번역과 동시에 또는 번역 이후에 변형 (가령, 포유동물 세포 내에서 발현 동안 N--X--S/T 모티프에서 N-연결된 당화)되거나, 또는 합성 수단에 의해 변형된다. 따라서 "돌연변이체", "변이체" 또는 "변형된" 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포는 부모 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포로부터 변경 또는 유래되었거나, 또는 어떤 방식으로든 상이하거나 변화된 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포를 의미한다. 돌연변이 또는 변형된 단백질 또는 효소는 반드시 그러한 것은 아니지만 일반적으로, 돌연변이 폴리뉴클레오티드 또는 유전자로부터 발현된다.

[0034] "돌연변이"는 돌연변이 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포를 발생시키는 임의의 공정 또는 기전을 의미한다. 여기에는 단백질, 효소, 폴리뉴클레오티드, 또는 유전자 서열이 변경되는 임의의 돌연변이, 그리고 이런 돌연변이에 기인하는 세포 내에서 임의의 검출가능한 변화가 포함된다. 전형적으로, 돌연변이는 단일 또는 복수 뉴클레오티드 잔기의 점 돌연변이, 결실, 또는 삽입에 의해, 폴리뉴클레오티드 또는 유전자 서열 내에서 발생한다. 돌연변이에는 유전자의 단백질-인코딩 영역 내에서 발생하는 폴리뉴클레오티드 변경, 그리고 단백질-인코딩 서열의 외부 영역, 예를 들면, 조절 서열 또는 프로모터 서열에서 변경이 포함된다. 유전자에서 돌연변이는 "침묵"일 수 있고, 다시 말하면, 발현 시에 아미노산 변경에 반영되지 않고, 상기 유전자의 "서열-보존성" 변이체를 산출할 수 있다. 이는 일반적으로, 하나의 아미노산이 하나 이상의 코돈에 대응할 때 발생한다.

[0035] 변형된 아미노산의 무제한적 실례에는 당화 아미노산, 황산화 아미노산, 프레닐화 (가령, 파르네실화, 게라닐게라닐화) 아미노산, 아세틸화 아미노산, 아실화 아미노산, 페길화 아미노산, 비오틴화 아미노산, 카르복실화 아미노산, 포스포릴화 아미노산 등이 포함된다. 아미노산의 변형에서 기술을 보도하는데 적합한 참고문헌은 본 명세서 전반에서 충분하다. 예시적인 프로토콜은 Walker (1998) Protein Protocols on CD-ROM (Humana Press, Towata, N.J.)에서 확인된다.

[0036] 본 발명의 변형된 CD 폴리펩티드와 폴리뉴클레오티드를 생산하고, 분리하고, 그리고 이용하기 위한 재조합 방법은 본 명세서에서 기술된다. 재조합 생산에 더하여, 이들 폴리펩티드는 고체-상 (solid-phase) 기술을 이용한 직접적인 펩티드 합성에 의해 생산될 수도 있다 (가령, Stewart *et al.* (1969) Solid-Phase Peptide Synthesis (WH Freeman Co, San Francisco); 그리고 Merrifield (1963) J. Am. Chem. Soc. 85: 2149-2154). 펩티드 합성은 수동 기술을 이용하여, 또는 자동화에 의해 수행될 수 있다. 자동화된 합성은 예로써, 제조업체에 의해 제공된 사용설명서에 따라서, Applied Biosystems 431A 펩티드 합성기 (Perkin Elmer, Foster City, Calif.)를 이용하여 달성된다.

[0037] 예시로써, 대장균 (*E. coli*) (Anderson *et al.*, 1992, Eur. J. Biochem 204, 51-56), 락토코커스 락티스 (*Lactococcus lactis*) (Martinussen and Hammer, 1994, J. Bacteriol. 176, 6457-6463), 미코박테리움 보비스 (*Mycobacterium bovis*) (Kim *et al.*, 1997, Biochem Mol. Biol. Int 41, 1117-1124), 그리고 바실루스 서브틸리스 (*Bacillus subtilis*) (Martinussen *et al.*, 1995, J. Bacteriol. 177, 271-274)의 UPRTase를 인코딩하는 핵산 서열이 본 발명의 배경에서 이용될 수 있다. 하지만, 효모 UPRTase, 특히 사카로마이세스 세레비지애 (*S. cerevisiae*)의 FUR1 유전자에 의해 인코딩되는 효모 UPRTase (이의 서열은 Kern *et al.* (1990, Gene 88, 149-157)에서 개시된다)의 이용은 본 발명에 참조로서 편입된다. 지시로써, 이들 유전자의 서열 및 상응하는 UPRTase의 서열은 기존 문헌 및 특수 데이터 뱅크 (SWISSPROT, EMBL, Genbank Medline 등)에서 확인될 수 있다.

[0038] 본 명세서에서 동의어로서 이용되는 용어 "단백질" 또는 "폴리펩티드"는 펩티드 결합으로 불리는 화학 결합에 의해 서로 연결되는 아미노산으로 불리는 화학적 빌딩 블록 (building block)의 하나 이상의 사슬을 포함한다. "효소"는 바람직하게는, 다소간 특이적으로 한 가지 이상의 화학적 또는 생화학적 반응을 촉진하거나 증진하는 단백질로 완전히 또는 대부분 구성되는 임의의 물질을 의미한다. 용어 "효소"는 또한, 촉매성 폴리뉴클레오티드 (가령, RNA 또는 DNA)를 지칭한다. "고유" 또는 "야생형" 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포는 자연에서 발생하는 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포를 의미한다.

[0039] 따라서 다양한 구체예에서, A23L, V108I, I140L과 이들의 임의의 조합으로 구성된 군에서 선택되는 돌연변이를 보유하는 SEQ ID NO: 2를 포함하는 분리된 또는 재조합 폴리펩티드가 제시된다. 또 다른 구체예에서, 폴리펩티드는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함하고 잔기: (a) 23, 108과 140을 제외하고 최대 50, 25, 10, 또는 5개의 보존성 아미노산 치환을 보유한다. 일반적으로, 본 발명에서 제시된 폴리펩티드는 5-FC를 5-FU로 전환시키는데 유용한 시토신 디아미나아제 활성을 나타낸다.

[0040] 본 발명에서는 또한, SEQ ID NO: 4에 최소한 80%, 90%, 95%, 98% 또는 99% 동일한 서열을 포함하는 폴리펩티드

를 제시하고, 여기서 상기 폴리펩티드는 23번 위치에서 류신, 108번 위치에서 이소류신, 그리고 140번 위치에서 류신을 보유하고, 그리고 상기 폴리펩티드는 시토신 디아미나아제 활성을 갖는다. 또 다른 구체예에서, 상기 폴리펩티드는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함한다.

[0041] 본 발명에서는 또한, 우라실 포스포리보실전이효소 (uracil phosphoribosyltransferase, UPRT) 또는 오로테이트 포스포리보실전이효소 (orotate phosphoribosyltransferase, OPRT)에 작동가능하게 연결된 상기한 폴리펩티드 중에서 임의의 한 가지를 포함하는 융합 구조체를 제시한다. 한 구체예에서, 융합 구조체는 UPRT 또는 OPRT 활성을 갖는 두 번째 폴리펩티드에 연결된, A23L, V108I, I140L과 이들의 임의의 조합으로 구성된 군에서 선택되는 돌연변이를 보유하는 SEQ ID NO: 2를 포함하는 첫 번째 폴리펩티드를 포함한다. 또 다른 구체예에서, 융합 구조체는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 갖고 잔기: 23, 108과 140을 제외하고 최대 50, 25, 10, 또는 5개의 보존성 아미노산 치환을 보유하는 첫 번째 폴리펩티드를 포함하고, 여기서 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 갖는 두 번째 폴리펩티드에 연결된다. 본 발명에서는 또한, 시토신 디아미나아제 활성을 갖고 SEQ ID NO: 4에 최소한 80%, 90%, 95%, 98% 또는 99% 동일한 서열을 포함하는 첫 번째 폴리펩티드를 포함하는 융합 구조체를 제시하고, 여기서 상기 폴리펩티드는 23번 위치에서 류신, 108번 위치에서 이소류신, 그리고 140번 위치에서 류신을 보유하고, 그리고 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 포함하는 두 번째 폴리펩티드에 연결된다. 한 구체예에서, UPRT 또는 OPRT 활성을 갖는 폴리펩티드는 각각, SEQ ID NO: 8과 10에서 열거된 바와 같은 서열, 또는 이들의 변이체를 포함한다. 또 다른 구체예에서, 시토신 디아미나아제 활성을 포함하는 첫 번째 폴리펩티드는 UPRT 또는 OPRT 활성을 갖는 폴리펩티드에 연결되고, 여기서 이들 폴리펩티드는 펩티드 링커에 의해 분리된다. 다른 구체예에서, 융합 구조체는 SEQ ID NO: 11, 12, 또는 13으로 구성된 군에서 선택되는 서열을 포함한다.

[0042] 용어 "작동가능하게 연결된" 또는 "작동가능하게 연관된"은 조절 서열과 상기 조절 서열에 의해 조절되는 폴리뉴클레오티드 사이에, 또는 2개의 별개의 폴리펩티드 또는 이들 폴리펩티드를 인코딩하는 폴리뉴클레오티드 사이에 기능적 연쇄 또는 연합을 지칭한다.

[0043] CD 활성을 갖는 폴리펩티드 및 UPRT 또는 OPRT 활성을 포함하는 폴리펩티드를 포함하는 융합 구조체는 단백질 회수를 보조하기 위한 절단 부위 (cleavage site) 또는 다른 링커 모이어티를 보유하도록 조작될 수 있다. 전형적으로, 링커는 펩티드 링커 모이어티일 것이다. 링커 모이어티의 길이는 CD 폴리펩티드와 UPRT 또는 OPRT 폴리펩티드의 생물학적 활성을 최적화하도록 선택되고 과도한 실험 없이 경험적으로 결정될 수 있다. 링커 모이어티는 기질이 CD 폴리펩티드 및 UPRT 또는 OPRT 폴리펩티드와 상호작용할 수 있도록 할 만큼 충분히 길고 유연해야 한다. 링커 모이어티는 대략 1개 내지 30개 아미노산 잔기, 전형적으로 대략 2개 내지 15개 아미노산 잔기 길이의 펩티드이다. 링커 모이어티의 실례는 --Gly--Gly--, GGGGS (SEQ ID NO: 23), $(GGGGS)_N$ (SEQ ID NO: 23), $(SGGGG)_N$ (SEQ ID NO: 24), GKSSGSGSESKS (SEQ ID NO: 25), GSTSGSGKSSEGKG (SEQ ID NO: 26), GSTSGSGKSSEGSGSTKG (SEQ ID NO: 27), GSTSGSGKPGSGEGSTKG (SEQ ID NO: 28), 또는 EGKSSGSGSESKEF (SEQ ID NO: 29)이다. 연결 모이어티는 예로써, Huston *et al.*, Proc. Nat'l Acad. Sci 85:5879, 1988; Whitlow *et al.*, Protein Engineering 6:989, 1993; 그리고 Newton *et al.*, Biochemistry 35:545, 1996에서 기술된다. 다른 적합한 펩티드 링커는 본 발명에 참조로서 편입되는 U.S. Pat. No. 4,751,180과 4,935,233에서 기술된 것들이다. 원하는 펩티드 링커를 인코딩하는 DNA 서열은 임의의 적절한 통상적인 기술을 이용하여, CD 폴리펩티드, 그 이후에 UPRT 또는 OPRT 폴리펩티드를 인코딩하는 폴리뉴클레오티드 사이에, 또는 상기 폴리뉴클레오티드와 동일한 개방 해독틀 (reading frame) 내에 삽입될 수 있다. 가령, 링커를 인코딩하는 화학적으로 합성된 올리고뉴클레오티드는 2개의 코딩 폴리뉴클레오티드 사이에 결찰될 수 있다. 특정 구체예에서, 융합 폴리펩티드는 펩티드 링커에 의해 분리되는 2개 내지 4개의 별개의 도메인 (가령, CD 도메인 및 UPRT 또는 OPRT)을 포함한다.

[0044] 특정 서열의 "보존성 아미노산 치환" 또는, 간단하게, "보존성 변이"는 본질적으로 동일한 아미노산 서열로 1개의 아미노산, 또는 일련의 아미노산의 치환을 지칭한다. 당업자가 인지하는 바와 같이, 인코딩된 서열 내에서 단일 아미노산 또는 일정한 비율의 아미노산을 변경, 부가 또는 결실시키는 개별 치환, 결실 또는 부가는 "보존성 변이"를 결과하는데, 여기서 이들 변경은 아미노산의 결실, 아미노산의 부가, 또는 화학적으로 유사한 아미노산으로 아미노산의 치환을 유발한다.

[0045] 기능적으로 유사한 아미노산을 제공하는 보존성 치환 표는 당분야에 널리 공지되어 있다. 가령, 보존성 치환 기에는 알라닌 (A), 세린 (S), 그리고 트레오닌 (T)이 포함된다. 다른 보존성 치환 기에는 아스파르트산 (D)과 글루타민산 (E)이 포함된다. 다른 보존성 치환 기에는 아스파라긴 (N)과 글루타민 (Q)이 포함된다. 또 다른 보존성 치환 기에는 아르기닌 (R)과 리신 (K)이 포함된다. 다른 보존성 치환 기에는 이소류신, (I) 류신 (L), 메티

오닌 (M), 그리고 발린 (V)이 포함된다. 다른 보존성 치환 기에는 페닐알라닌 (F), 티로신 (Y), 그리고 트립토판 (W)이 포함된다.

[0046] 따라서 열거된 폴리펩티드 서열 (가령, SEQ ID NO: 2, 4, 6, 8, 10, 11, 12, 13 등)의 "보존성 아미노산 치환"에는 동일한 보존성 치환 기의 보존적으로 선택된 아미노산으로, 상기 폴리펩티드 서열의 일정한 비율, 전형적으로 10% 이하의 아미노산의 치환이 포함된다. 따라서 본 발명의 폴리펩티드의 보존성 치환된 변이는 동일한 보존성 치환 기의 보존성 치환된 변이로, 100, 75, 50, 25, 또는 10개의 치환을 보유할 수 있다.

[0047] 효소의 "활성"은 반응을 촉진하는, 다시 말하면, "기능하는" 능력의 척도이고, 상기 반응의 산물이 생산되는 속도로서 표시될 수 있다. 가령, 효소 활성은 시간 단위당 또는 효소 단위 (가령, 농도 또는 중량)당 생산되는 산물의 양으로서, 또는 친화력 또는 해리 상수의 항으로 표시될 수 있다. 본 명세서에서 동의어로서, "시토신 디아미나아제 활성", "시토신 디아미나아제의 생물학적 활성" 또는 "시토신 디아미나아제의 기능적 활성"은 표준 기술에 따라 생체내 또는 시험관내에서 측정될 때, 시토신 디아미나아제 기질에 대해 시토신 디아미나아제 단백질, 또는 본 발명의 폴리펩티드에 의해 발휘되는 활성을 지칭한다. 시토신 디아미나아제 활성을 측정하는 검사법은 당분야에 공지되어 있다. 가령, 시토신 디아미나아제 활성은 5-FC의 5-FU로의 전환 또는 시토신의 우라실로의 전환의 속도를 결정함으로써 측정될 수 있다. 5-FC, 5-FU, 시토신과 우라실의 검출은 색층분석 및 당분야에 공지된 다른 방법에 의해 수행될 수 있다.

[0048] 당업자가 인지하는 바와 같이, 개시되는 핵산 구조체의 많은 보존성 변이는 기능적으로 동일한 구조체를 산출한다. 가령, 앞서 언급된 바와 같이, 유전자 코드의 축중성 (degeneracy)으로 인하여, "침묵 치환" (즉, 인코딩된 폴리펩티드에서 변경을 유발하지 않는 핵산 서열에서 치환)은 아미노산을 인코딩하는 모든 핵산 서열의 함축된 특징이다. 당업자가 인지하는 바와 같이, 유전자 코드의 축중성으로 인하여, 본 발명의 변형된 시토신 디아미나아제 폴리펩티드를 인코딩하는 다수의 뉴클레오티드 서열이 생산될 수 있는데, 이들 중에서 일부는 본 명세서에 명시된 핵산 서열에 실질적인 동일성을 갖는다. 가령, 코돈 AGA, AGG, CGA, CGC, CGG, 그리고 CGU는 모두 아미노산 아르기닌을 인코딩한다. 따라서 본 발명의 핵산 내에서 아르기닌이 코돈에 의해 지정되는 모든 위치에서, 상기 코돈은 인코딩된 폴리펩티드를 변경하지 않으면서, 앞서 기술된 상응하는 코돈 중에서 한 가지로 변경될 수 있다. RNA 서열에서 U는 DNA 서열에서 T에 해당하는 것으로 이해된다.

[0049] 유사하게, 아미노산 서열 내에 1개 또는 여러 개의 아미노산이 매우 유사한 특성을 갖는 상이한 아미노산으로 치환되는 "보존성 아미노산 치환"은 개시된 구조체에도 매우 유사한 것으로 용이하게 확인된다. 각 개시된 서열의 이런 보존성 변이는 본 발명에서 제시된 폴리펩티드의 특징이다.

[0050] "보존성 변이체"는 극성 또는 비-극성 특성, 크기, 형상과 전하를 비롯한 유사한 특성을 갖는 아미노산으로 소정의 아미노산의 치환이 포함되지만 이들에 국한되지 않는 변이에 의해, 단백질 또는 효소의 전반적인 형태와 기능의 변경 없이, 소정의 아미노산 잔기가 변화된 단백질 또는 효소이다. 보존성인 것으로 지시되는 것들 이외의 아미노산은 유사한 기능의 임의의 2개의 단백질 사이에 단백질 또는 아미노산 서열 유사성 비율이 변하고, 예로써 정렬 설계 (alignment scheme)에 따라 측정될 때 최소한 30%, 최소한 50%, 최소한 70%, 최소한 80%, 또는 최소한 90%가 될 수 있도록 단백질 또는 효소 내에서 상이할 수 있다. 본 명세서에서, "서열 유사성"은 뉴클레오티드 또는 단백질 서열이 관련되는 정도를 의미한다. 2개의 서열 사이에 유사성의 정도는 서열 동일성 및/또는 보존 비율에 기초될 수 있다. 본 명세서에서, "서열 동일성"은 2개의 뉴클레오티드 또는 아미노산 서열이 불변하는 정도를 의미한다. "서열 정렬"은 유사성 정도를 평가하기 위한 목적으로 동일성 (그리고, 아미노산 서열의 경우에, 보존)의 최대 수준을 달성하기 위하여 2개 이상의 서열을 일렬로 세우는 과정을 의미한다. 서열을 정렬하고 유사성/동일성을 평가하기 위한 다수의 방법, 예를 들면, Cluster 방법은 당분야에 공지되어 있고, 여기서 유사성은 MEGALIGN 알고리즘, 그리고 BLASTN, BLASTP와 FASTA에 기초된다 (Lipman and Pearson, 1985; Pearson and Lipman, 1988). 이들 프로그램을 모두 이용할 때, 바람직한 설정은 최대 서열 유사성을 결과하는 설정이다. 가령, 정렬된 아미노산 서열의 특정 쌍에 대하여 "동일성" 또는 "동일성 비율" 은 ClustalW 분석 (version W 1.8, European Bioinformatics Institute, Cambridge, UK)에 의해, 정렬 내에서 동일한 매치 (match)의 총수를 계산하고 동일한 매치의 이러한 총수를 (i) 정렬된 서열의 길이, 그리고 (ii) 96 중에서 더 큰 것으로 나누고, 그리고 느린/정확한 쌍별 정렬 (pairwise alignment)을 달성하기 위한 아래의 디폴트 ClustalW 파라미터 -- 갭 시작 페널티 (Gap Open Penalty): 10; 갭 연장 페널티 (Gap Extension Penalty): 0.10; 단백질 중량 매트릭스: Gonnet 시리즈; DNA 중량 매트릭스: IUB; Toggle 느린/빠른 쌍별 정렬=SLOW 또는 FULL Alignment를 이용함으로써 획득되는 아미노산 서열 동일성 비율을 지칭할 수 있다.

[0051] 2개의 서열은 서열 쌍에 대한 가능한 최대 스코어에 도달하기 위하여 그들이 정의된 아미노산 치환 매트릭스

(가령, BLOSUM62), 갭 존재 페널티 (gap existence penalty)와 갭 연장 페널티 (gap extension penalty)를 이용한 유사성 채점 (similarity scoring)을 위하여 정렬될 때, "최적으로 정렬된다". 아미노산 치환 매트릭스 및 2개의 서열 간에 유사성의 정량에서 이들의 이용은 당분야에 널리 공지되어 있고, 예로써 Dayhoff *et al.* (1978) "A model of evolutionary change in proteins" in "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff*et al.* (1992) Proc. Nat'l. Acad. Sci. USA 89: 10915-10919에서 기술된다. BLOSUM62 매트릭스 (도 10)는 서열 정렬 프로토콜, 예를 들면, Gapped BLAST 2.0에서 디폴트 채점 치환 매트릭스로서 종종 이용된다. 갭 존재 페널티는 정렬된 서열 중에서 하나에서 단일 아미노산 갭의 도입을 위하여 강제되고, 그리고 갭 연장 페널티는 이미 열린 갭 내로 삽입된 각 추가의 빈 아미노산 위치에 대하여 강제된다. 정렬은 이러한 정렬이 시작되고 끝나는 각 서열의 아미노산 위치에 의해, 그리고 선택적으로, 최대 가능 스코어에 도달하기 위한 한쪽 또는 양쪽 서열 내에 단일 갭 또는 복수 갭의 삽입에 의해 정의된다. 최적 정렬과 채점이 수동으로 달성될 수 있긴 하지만, 이러한 과정은 Altschul *et al.* (1997) Nucl. Acids Res. 25: 3389-3402에서 기술되고, 그리고 the National Center for Biotechnology Information (NCBI) Website ()에서 공개적으로 가용한 컴퓨터-기반 정렬 알고리즘, 예를 들면, gapped BLAST 2.0의 이용에 의해 조장된다. 복수 정렬을 비롯한 최적 정렬은 예로써, NCB1 웹사이트를 통해 가용하고, 그리고 Altschul *et al.* (1997) Nucl. Acids Res. 25:3389-3402에서 기술된 PSI-BLAST를 이용하여 준비될 수 있다.

[0052] 참고 서열과 최적으로 정렬되는 아미노산 서열에 대하여, 아미노산 잔기는 상기 잔기가 정렬 내에서 쌍을 이루는 참고 서열 내에 위치에 "대응한다." "위치"는 N-말단에 상대적인 위치에 기초된, 참고 서열 내에 각 아미노산을 순차적으로 확인하는 번호에 의해 표시된다. 최적 정렬을 결정할 때 고려되어야 하는 결실, 삽입, 절두, 융합 등으로 인하여, 일반적으로, N-말단으로부터 단순히 계산함으로써 결정된 검사 서열에서 아미노산 잔기 번호는 참고 서열에서 상응하는 위치의 번호와 반드시 동일하지는 않을 것이다. 가령, 정렬된 검사 서열 내에 결실이 존재하는 경우에, 결실 부위에는 참고 서열에서 위치에 대응하는 아미노산이 존재하지 않을 것이다. 정렬된 참고 서열 내에 삽입이 존재하는 경우에, 이러한 삽입은 참고 서열에서 어떤 아미노산 위치에도 대응하지 않을 것이다. 절두 또는 융합의 경우에, 상응하는 서열에서 어떤 아미노산에도 대응하지 않는, 참고 서열 또는 정렬된 서열 내에 일련의 아미노산이 존재할 수 있다.

[0053] 특정 폴리펩티드의 비-보존성 변형은 보존성 치환으로서 특징되지 않는 임의의 아미노산을 치환하는 변형, 예를 들면, 앞서 설명된 6가지 군의 경계를 뛰어넘는 임의의 치환이다. 이들에는 중성 아미노산에 대한 염기성 또는 산성 아미노산의 치환 (가령, Val, Ile, Leu 또는 Met에 대한 Asp, Glu, Asn, 또는 Gln), 염기성 또는 산성 아미노산에 대한 방향족 아미노산 (가령, Asp, Asn, Glu 또는 Gln에 대한 Phe, Tyr 또는 Trp), 또는 아미노산을 유사한 아미노산으로 대체하지 않는 임의의 다른 치환이 포함된다. 염기성 측쇄에는 리신 (K), 아르기닌 (R), 히스티딘 (H)이 포함된다; 산성 측쇄에는 아스파르트산 (D), 글루타민산 (E)이 포함된다; 하전되지 않은 극성 측쇄에는 글리신 (G), 아스파라긴(N), 글루타민(Q), 세린 (S), 트레오닌 (T), 티로신 (Y), 시스테인 (C)이 포함된다; 비-극성 측쇄에는 알라닌 (A), 발린 (V), 류신 (L), 이소류신 (I), 프롤린 (P), 페닐알라닌 (F), 메티오닌 (M), 트립토판 (W)이 포함된다; 베타-가지화 측쇄에는 트레오닌 (T), 발린 (V), 이소류신 (I)이 포함된다; 방향족 측쇄에는 티로신 (Y), 페닐알라닌 (F), 트립토판 (W), 히스티딘 (H)이 포함된다.

[0054] 따라서 폴리펩티드 내에 특정 위치에서 일부 아미노산 잔기는 보존성 아미노산 치환으로부터 "배제된다." 가령, 본 발명에서는 SEQ ID NO: 4에서 열거된 바와 같은 서열을 포함하는 폴리펩티드를 제시한다. 일부 구체예에서, 상기 폴리펩티드는 앞서 기술된 바와 같은 보존성 아미노산 치환으로 변경될 수 있지만, 일정한 잔기, 예를 들면, 위치 23, 108과 140에서 잔기는 가급적 치환되지 않는다.

[0055] "부모" 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포는 임의의 방법, 도구 또는 기술을 이용하고, 그리고 부모가 그 자체로 고유 또는 돌연변이인 지에 상관없이, 임의의 다른 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포가 유래되거나 만들어지는 임의의 단백질, 효소, 폴리뉴클레오티드, 유전자, 또는 세포이다. 부모 폴리뉴클레오티드 또는 유전자는 부모 단백질 또는 효소를 인코딩한다.

[0056] 폴리뉴클레오티드, 폴리펩티드, 또는 다른 성분은 정상적으로 연관되는 성분 (다른 단백질, 핵산, 세포, 합성 시약 등)으로부터 부분적으로 또는 완전하게 분리될 때 "분리된다." 핵산 또는 폴리펩티드는 인공이거나 조작될 때, 또는 인공 또는 조작된 단백질 또는 핵산으로부터 유래될 때 "재조합"이다. 가령, 자연에서 관찰되는 바와 같이 정상적으로 폴리뉴클레오티드에 접하는 뉴클레오티드 서열과 연관되지 않도록, 벡터 내로, 또는 예로써, 재조합 생물체의 게놈 내에 임의의 다른 이종성 위치 내로 삽입되는 폴리뉴클레오티드는 재조합 폴리뉴클레오티드이다. 재조합 폴리뉴클레오티드로부터 시험관내에서 또는 생체내에서 발현된 단백질은 재조합 폴리펩티드의

실례이다. 유사하게, 자연에서 나타나지 않는 폴리뉴클레오티드 서열, 예를 들면, 자연 발생 유전자의 변이체는 재조합이다.

[0057] 다른 구체예에서, 본 발명의 시토신 디아미나아제 폴리펩티드 또는 융합 구조체를 인코딩하는 분리된 폴리뉴클레오티드가 제시된다. 한 측면에서, 본 발명에서는 본 명세서에서 "CD 최적화된 폴리뉴클레오티드", "CD 폴리뉴클레오티드" 또는 "CD-융합 폴리뉴클레오티드"로 지칭되는 분리된 또는 재조합 폴리뉴클레오티드를 제시한다. 용어 "폴리뉴클레오티드", "뉴클레오티드 서열", 그리고 "핵산 분자"는 관련된 배경에 따라, 뉴클레오티드 (A, C, T, U, G 등, 또는 자연 발생 또는 인공 뉴클레오티드 유사체), 예를 들면, DNA 또는 RNA의 중합체, 또는 이들의 표현, 예를 들면, 문자 열 (character string) 등을 지칭하는데 이용된다. 소정의 폴리뉴클레오티드 또는 상보성 폴리뉴클레오티드는 임의의 특정된 뉴클레오티드 서열로부터 결정될 수 있다.

[0058] 본 발명의 한 구체예는 시토신 디아민 또는 돌연변이 시토신 디아미나아제 폴리펩티드, 또는 이들의 생물학적으로 활성 부분을 인코딩하는 분리된 폴리뉴클레오티드에 관계한다. 본 명세서에서, 용어 "핵산 분자" 또는 "폴리뉴클레오티드"는 DNA 분자 (가령, cDNA 또는 게놈 DNA)와 RNA 분자 (가령, mRNA), 그리고 뉴클레오티드 유사체를 이용하여 산출된 DNA 또는 RNA의 유사체를 포함하는 것으로 의도된다. 핵산 분자는 단일-가닥 또는 이중-가닥일 수 있다.

[0059] 한 구체예에서, CD 최적화된 폴리뉴클레오티드 또는 CD 폴리뉴클레오티드는 생물체, 예를 들면, 세균 또는 효모 균주로부터 분리된 자연 발생 핵산의 재조합, 조작된 또는 분리된 형태를 포함한다. 전형적인 CD 폴리뉴클레오티드에는 SEQ ID NO: 2에 열거된 야생형 폴리펩티드를 인코딩하는 것들이 포함된다. 본 발명의 다른 구체예에서, CD 폴리뉴클레오티드는 하나 이상의 자연 발생, 분리된, 또는 재조합 CD 폴리뉴클레오티드를 다양화시킴으로써, 예를 들면, 재조합하고 및/또는 돌연변이시킴으로써 생산된다. 본 명세서의 다른 곳에서 더욱 상세하게 기술된 바와 같이, 다양한 과정에서 기질 또는 부모로서 이용되는 CD 폴리뉴클레오티드보다 우수한 기능적 속성, 예를 들면, 증가된 촉매 기능, 증가된 안정성, 또는 더욱 높은 발현 수준을 갖는 CD 폴리펩티드를 인코딩하는 다양화된 CD 폴리뉴클레오티드를 산출하는 것이 가능하다. 전형적인 폴리뉴클레오티드는 SEQ ID NO: 3과 5를 포함하고, 그리고 여기에는 본 발명의 CD 변이체 폴리펩티드를 인코딩하는 것들이 포함된다.

[0060] 본 발명의 폴리뉴클레오티드는 예로써, 본 발명의 CD 폴리펩티드의 재조합 생산 (즉, 발현)에서, 그리고 새로운 및/또는 향상된 CD 동족체 등을 생산하는 진전된 다양성 산출, 예를 들면, 재조합 반응 또는 돌연변이 반응을 위한 기질로서 다양한 용도를 갖는다.

[0061] 중요하게는, CD 폴리뉴클레오티드의 일정한 명확하고, 실질적이고, 신뢰성 있는 유용성은 상기 폴리뉴클레오티드가 실질적인 CD 활성 또는 심지어 상이한 CD 활성을 갖는 폴리펩티드를 인코딩하도록 요구하지 않는다. 가령, 활성 효소를 인코딩하지 않는 CD 폴리뉴클레오티드는 원하는 기능적 특성 (가령, 높은 k_{cat} 또는 k_{cat}/K_m, 낮은 K_m, 열 또는 다른 환경 인자에 대한 높은 안정성, 높은 전사 또는 번역 속도, 단백분해 절단에 대한 내성 등)을 갖는 CD 폴리뉴클레오티드 변이체, 또는 비-CD 폴리뉴클레오티드에 도달하기 위한 다양화 절차에 이용되는 부모 폴리뉴클레오티드의 귀중한 공급원일 수 있다.

[0062] CD 폴리펩티드와 이들의 변이체, CD 폴리펩티드의 단편, 관련된 융합 단백질, 또는 이들의 기능적 등가물을 인코딩하는 뉴클레오티드 서열을 비롯한 CD 폴리뉴클레오티드는 적절한 숙주 세포, 예를 들면, 세균 세포에서 이들 CD 폴리펩티드의 발현을 주동하는 재조합 DNA 분자에 이용된다. 유전자 코드의 본질적인 축중성으로 인하여, 실질적으로 동일하거나 또는 기능적으로 등가인 아미노산 서열을 인코딩하는 다른 핵산 서열 역시 CD 폴리뉴클레오티드를 클로닝하고 발현하는데 이용될 수 있다. 본 명세서에서, 용어 "숙주 세포"는 핵산 구조체로 형질전환 (transformation)될 수 있는 임의의 세포 유형을 포함한다. 용어 "형질전환"은 외래 (즉, 외부 또는 세포외) 유전자, DNA 또는 RNA 서열의 숙주 세포로의 도입을 의미하고, 여기서 숙주 세포는 도입된 유전자 또는 서열을 발현하여 원하는 물질, 전형적으로 상기 도입된 유전자 또는 서열에 의해 코딩된 단백질 또는 효소를 생산할 것이다. 도입된 유전자 또는 서열은 조절 또는 제어 서열, 예를 들면, 출발, 중단, 프로모터, 신호, 분비, 또는 세포의 유전자 기구 (genetic machinery)에 의해 사용되는 기타 서열을 포함할 수 있다. 도입된 DNA 또는 RNA를 수용하고 발현하는 숙주 세포는 "형질전환되고", 그리고 "형질전환체" 또는 "클론"이다. 숙주 세포에 도입되는 DNA 또는 RNA는 숙주 세포와 동일한 속 (genus) 또는 종 (species)의 세포, 또는 상이한 속 또는 종의 세포를 비롯한 임의의 공급원으로부터 제공될 수 있다.

[0063] 당업자가 인지하는 바와 같이, 특정 숙주 내에서 발현을 강화시키기 위하여 코딩 서열을 변형하는 것이 유리할 수 있다. 유전자 코드는 64개의 가능한 코돈으로 중복되지만, 대부분의 생물체는 이들 코돈의 부분집합을 선호

적으로 사용한다. 소정의 종에서 가장 빈번하게 사용되는 코돈은 최적 코돈으로 불리고, 그리고 빈번하게 사용되지 않는 코돈은 드문 또는 낮은 사용빈도 코돈으로 분류된다 (참조: Zhang *et al.* (1991) Gene 105:61-72). 코돈은 숙주의 선호되는 코돈 사용빈도 (codon usage)를 반영하기 위하여 치환될 수 있는데, 이러한 과정은 때때로, "코돈 최적화" 또는 "종 코돈 편향을 위한 제어"라고 불린다.

[0064] 특정 원핵 또는 진핵 숙주에 의해 선호되는 코돈을 포함하는 최적화된 코딩 서열 (참조: Murray *et al.* (1989) Nucl. Acids Res. 17:477-508)은 예로써, 비-최적화된 서열로부터 생산된 전사체와 비교하여, 번역의 속도를 증가시키기 위하여, 또는 원하는 특성, 예를 들면, 더욱 긴 반감기를 갖는 재조합 RNA 전사체를 생산하기 위하여 준비될 수 있다. 번역 종결 코돈 역시 숙주 선호 (host preference)를 반영하기 위하여 변형될 수 있다. 가령, 사카로마이세스 세레비지애 (*S. cerevisiae*)와 포유동물에 대한 선호되는 종결 코돈은 각각, UAA와 UGA이다. 외떡잎식물 (monocotyledonous plant)에 대한 선호되는 종결 코돈은 UGA인 반면, 곤충과 대장균 (*E. coli*)은 종결 코돈으로서 UAA의 사용을 선호한다 (Dalphin *et al.* (1996) Nucl. Acids Res. 24: 216-218).

[0065] 코돈 사용빈도 편향 (codon usage bias)은 단백질-코딩 DNA 서열 (유전자)에서 코돈의 발생 빈도에서 생물체 간에 차이를 지칭한다. 코돈은 폴리펩티드 사슬 내에서 특정 아미노산 잔기를 인코딩하는 일련의 3개의 뉴클레오티드 (트리플렛)이다. DNA 내에 4가지 뉴클레오티드, 아데닌 (A), 구아닌 (G), 시토신 (C), 그리고 티민 (T)이 존재하기 때문에, 20개의 아미노산을 인코딩하는 64개의 가능한 트리플렛, 그리고 3개의 번역 정지 (nonsense) 코돈이 존재한다. 이러한 중복성 (redundancy)으로 인하여, 2가지 아미노산을 제외한 모든 아미노산은 1개 이상의 트리플렛에 의해 인코딩된다. 상이한 생물체는 종종, 동일한 아미노산을 인코딩하는 여러 코돈 중에서 한 가지에 대한 특별한 선호를 보인다. 이들 선호가 어떻게 발생하는 지는 많은 논의가 진행되고 있는 분자 진화의 분야이다.

[0066] 코돈 선호는 돌연변이 편향과 번역 최적화를 위한 자연 선택 (natural selection) 사이에 균형을 반영하는 것으로 일반적으로 인식되고 있다. 빠르게 성장하는 미생물, 예를 들면, 대장균 (*Escherichia coli*) 또는 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)(빵 효모)에서 최적 코돈은 그들의 개별 게놈 tRNA 풀(pool)의 조성을 반영한다. 최적 코돈은 더욱 빠른 번역 속도와 높은 정확도를 달성하는데 도움을 주는 것으로 생각된다. 이들 인자의 결과로써, 번역 선택은 앞서-언급된 생물체의 경우에서처럼, 고도로 발현된 유전자에서 더욱 강할 것으로 예상된다. 높은 성장 속도를 보이지 않거나, 또는 작은 게놈을 제공하는 다른 생물체에서, 코돈 사용빈도 최적화는 정상적으로 부재하고, 그리고 코돈 선호는 상기 특정 게놈에서 관찰되는 특징적인 돌연변이 편향에 의해 결정된다. 이의 실례는 호모 사피엔스 (*Homo sapiens*) (인간)와 헬리코박터 파일로리 (*Helicobacter pylori*)이다. 중간 수준의 코돈 사용빈도 최적화를 보이는 생물체에는 노랑초파리 (*Drosophila melanogaster*) (fruit fly), 예쁜 꼬마 선충 (*Caenorhabditis elegans*) (nematode worm) 또는 애기장대 (*Arabidopsis thaliana*) (wall cress)가 포함된다.

[0067] 용어 "코돈 최적화된 서열"은 일반적으로, 대략 20% 이하의 사용빈도를 갖는 임의의 코돈을 대체함으로써 특정 숙주 종에 대하여 최적화된 뉴클레오티드 서열을 지칭한다. 의사 폴리아데닐화 서열의 제거, 엑손/인트론 접합 신호의 제거, 트랜스포손 (transposon)-유사 반복의 제거 및/또는 코돈 최적화에 더하여 GC 함량의 최적화에 의해 소정의 숙주 종에서 발현을 위하여 최적화된 뉴클레오티드 서열은 본 명세서에서, "발현 강화된 서열"로 지칭된다.

표 1

[0068] 인간 코돈 사용빈도와 코돈 선호.

인간 코돈 사용빈도															
Gly	GGG	17.08	0.23	Arg	AGG	12.09	0.22	Trp	TGG	14.74	1.00	Arg	CGG	10.40	0.19
Gly	GGA	19.31	0.26	Arg	AGA	11.73	0.21	End	TGA	2.64	0.61	Arg	CGA	5.63	0.10
Gly	GGT	13.66	0.18	Ser	AGT	10.18	0.14	Cys	TGT	9.99	0.42	Arg	CGT	5.16	0.09
Gly	GGC	24.94	0.33	Ser	AGC	18.54	0.25	Cys	TGC	13.86	0.58	Arg	CGC	10.82	0.19
Glu	GAG	38.82	0.59	Lys	AAG	33.79	0.60	End	TAG	0.73	0.17	Gln	CAG	32.95	0.73
Glu	GAA	27.51	0.41	Lys	AAA	22.32	0.40	End	TAA	0.95	0.22	Gln	CAA	11.94	0.27
Asp	GAT	21.45	0.44	Asn	AAT	16.43	0.44	Tyr	TAT	11.80	0.42	His	CAT	9.56	0.41
Asp	GAC	27.06	0.56	Asn	AAC	21.30	0.56	Tyr	TAC	16.48	0.58	His	CAC	14.00	0.59
Val	GTG	28.60	0.48	Met	ATG	21.86	1.00	Leu	TTG	11.43	0.12	Leu	CTG	39.93	0.43
Val	GTA	6.09	0.10	Ile	ATA	6.05	0.14	Leu	TTA	5.55	0.06	Leu	CTA	6.42	0.07

Val	GTT	10.30	0.17	Ile	ATT	15.03	0.35	Phe	TTT	15.36	0.43	Leu	CTT	11.24	0.12
Val	GTC	15.01	0.25	Ile	ATC	22.47	0.52	Phe	TTC	20.72	0.57	Leu	CTC	19.14	0.20
Ala	GCG	7.27	0.10	Thr	ACG	6.80	0.12	Ser	TCG	4.38	0.06	Pro	CCG	7.02	0.11
Ala	GCA	15.50	0.22	Thr	ACA	15.04	0.27	Ser	TCA	10.96	0.15	Pro	CCA	17.11	0.27
Ala	GCT	20.23	0.28	Thr	ACT	13.24	0.23	Ser	TCT	13.51	0.18	Pro	CCT	18.03	0.29
Ala	GCC	28.43	0.40	Thr	ACC	21.52	0.38	Ser	TCC	17.37	0.23	Pro	CCC	20.51	0.33

[0069] 각 코돈에 대하여, 상기 표는 인간 코딩 영역 내에서 각 코돈 (1000개당)의 사용빈도 (첫 번째 칼럼), 그리고 동의적 코돈 사이에 각 코돈의 상대적 빈도 (두 번째 칼럼)를 제시한다.

[0071] 따라서 일부 구체예에서, 본 발명의 폴리뉴클레오티드는 인간 세포에서 번역을 위하여 최적화된 분자 코돈을 포함한다. 가령, SEQ ID NO: 3과 5는 인간 숙주 세포에서 시토신 디아미나아제를 생산하기 위하여 최적화된 서열을 포함한다.

[0072] 이런 이유로, 본 발명에서는 시토신 디아미나아제 활성을 갖는 폴리펩티드를 인코딩하는 인간 코돈 최적화된 폴리뉴클레오티드를 제시한다. 본 발명의 인간 코돈 최적화된 폴리뉴클레오티드 (가령, SEQ ID NO: 5)는 보존성 아미노산 치환 및/또는 인코딩된 폴리펩티드의 향상된 활성 또는 안정성을 결과하는 부가적인 돌연변이를 더욱 보유할 수 있다. 가령, SEQ ID NO: 6을 포함하는 폴리펩티드의 위치 23, 108과 140에서 돌연변이는 증가된 열안정성을 제공한다. 따라서 본 발명에서는 시토신 디아미나아제 활성 및 증가된 열안정성을 갖는 폴리펩티드를 인코딩하는 인간 코돈 최적화된 폴리뉴클레오티드를 제시한다. 한 구체예에서, 인간 코돈 최적화된 폴리뉴클레오티드는 SEQ ID NO: 4의 폴리펩티드를 인코딩하는 서열을 포함하고, 여기서 상기 폴리펩티드의 아미노산을 인코딩하는 코돈은 인간 세포에서 발현을 위하여 최적화된다. 다른 구체예에서, 본 발명의 폴리뉴클레오티드는 SEQ ID NO: 3 또는 5를 포함한다.

[0073] 한 구체예에서, 본 발명에서는 UPRT 또는 OPRT 활성을 갖는 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드에 연결된 시토신 디아미나아제 폴리뉴클레오티드 또는 코돈 최적화된 폴리뉴클레오티드 또는 돌연변이체를 포함하는 폴리뉴클레오티드를 제시한다.

[0074] 예로써, SEQ ID NO: 4, 8, 10, 12, 14, 16, 또는 18의 폴리펩티드를 인코딩하는; 또는 SEQ ID NO: 3, 5, 7, 9, 11, 13, 15, 또는 17 중에서 한 가지에 열거되는 뉴클레오티드 서열, 또는 이의 일부분을 갖는 서열을 포함하는 본 발명의 폴리뉴클레오티드는 표준 분자 생물학 기술 및 본 명세서에 제공된 서열 정보를 이용하여 분리되고 산출될 수 있다.

[0075] 본 발명의 폴리뉴클레오티드는 표준 PCR 증폭 기술에 따라, 주형으로서 cDNA, mRNA 또는 대안으로, 게놈 DNA, 그리고 적절한 올리고뉴클레오티드 프라이머를 이용하여 증폭될 수 있다. 이렇게 증폭된 핵산은 적절한 벡터 내로 클로닝되고 DNA 서열 분석에 의해 특징될 수 있다. 더 나아가, 뉴클레오티드 서열에 상응하는 올리고뉴클레오티드는 예로써, 자동화된 DNA 합성기를 이용한 표준 합성 기술에 의해 제조될 수 있다. 일부 구체예에서, 본 발명의 분리된 핵산 분자는 SEQ NO:2, 4, 6, 8, 10, 12, 14, 16, 또는 18 중에서 한 가지에서 열거된 폴리펩티드를 인코딩하는, 또는 SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 또는 17 중에서 한 가지에서 열거되는 뉴클레오티드 서열을 갖는 뉴클레오티드 서열의 보체 (complement)인 핵산 분자를 포함한다.

[0076] 다른 구체예에서, 본 발명의 분리된 폴리뉴클레오티드는 SEQ NO:4, 12, 14, 16 또는 18 중에서 한 가지에서 열거된 폴리펩티드를 인코딩하는 뉴클레오티드 서열로 구성되거나, 또는 SEQ ID NO: 3, 5, 11, 13, 15, 또는 17 중에서 한 가지에서 열거된 뉴클레오티드 서열로 구성되는 핵산 분자에 엄밀한 조건 하에 혼성화하는 핵산 분자를 포함한다. 다른 구체예에서, 핵산은 최소한 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 뉴클레오티드 길이를 갖는다. 핵산 분자는 한 폴리뉴클레오티드의 최소한 하나의 가닥이 정의된 엄밀도 조건 하에 다른 폴리뉴클레오티드에 어닐링될 수 있을 때, 서로에 "혼성화가능"하다. 혼성화의 엄밀도는 예로써, (a) 혼성화 및/또는 세척이 수행되는 온도, 그리고 (b) 혼성화와 세척 용액의 이온 강도와 극성 (가령, 포름아미드) 등과 같은 파라미터에 의해 결정된다. 혼성화는 2개의 폴리뉴클레오티드가 실질적으로 상보성 서열을 포함하도록 요구한다; 하지만, 혼성화의 엄밀도에 따라, 미스매치가 관용될 수도 있다. 전형적으로, 높은 엄밀도 (가령, 65℃에서 0.5 X SSC의 수성 용액에서)에서 2개의 서열의 혼성화는 이들 서열이 그들의 전체 서열에 걸쳐 상당히 높은 상보성을 나타내도록 요구한다. 중간 엄밀도 (가령, 65℃에서 2 X SSC의 수성 용액) 및 낮은 엄밀도 (가령, 55℃에서 2 X SSC의 수성 용액)의 조건은 혼성화 서열 간에 전체적으로 부합하게 더욱 낮은 상보성을 요구한다 (1 X SSC는 0.15 M NaCl, 0.015 M Na 구연산염). 혼성화하는 핵산 분자에는 적절한 엄밀도 조건

하에 어닐링되고, 그리고 동일한 기능, 예를 들면, 5-플루시토신 (5-FC)의 5-플루오르우라실 (5-FU)로의 전환을 촉진하는 능력을 갖는 본 발명의 폴리펩티드 또는 효소를 인코딩하는 것들이 포함된다. 게다가, 용어 "엄밀한 조건 하에 혼성화한다"는 서로에 최소한 30%, 40%, 50%, 또는 60% 상동한 뉴클레오티드 서열이 전형적으로, 서로에 혼성화된 상태로 남아있는 혼성화와 세척을 위한 조건을 기술하는 것으로 의도된다. 바람직하게는, 이들 조건은 서로에 최소한 대략 70%, 더욱 바람직하게는 최소한 대략 80%, 이보다 더욱 바람직하게는 최소한 대략 85% 또는 90% 상동한 서열이 전형적으로, 서로에 혼성화된 상태로 남아있도록 한다. 일부 경우에, 본 발명의 분리된 핵산 분자는 SEQ ID NO: 4, 12, 14, 16, 또는 18 중에서 한 가지에서 열거된 폴리펩티드를 인코딩하는, 또는 SEQ ID NO: 3, 5, 11, 13, 15, 또는 17 중에서 한 가지에서 열거되는 뉴클레오티드 서열을 갖는 핵산 서열에 엄밀한 조건 하에 혼성화한다. 본 명세서에서, "자연-발생" 핵산 분자는 자연에서 발생하는 (가령, 자연 단백질을 인코딩하는) 뉴클레오티드 서열을 갖는 RNA 또는 DNA 분자를 지칭한다.

[0077] 당업자가 인지하는 바와 같이, 변화는 SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 또는 18 중에서 한 가지에서 열거된 폴리펩티드를 인코딩하는, 또는 SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 또는 17 중에서 한 가지에서 열거되는 뉴클레오티드 서열을 갖는 임의의 폴리뉴클레오티드의 뉴클레오티드 서열 내로 돌연변이에 의해 도입될 수 있고, 따라서 이들 인코딩된 단백질의 아미노산 서열에서 변화를 유발한다. 일부 경우에, 이러한 변화는 폴리펩티드의 변경된 기능을 유발할 것이다. 다른 경우에, 이러한 변화는 인코딩된 폴리펩티드의 기능적 능력을 변경하지 않을 것이다. 일반적으로, 폴리펩티드의 기능을 변경하지 않는 치환에는 "비-필수" 아미노산 잔기에서 아미노산 치환을 유발하는 뉴클레오티드 치환이 포함된다. "비-필수" 아미노산 잔기는 예로써, 5-FC의 5-FU로의 전환을 촉진함으로써, 생성된 폴리펩티드의 생물학적 활성을 변경하지 않으면서 부모 서열로부터 변경될 수 있는 잔기이다.

[0078] 또한, 폴리펩티드가 특정 기질에서 새로운 또는 증가된 활성, 또는 증가된 안정성 또는 감소된 분해를 갖도록 부모 폴리펩티드의 활성을 변경하는 것이 바람직한 상황이 고려된다. 이들 아미노산 치환은 일반적으로, "보존성" 치환을 구성하지 않는 것으로 이해된다. 실제로, 이들 치환은 새로운 또는 향상된 활성을 획득하기 위하여 서열 내로 도입된 비-보존성 치환을 구성한다. 가령, SEQ ID NO: 1에서는 사카로마이세스 세레비지애 (*S. cervisae*) 시토신 디아미나아제에 대한 부모 핵산 서열을 제공한다 (SEQ ID NO: 2는 상응하는 폴리펩티드를 제공한다). SEQ ID NO: 3에서는 폴리펩티드에 증가된 안정성을 부여하는 아미노산 치환을 보유하는 돌연변이 서열의 핵산 서열을 제공한다. 따라서 SEQ ID NO: 2의 아미노산 서열을 인코딩하는 핵산 분자는 아미노산 치환을 보유하는 변형된 폴리펩티드를 인코딩하는 핵산 분자를 획득하기 위하여 돌연변이가 만들어질 수 있는 "부모" 핵산 분자를 제공한다.

[0079] 또한, 변형된 폴리펩티드는 부가적인 치환이 만들어질 수 있는 "부모" 폴리펩티드를 구성할 수 있는 것으로 이해된다. 따라서 부모 폴리펩티드, 그리고 부모 폴리펩티드를 인코딩하는 핵산 분자는 변형된 폴리펩티드를 포함하고, 오로지 "야생형" 서열인 것만은 아니다. 가령, SEQ ID NO: 5의 폴리뉴클레오티드는 SEQ ID NO: 1 (즉, "부모" 폴리뉴클레오티드)에 대하여 변형된 폴리뉴클레오티드이다. 유사하게, SEQ ID NO: 3의 폴리뉴클레오티드는 SEQ ID NO: 5에 대하여 변형된 폴리뉴클레오티드이다. 따라서 SEQ ID NO: 5는 SEQ ID NO: 3의 부모 서열이다.

[0080] 다양성을 산출하는 돌연변이 방법에는 예로써, 부위-특이적 돌연변이유발 (site-directed mutagenesis) (Ling *et al.* (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale *et al.* (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol. Biol. 57:369-374; Smith (1985) "*In vitro* mutagenesis mutagenesis" Ann. Rev. Genet. 19:423-462; Botstein & Shortle (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; Carter (1986) "Site-directed mutagenesis" Biochem. J. 237:1-7; 그리고 Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); 우라실 보유 주형을 이용한 돌연변이유발 (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel *et al.* (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Methods in Enzymol. 154, 367-382; 그리고 Bass *et al.* (1988) "Mutant Trp repressors with new DNA-binding specificities" Science 242:240-245); 올리고뉴클레오티드-특이적 돌연변이유발 (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" Nucleic Acids Res. 10:6487-6500; Zoller &

Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" Methods in Enzymol. 100:468-500; 그리고 Zoller & Smith (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" Methods in Enzymol. 154:329-350); 포스포로티오에이트-변형된 DNA 돌연변이유발 (Taylor *et al.* (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" Nucl. Acids Res. 13: 8749-8764; Taylor *et al.* (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucl. Acids Res. 13: 8765-8787; Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers *et al.* (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" Nucl. Acids Res. 16:791-802; 그리고 Sayers *et al.* (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucl. Acids Res. 16: 803-814); 갭 (gap)이 있는 이중나선 DNA를 이용한 돌연변이유발 (Kramer *et al.* (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer *et al.* (1988) "Improved enzymatic *in vitro* reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; 그리고 Fritz *et al.* (1988) "oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*" Nucl. Acids Res. 16: 6987-6999)이 포함된다.

[0081] 부가적인 적합한 방법에는 점 미스매치 회복 (Kramer *et al.* (1984) "Point Mismatch Repair" Cell 38:879-887), 회복-결함성 숙주 균주를 이용한 돌연변이유발 (Carter *et al.* (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; 그리고 Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), 결실 돌연변이유발 (Eghtedarzadeh & Henikoff (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), 제한-선택과 제한-정제 (Wells *et al.* (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), 전체 유전자 합성에 의한 돌연변이유발 (Nambiar *et al.* (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14: 6361-6372; Wells *et al.* (1985) "Casette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; 그리고 Grundstrom *et al.* (1985) "Oligonucleotide-directed mutagenesis by microscale `shot-gun` gene synthesis" Nucl. Acids Res. 13: 3305-3316); 이중-가닥 파괴 회복 (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455; 그리고 "Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181)이 포함된다. 이들 방법에 대한 추가적인 상세는 Methods in Enzymology Volume 154에서 확인될 수 있는데, 상기 문헌은 또한, 다양한 돌연변이유발 방법으로 트러블-슈팅 (trouble-shooting) 문제점에 대한 유용한 제어를 기술한다.

[0082] 여러 가지 다양성 산출 방법에 관한 부가적인 상세는 아래의 U.S. 특허, PCT 공보, 그리고 EPO 공보에서 확인될 수 있다: U.S. Pat. No. 5,605,793 (Stemmer, (Feb. 25, 1997), "Methods for *In vitro* Recombination;" U.S. Pat. No. 5,811,238 (Stemmer *et al.,* Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 (Stemmer *et al.,* Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, *et al.* (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 (Minshull, *et al.,* Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207, Stemmer and Lipschutz, "End Complementary Polymerase Chain Reaction;" WO 97/20078, Stemmer and Crameri, "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966, Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402, Punnonen *et al.* "Targeting of Genetic Vaccine Vectors;" WO

99/41383, Punnonen *et al.*, "Antigen Library Immunization;" WO 99/41369, Punnonen *et al.*, "Genetic Vaccine Vector Engineering;" WO 99/41368, Punnonen *et al.*, "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008, Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670, Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107, Stemmer *et al.*, "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979, Apt *et al.*, "Human Papillomavirus Vectors;" WO 98/31837, del Cardayre *et al.*, "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230, Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/13487, Stemmer *et al.*, "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection;" WO 00/00632, "Methods for Generating Highly Diverse Libraries;" WO 00/09679, "Methods for Obtaining *in vitro* Recombined Polynucleotide Sequence Banks and Resulting Sequences;" WO 98/42832, Arnold *et al.*, "Recombination of Polynucleotide Sequences Using Random or Defined Primers;" WO 99/29902, Arnold *et al.*, "Method for Creating Polynucleotide and Polypeptide Sequences;" WO 98/41653, Vind, "An *in vitro* Method for Construction of a DNA Library;" WO 98/41622, Borchert *et al.*, "Method for Constructing a Library Using DNA Shuffling;" WO 98/42727, Pati and Zarling, "Sequence Alterations using Homologous Recombination;" WO 00/18906, Patten *et al.*, "Shuffling of Codon-Altered Genes;" WO 00/04190, del Cardayre *et al.* "Evolution of Whole Cells and Organisms by Recursive Recombination;" WO 00/42561, Crameri *et al.*, "Oligonucleotide Mediated Nucleic Acid Recombination;" WO 00/42559, Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations;" WO 00/42560, Selifonov *et al.*, "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics;" WO 01/23401, Welch *et al.*, "Use of Codon-Varied Oligonucleotide Synthesis for Synthetic Shuffling;" 그리고 WO 01/64864, Affholter, "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation".

[0083] 또한, 앞서 폭넓게 기술된 바와 같은 하나 이상의 핵산 서열을 포함하는 재조합 구조체가 제시된다. 이들 구조체는 본 발명의 폴리뉴클레오티드가 전방 또는 후방 방향으로 삽입되는 벡터, 예를 들면, 플라스미드, 코스미드, 파지, 바이러스, 세균 인공 염색체 (BAC), 효모 인공 염색체 (YAC), 바이러스 벡터 등을 포함한다. 많은 숫자의 적절한 벡터와 프로모터는 당업자에게 공지되어 있고, 그리고 상업적으로 가용하다. 한 구체예에서, 바이러스 벡터는 레트로바이러스 벡터이다.

[0084] 용어 "벡터", "벡터 구조체"와 "발현 벡터"는 숙주를 형질전환시키고 도입된 서열의 발현 (가령, 전사와 번역)을 촉진하기 위하여, DNA 또는 RNA 서열 (가령, 외래 유전자)이 숙주 세포 내로 도입될 수 있는 운반제를 의미한다. 벡터는 전형적으로, 단백질을 인코딩하는 외래 DNA가 제한 효소 기술에 의해 삽입되는 전파가능 병원체 (transmissible agent)의 DNA를 포함한다. 일반적인 유형의 벡터는 "플라스미드"인데, 이는 일반적으로, 부가적인 (외래) DNA를 용이하게 수용할 수 있고 적절한 숙주 세포 내로 용이하게 도입될 수 있는 이중-가닥 DNA의 자기-한정된 분자이다. 다양한 진핵과 원핵 숙주에서 복제 및/또는 발현을 위한, 플라스미드와 진균 벡터를 비롯한 많은 숫자의 벡터가 기술되었다. 무제한적 실례에는 pKK 플라스미드 (Clonetech), pUC 플라스미드, pET 플라스미드 (Novagen, Inc., Madison, Wis.), pRSET 또는 pREP 플라스미드 (Invitrogen, San Diego, Calif.), 또는 pMAL 플라스미드 (New England Biolabs, Beverly, Mass.), 그리고 본 명세서에 개시되거나 언급된 방법, 또는 당업자에게 달리 공지된 방법을 이용한 많은 적절한 숙주 세포가 포함된다. 재조합 클로닝 벡터는 종종, 클로닝 또는 발현을 위한 하나 이상의 복제 시스템, 숙주에서 선별을 위한 하나 이상의 마커, 예를 들면, 항생제 내성, 그리고 하나 이상의 발현 카세트를 포함할 것이다.

[0085] 용어 "발현한다"와 "발현"은 유전자 또는 DNA 서열 내에 정보가 명백해지도록 하는 것을 의미한다, 가령, 상응하는 유전자 또는 DNA 서열의 전사와 번역에 관여하는 세포 기능을 활성화시킴으로써 단백질을 생산하는 것을 의미한다. DNA 서열은 "발현 산물", 예를 들면, 단백질을 형성하기 위하여 세포 내에서 또는 세포에 의해 발현된다. 발현 산물 자체, 예를 들면, 생성 단백질은 또한, 세포에 의해 "발현되는" 것으로 일컬어질 수 있다. 폴리뉴클레오티드 또는 폴리펩티드는 재조합 방식으로, 예를 들면, 외래 또는 고유 프로모터의 제어 하에 외래 숙주 세포에서, 또는 외래 프로모터의 제어 하에 고유 숙주 세포에서 발현되거나 생산될 때, 발현된다.

[0086] 본 발명에 제시된 폴리뉴클레오티드는 폴리펩티드를 발현하는데 적합한 다양한 발현 벡터 중에서 임의의 한 가지 내로 통합될 수 있다. 적절한 벡터에는 염색체, 비-염색체와 합성 DNA 서열, 예를 들면, SV40의 유도체; 세균 플라스미드; 파지 DNA; 베큘로바이러스; 효모 플라스미드; 플라스미드와 파지 DNA, 바이러스 DNA, 예를

들면, 우두, 아데노바이러스, 계두 바이러스, 위광견병 바이러스, 아데노바이러스, 아데노-연관된 바이러스, 레트로바이러스 등의 조합으로부터 유래된 벡터가 포함된다. 유전 물질을 세포 내로 형질도입하고, 그리고 복제가 바람직하면, 관련된 숙주 내에서 복제가능하고 생존하는 임의의 벡터가 이용될 수 있다. 이들 핵산 벡터는 예로써, P.Midoux et al. Brit J Pharm 157: 166-178 2009 또는 Kamimura & Liu J AAPS 10:589-595 (2008)에서 기술된 비-바이러스 전달 시스템, 또는 하기에 기술된 바와 같은 바이러스 시스템을 이용하여 전달될 수 있다.

[0087] 다른 측면에서, 시토신 디아미나아제를 인코딩하는 폴리뉴클레오티드 또는 이의 돌연변이체는 바이러스 또는 바이러스 유래된 벡터인 유전자 전달 시스템 내로 전달된다. 바이러스 벡터는 복제성 또는 비-복제성일 수 있고, 그리고 아데노바이러스 벡터, 홍역 벡터, 포진 벡터, 레트로바이러스 벡터 (렌티바이러스 벡터 포함), 랍도바이러스 벡터, 예를 들면, 수포성 구내염 바이러스 벡터, 레오바이러스 벡터, 세네카 밸리 바이러스 벡터, 폭스바이러스 벡터 (수두 또는 우두 유래된 벡터 포함), 파보바이러스 벡터 (AAV 벡터 포함), 알파바이러스 벡터 또는 당업자에게 공지된 기타 바이러스 벡터일 수 있다 (참조: *Concepts in Genetic Medicine*, ed. Boro Dropulic and Barrie Carter, Wiley, 2008, Hoboken, NJ.; *The Development of Human Gene Therapy*, ed. Theodore Friedmann, Cold Springs Harbor Laboratory Press, Cold springs Harbor, New York, 1999; *Gene and Cell Therapy*, ed. Nancy Smyth Templeton, Marcel Dekker Inc., New York, New York, 2000 and *Gene Therapy: Therapeutic Mechanism and Strategies*, ed. Nancy Smyth Templetone and Danilo D Lasic, Marcel Dekker, Inc., New York, New York, 2000; 이들의 내용은 본 발명에 참조로서 편입됨).

[0088] 한 구체예에서, 바이러스 벡터는 복제성 포유동물 세포만을 감염시킬 수 있는 복제 적격성 레트로바이러스 벡터일 수 있다. 레트로바이러스는 다양한 방법으로 분류되지만 명명법이 지난 10년 동안 표준화되었다 (참조: ICTVdB - The Universal Virus Database, v4, World Wide Web (www)에서 ncbi.nlm.nih.gov/ICTVdb/ICTVdB/, 그리고 교재 "Retroviruses" Eds Coffin, Hughs and Varmus, Cold Spring Harbor Press 1997; 이들의 내용은 본 발명에 참조로서 편입됨). 복제 적격성 레트로바이러스 벡터는 오르토레트로바이러스 또는 더욱 전형적으로, 감마 레트로바이러스 벡터를 포함할 수 있다. 한 측면에서, 복제 적격성 레트로바이러스 벡터는 시토신 디아미나아제를 인코딩하는 폴리뉴클레오티드에 대하여 5'에서 내부 리보솜 결합 부위 (IRES)를 포함한다. 한 구체예에서, 시토신 디아미나아제를 인코딩하는 폴리뉴클레오티드는 레트로바이러스 벡터의 ENV 폴리뉴클레오티드에 대하여 3'에 위치한다.

[0089] 따라서 다른 구체예에서, 본 발명의 폴리뉴클레오티드를 포함하는 벡터가 제시된다. 다른 구체예에서, 본 발명의 핵산 분자, 또는 본 발명의 폴리뉴클레오티드를 포함하는 벡터로 형질감염된 숙주 세포가 제시된다. 숙주 세포에는 진핵 세포, 예를 들면, 효모 세포, 곤충 세포, 또는 동물 세포가 포함된다. 숙주 세포에는 또한, 원핵 세포, 예를 들면, 세균 세포가 포함된다.

[0090] 앞서 논의된 바와 같이, 벡터, 프로모터 및 많은 다른 관련된 원리의 이용을 비롯한 본 발명에 유용한 분자 생물학적 기술을 설명하는 일반적인 문서에는 Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology Volume 152, (Academic Press, Inc., San Diego, Calif.) ("Berger"); Sambrook *et al.*, Molecular Cloning--A Laboratory Manual, 2d ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1999년에 보완됨) ("Ausubel")이 포함된다. 예로써, 본 발명의 상동한 핵산의 생산을 위하여 중합효소 연쇄 반응 (PCR), 리가아제 연쇄 반응 (LCR), Qβ-레플리카아제 증폭 및 기타 RNA 중합효소 매개된 기술 (가령, NASBA)을 비롯한 시험관내에서 증폭 방법을 통하여, 당업자를 안내할 만큼 충분한 프로토콜의 실례는 Berger, Sambrook, and Ausubel, 그리고 Mullis *et al.* (1987) U.S. Pat. No. 4,683,202; Innis *et al.*, eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press Inc. San Diego, Calif.) ("Innis"); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3: 81-94; Kwoh *et al.* (1989) Proc. Natl. Acad. Sci. USA 86: 1173; Guatelli *et al.*(1990) Proc. Nat'l. Acad. Sci. USA 87: 1874; Lomell *et al.* (1989) J. Clin. Chem 35: 1826; Landegren *et al.* (1988) Science 241: 1077-1080; Van Brunt (1990) Biotechnology 8: 291-294; Wu and Wallace (1989) Gene 4:560; Barringer *et al.*(1990) Gene 89:117; 그리고 Sooknanan and Malek (1995) Biotechnology 13: 563-564. Improved methods for cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039에서 확인된다. PCR로 대형 핵산을 증폭하기 위한 향상된 방법은 Cheng *et al.* (1994) Nature 369: 684-685 및 여기에 인용된 참고문헌에서 요약되고, 이에 의해 최대 40 kb의 PCR 앰플리콘 (amplicon)이 산출된다. 당업자가 인지하는 바와 같이, 본질적으로 임의의 RNA가 제한 절단 (restriction digestion), PCR 확대,

그리고 역전사효소와 중합효소를 이용한 염기서열분석 (sequencing)에 적합한 이중 가닥 DNA로 전환될 수 있다 (참조: Ausubel, Sambrook and Berger, all supra).

[0091] 또한, 본 발명에 제시된 벡터 (가령, 클로닝 벡터 또는 발현 벡터)로 형질도입 (형질전환 또는 형질감염)된 조작된 숙주 세포, 그리고 재조합 기술에 의한 본 발명의 폴리펩티드의 생산이 제시된다. 상기 벡터는 예로써, 플라스미드, 바이러스 입자, 파지 등일 수 있다. 조작된 숙주 세포는 프로모터를 활성화시키거나, 형질전환체를 선별하거나, 또는 코딩 폴리뉴클레오티드를 증폭하기 위하여 적절하게 변형된 전통적인 영양 배지에서 배양될 수 있다. 배양 조건, 예를 들면, 온도, pH 등은 발현을 위하여 선별된 숙주 세포에서 기존에 이용된 조건이고, 그리고 예로써, Sambrook, Ausubel and Berger, and Freshney (1994) Culture of Animal Cells: A Manual of Basic Technique, 3rd ed. (Wiley-Liss, New York) 및 여기에 인용된 참고문헌을 비롯한 본 명세서에서 인용된 참고문헌으로부터 당업자에게 명백할 것이다.

[0092] 벡터는 숙주가 단백질 또는 폴리펩티드를 발현하도록 하기 위하여 적절한 숙주를 형질전환시키는데 이용될 수 있다. 적절한 발현 숙주의 실례에는 세균 세포, 예를 들면, 대장균 (*E. coli*), 바실루스 서브틸리스 (*B. subtilis*), 스트렙토마이세스 (*Streptomyces*), 그리고 살모넬라 타이피뮤리움 (*Salmonella typhimurium*); 진균 세포, 예를 들면, 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*), 피치아 파스토리스 (*Pichia pastoris*), 그리고 빵곰팡이 (*Neurospora crassa*); 곤충 세포, 예를 들면, 초파리 (*Drosophila*)와 도둑나방 (*Spodoptera frugiperda*); 포유동물 세포, 예를 들면, CHO, COS, BHK, HEK 293 또는 Bowes 흑색종; 또는 식물 세포 또는 외식체 등이 포함된다.

[0093] 세균 시스템에서, 다수의 발현 벡터는 시토신 디아미나아제 폴리펩티드의 의도된 용도에 따라 선택될 수 있다. 가령, 대량의 CD 폴리펩티드 또는 이의 단편이 상업적 생산 또는 항체의 유도에 요구될 때, 용이하게 정제되는 융합 단백질의 높은 발현 수준을 주동하는 벡터가 바람직할 수 있다. 이런 벡터에는 다중기능적 대장균 (*E. coli*) 클로닝과 발현 벡터, 예를 들면, BLUESCRIPT (Stratagene), 여기서 CD 폴리펩티드 코딩 서열이 하이브리드 단백질이 생산되도록 베타-갈락토시다아제의 아미노-말단 Met 및 차후 7 잔기에 대한 서열과 인-프레임 (in-frame)으로 상기 벡터 내로 결찰된다; pIN 벡터 (Van Heeke & Schuster (1989) J. Biol. Chem. 264: 5503-5509); pET 벡터 (Novagen, Madison Wis.) 등이 포함되지만 이들에 국한되지 않는다.

[0094] 유사하게, 효모 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)에서, 구조성 또는 유도성 프로모터, 예를 들면, 알파 인자, 알코올 옥시다아제와 PGH를 포함하는 다수의 벡터가 본 발명의 CD 폴리펩티드의 생산에 이용될 수 있다. 검토를 위하여, Ausubel (*supra*)와 Grant *et al.* (1987) Methods in Enzymology 153:516-544를 참조한다.

[0095] 본 발명에서는 또한, 기존 레트로바이러스 벡터에 비하여 증가된 안정성을 갖는 복제 적격성 레트로바이러스 벡터를 제시한다. 감염과 복제 동안 이런 증가된 안정성은 세포 증식성 질환의 치료에 중요하다. 형질도입 효율, 이식유전자 (transgene) 안정성과 표적 선택성의 조합은 복제 적격성 레트로바이러스에 의해 제공된다. 이들 조성과 방법은 삽입물 (insert) 안정성을 제공하고 이식유전자의 전사 활성 및 인코딩된 폴리펩티드의 번역 생존력을 유지시킨다.

[0096] 레트로바이러스는 수평과 수직으로 전염될 수 있다. 레트로바이러스의 효율적인 감염성 전염은 비록 바이러스가 낮은 효율에서 수용체-독립적인 비-특이적 침투 경로를 이용할 수도 있지만, 바이러스 외피 단백질을 특이적으로 인식하는 수용체의 표적 세포 상에서 발현을 요구한다. 이에 더하여, 표적 세포 유형은 바이러스가 결합하고 침투한 이후에, 복제 주기 (replication cycle)의 모든 단계를 뒷받침할 수 있어야 한다. 수직 전염은 바이러스 게놈이 숙주의 생식세포주 (germ line)에 통합될 때 발생한다. 프로바이러스 (provirus)는 이후, 마치 세포 유전자인 것처럼 세대에 걸쳐 유전될 것이다. 이로써, 빈번하게 잠복 상태로 있지만 숙주가 적당한 작인에 노출될 때 활성화될 수 있는 내인성 프로바이러스가 확립된다.

[0097] 앞서 언급된 바와 같이, 통합된 DNA 중간물은 프로바이러스로 지칭된다. 기존 유전자 요법 또는 유전자 전달 시스템은 적절한 보조 바이러스의 존재에서, 또는 오염성 보조 바이러스의 동시 생산 없이 외피화 (encapsidation)를 가능하게 하는 적절한 서열을 포함하는 세포주 내에서 프로바이러스의 전사 및 감염성 바이러스 내로 조립을 요구하는 방법과 레트로바이러스를 이용한다. 하기에 기술된 바와 같이, 보조 바이러스는 본 발명의 재조합 레트로바이러스의 생산에 요구되지 않는데, 그 이유는 외피화를 위한 서열이 게놈 내에 제공되고, 따라서 유전자 전달 또는 요법을 위한 복제 적격성 레트로바이러스 벡터를 제공하기 때문이다.

[0098] 본 발명의 방법과 조성물에 유용한 레트로바이러스 게놈은 최소한 3가지 유전자: *gag*, *pol*, 그리고 *env*를 보유

하는 프로바이러스 DNA를 포함하고, 이들 유전자는 1개 또는 2개의 긴 말단 (LTR) 반복이 측면에 위치하고, 또는 프로바이러스에서 2개의 긴 말단 반복 (LTR) 및 cis-작용 서열, 예를 들면, *psi*를 포함하는 서열이 측면에 위치한다. *gag* 유전자는 내부 구조 (매트릭스, 캡시드, 그리고 뉴클레오캡시드) 단백질을 인코딩한다; *pol* 유전자는 RNA-지시된 DNA 중합효소 (역전사효소), 프로테아제와 인테그라아제를 인코딩한다; 그리고 *env* 유전자는 바이러스 외피 당단백질을 인코딩한다. 5' 및/또는 3' LTR은 비리온 RNA의 전사와 폴리아데닐화를 촉진하는 기능을 한다. LTR은 바이러스 복제에 필요한 모든 다른 cis-작용 서열을 포함한다. 렌티바이러스는 *vif*, *vpr*, *tat*, *rev*, *vpu*, *nef*, 그리고 *vpx* (HIV-1, HIV-2 및/또는 SIV에서)를 비롯한 부가적인 유전자를 보유한다.

[0099] 5' LTR에 인접하여, 게놈의 역전사 (tRNA 프라이머 결합 부위) 및 바이러스 RNA의 입자 내로의 효과적인 외피화 (*Psi* 부위)에 필요한 서열이 존재한다. 외피화 (또는 레트로바이러스 RNA의 감염성 비리온 내로의 포장)에 필요한 서열이 바이러스 게놈으로부터 상실되면, 게놈 바이러스 RNA의 외피화를 예방하는 cis 결함이 발생한다. 이러한 유형의 변형된 벡터는 기존 유전자 전달 시스템 (가령, 비리온의 외피화에 요구되는 요소가 부재하는 시스템)에 전형적으로 이용되는 것이다.

[0100] 한 구체예에서, 본 발명에서는 비-분열 세포, 분열 세포, 또는 세포 증식성 질환을 앓는 세포를 감염시킬 수 있는 재조합 레트로바이러스를 제시한다. 본 발명의 재조합 복제 적격성 레트로바이러스는 바이러스 GAG, 바이러스 POL, 바이러스 ENV, 그리고 비리온 내에서 외피화된 내부 리보솜 결합 부위 (IRES)가 선행하는 이종성 폴리뉴클레오티드를 인코딩하는 폴리뉴클레오티드 서열을 포함한다.

[0101] 이종성 핵산 서열은 IRES에 작동가능하게 연결된다. 본 명세서에서, 용어 "이종성" 핵산 서열 또는 이식유전자는 (i) 야생형 레트로바이러스 내에 정상적으로 존재하지 않는 서열, (ii) 외래 종으로부터 기원하는 서열, 또는 (iii) 동일한 종으로부터 기원하면, 최초 형태로부터 실질적으로 변형된 것을 지칭한다. 대안으로, 세포 내에서 정상적으로 발현되지 않는 변화되지 않은 핵산 서열은 이종성 핵산 서열이다. 특정 구체예에서, 이종성 폴리뉴클레오티드는 시토신 디아미나아제 활성을 갖는 본 발명의 폴리펩티드이다.

[0102] 구(句) "비-분열" 세포는 유사분열 (mitosis)을 겪지 않는 세포를 지칭한다. 비-분열 세포는 상기 세포가 활동적으로 분열하지 않으면, 세포 주기의 임의의 시점 (가령, G_0/G_1, $G_{1/S}$, $G_{2/M}$)에서 차단될 수 있다. 탈체에서 감염의 경우에, 분열 세포는 방사 (irradiation), 아피도콜린 (aphidocolin) 치료, 혈청 기아 (serum starvation), 그리고 접촉 저해 (contact inhibition)를 비롯하여 당업자에 의해 이용되는 표준 기술에 의해 세포 분열이 차단되도록 처리될 수 있다. 하지만, 탈체에서 감염은 종종, 세포를 차단하지 않으면서 수행되는 것으로 이해되는데, 그 이유는 많은 세포가 이미 정지되어 있기 때문이다 (가령, 줄기 세포). 가령, 재조합 렌티바이러스 벡터는 세포 분열을 차단하는데 이용되는 기전, 또는 세포가 차단되는 세포 주기에서 시점에 상관없이, 임의의 비-분열 세포를 감염시킬 수 있다. 체내에 미리 존재하는 비-분열 세포의 실례에는 신경, 근육, 간, 피부, 심장, 폐, 그리고 골수 세포, 그리고 이들의 유도체가 포함된다. 분열 세포의 경우에, 온코-레트로바이러스 벡터가 이용될 수 있다.

[0103] "분열" 세포는 활성 유사분열, 또는 감수분열 (meiosis)을 겪는 세포를 의미한다. 이런 분열 세포에는 줄기 세포, 피부 세포 (가령, 섬유아세포와 각질세포), 생식체 (gamete), 그리고 당분야에 공지된 기타 분열 세포가 포함된다. 특히, 용어 분열 세포에는 세포 증식성 질환을 앓는 세포, 예를 들면, 종양 세포가 포함된다. 용어 "세포 증식성 질환"은 비정상적으로 많은 세포로 특징되는 장애를 지칭한다. 이러한 장애는 비대성 (조직 내에서 세포 집단의 과형성을 유발하는 세포의 연속적인 증가)과 발육부전성 (조직 내에서 세포의 부재 또는 결핍) 세포 성장 또는, 신체 부위 내로 세포의 과도한 유입 또는 이동을 수반할 수 있다. 이들 세포 집단은 반드시 형질전환된, 발암성 또는 악성 세포인 것은 아니며 정상 세포 역시 포함할 수 있다. 세포 증식성 질환에는 연결 조직 (connective tissue)의 과형성과 연관된 질환, 예를 들면, 경피증, 관절염과 간경변을 비롯한 다양한 섬유성 장애 (fibrotic condition)가 포함된다. 세포 증식성 질환에는 종양 질환, 예를 들면, 두경부 암종이 포함된다. 두경부 암종은 예로써, 입, 식도, 인후, 후두, 갑상선, 혀, 입술, 타액선, 코, 부비동, 비인두의 암종, 상부 비천정 (nasal vault)과 부비동 종양, 후신경모세포종 (esthesioneuroblastoma), 편평 세포 암, 악성 흑색종, 비강상악동 미분화 암종 (sinonasal undifferentiated carcinoma, SNUC), 뇌 (아교모세포종 포함) 또는 혈액 신조직 형성을 포함할 것이다. 또한, 경부 림프절 (cervical lymph node), 후두앞 림프절 (prelaryngeal lymph node), 허파식도옆 림프절 (pulmonary juxtaesophagal lymph node)과 악하 림프절 (submandibular lymph node)을 비롯한 국부 림프절의 암종이 포함된다 (Harrison's Principles of Internal Medicine (eds., Isselbacher, *et al.*, McGraw-Hill, Inc., 13th Edition, pp1850-1853, 1994). 다른 암 유형에는 폐 암, 결장-직장 암, 유방 암, 전립선 암, 요도관 암, 자궁 암 림프종, 구강 암, 췌장 암, 백혈병, 흑색종, 위 암, 피부 암

과 난소 암이 포함되지만 이들에 국한되지 않는다.

[0104] 한 구체예에서, 벡터 내에서 이종성 폴리뉴클레오티드는 인간 세포에서 발현을 위하여 최적화된 시토신 디아미나아제를 포함한다. 진전된 구체예에서, 시토신 디아미나아제는 인간 코돈 최적화된 서열을 포함하고, 그리고 야생형 시토신 디아미나아제와 비교하여 시토신 디아미나아제의 안정성 (가령, 감소된 분해 또는 증가된 열-안정성)을 증가시키는 돌연변이를 보유한다. 또 다른 구체예에서, 이종성 폴리뉴클레오티드는 UPRT 또는 OPRT 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오티드에 작동가능하게 연결된 시토신 디아미나아제 (인간 코돈 최적화된 또는 비-최적화된, 돌연변이된 또는 비-돌연변이된)를 포함하는 융합 구조체를 인코딩한다. 다른 구체예에서, 이종성 폴리뉴클레오티드는 본 발명의 CD 폴리뉴클레오티드 또는 융합 구조체 (가령, SEQ ID NO: 3, 5, 11, 13, 15, 또는 17)를 포함한다.

[0105] 다른 구체예에서, 복제 적격성 레트로바이러스 벡터는 시토신 디아미나아제를 포함하는 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드 (본 명세서에서 기술된 바와 같음)를 포함할 수 있고, 그리고 세포-유형 또는 조직 특이적 프로모터에 연결된 miRNA 또는 siRNA 분자를 포함하는 폴리뉴클레오티드를 더욱 포함할 수 있다.

[0106] 용어 "조절 핵산 서열"은 수용자 세포 내에서 코딩 서열의 복제, 전사와 번역을 집합적으로 제공하는 프로모터 서열, 폴리아데닐화 신호, 전사 종료 서열, 상류 조절 도메인, 복제 기점, 인핸서 등을 집합적으로 지칭한다. 선택된 코딩 서열이 적절한 숙주 세포 내에서 복제되고, 전사되고, 그리고 번역될 수 있다면, 이들 조절 서열 모두가 항상 존재해야 하는 것은 아니다. 당업자는 공개된 데이터베이스와 자료로부터 조절 핵산 서열을 용이하게 확인할 수 있다. 더 나아가, 당업자는 의도된 용도, 예를 들면, 생체내, 탈체, 또는 시험관내에 적용가능한 조절 서열을 확인할 수 있다.

[0107] 본 명세서에서, 용어 "프로모터 영역"은 통상적인 의미에서, DNA 조절 서열을 포함하는 뉴클레오티드 영역을 지칭하고, 여기서 상기 조절 서열은 RNA 중합효소에 결합할 수 있고 하류 (3'-방향) 코딩 서열의 전사를 개시할 수 있는 유전자로부터 유래된다. 조절 서열은 원하는 유전자 서열에 상동성 또는 이종성이다. 가령, 앞서 기술된 바와 같은 바이러스 또는 포유동물 프로모터를 비롯한 넓은 범위의 프로모터가 이용될 수 있다.

[0108] 내부 리보솜 결합 부위 ("IRES")는 코딩 서열의 번역 동안, 통상적으로 IRES에 대하여 3'에 리보솜의 결합 또는 유지를 촉진하는 핵산의 조각을 지칭한다. 일부 구체예에서, IRES는 접합 수용자/공여자를 포함하지만, 선호되는 IRES는 접합 수용자/공여자 부위가 부재한다. 정상적으로, 메신저 RNA 내로 리보솜의 결합은 모든 진핵 mRNA의 5' 말단에 위치된 캡 (cap)을 통해 발생한다. 하지만, 이러한 보편적인 규칙에 예외가 존재한다. 일부 바이러스 mRNA에서 캡의 부재는 이들 RNA의 내부 부위에서 리보솜의 결합을 허용하는 대안적 구조의 존재를 암시한다. 현재까지, 기능 때문에 IRES로 명명된 다수의 이런 구조가 예로써, 특히 피코르나바이러스 (picornavirus), 예를 들면, 회백수염 바이러스 (poliomyelitis virus) (Pelletier *et al.*, 1988, Mol. Cell. Biol., 8, 1103-1112)와 EMCV 바이러스 (encephalo-myocarditis virus) (Jang *et al.*, J. Virol., 1988, 62, 2636-2643)의 캡이 없는 (uncapped) 바이러스 mRNA의 5' 비-코딩 영역에서 확인되었다. 본 발명에서는 복제-적격성 레트로바이러스 벡터의 배경에서 IRES의 용도를 제시한다.

[0109] 이종성 핵산 서열은 전형적으로, 바이러스 LTR 프로모터-인핸서 신호 또는 내부 프로모터의 제어 하에 있고, 그리고 레트로바이러스 LTR 내에서 유지된 신호는 여전히, 숙주 세포 게놈 내로 벡터의 효과적인 통합을 유발할 수 있다. 따라서 본 발명의 재조합 레트로바이러스 벡터, 원하는 서열, 유전자 및/또는 유전자 단편은 상이한 조절 서열 하에서, 여러 부위에 삽입될 수 있다. 가령, 삽입 부위는 바이러스 인핸서/프로모터 근위 부위 (즉, 5' LTR-주동된 유전자 좌위)일 수 있다. 대안으로, 원하는 서열은 원위 부위 (가령, env 유전자에 대하여 3'에 IRES 서열) 내로 삽입될 수 있고, 또는 2개 이상의 이종성 서열이 존재하는 경우에, 첫 번째 이종성 서열은 첫 번째 조절 영역의 제어 하에 있고 두 번째 이종성 서열은 두 번째 조절 영역의 제어 하에 있을 수 있다. 다른 원위 부위에는 바이러스 프로모터 서열이 포함되고, 여기서 프로모터 근위 시스트론, SV40 또는 CMV와 같은 내부 이종성 프로모터, 또는 내부 리보솜 결합 부위 (IRES)의 접합을 통한 원하는 서열 또는 서열들의 발현이 이용될 수 있다.

[0110] 한 구체예에서, 본 발명의 레트로바이러스 게놈은 원하는 폴리뉴클레오티드 서열의 삽입을 위한 클로닝 부위를 포함하는 IRES를 내포한다. 한 구체예에서, IRES는 레트로바이러스 벡터 내에서 env 유전자에 대하여 3'에 위치하지만 원하는 이종성 핵산에 대하여 5'에 위치한다. 따라서 원하는 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드 서열은 IRES에 작동가능하게 연결된다.

[0111] 다른 구체예에서, 표적화 폴리뉴클레오티드 서열은 본 발명의 재조합 레트로바이러스 벡터의 일부로서

포함된다. 표적화 폴리뉴클레오티드 서열은 표적화 리간드 (가령, 펩티드 호르몬, 예를 들면, 헤레굴린 (heregulin), 단일-사슬 항체, 수용체 또는 수용체에 대한 리간드), 조직-특이적 또는 세포-유형 특이적 조절 요소 (가령, 조직-특이적 또는 세포-유형 특이적 프로모터 또는 인핸서), 또는 표적화 리간드와 조직-특이적/세포-유형 특이적 조절 요소의 조합이다. 바람직하게는, 표적화 리간드는 레트로바이러스의 env 단백질에 작동가능하게 연결되고, 키메라 레트로바이러스 env 단백질을 산출한다. 바이러스 GAG, 바이러스 POL과 바이러스 ENV 단백질은 임의의 적절한 레트로바이러스 (가령, MLV 또는 렌티바이러스-유래된)로부터 유래될 수 있다. 다른 구체예에서, 바이러스 ENV 단백질은 비-레트로바이러스-유래된다 (가령, CMV 또는 VSV).

[0112] 이런 이유로, 본 발명의 재조합 레트로바이러스는 핵산 게놈이 표적 비-분열 세포, 표적 분열 세포, 또는 세포 증식성 질환을 앓는 표적 세포에 전달되도록, 상기 바이러스가 특정 세포 유형 (가령, 평활근 세포, 간 세포, 신장 세포, 섬유아세포, 각질세포, 중간엽 줄기 세포, 골수 세포, 연골 세포, 상피 세포, 장내 세포, 종양 세포, 신경교종 세포, 뉴런 세포, 그리고 당분야에 공지된 기타 세포)에 표적화되는 방식으로 유전적으로 변형된다. 표적화는 2가지 방식으로 달성될 수 있다. 첫 번째 방식은 외부 표면 상에 분자를 보유하는 세포에 결합함으로써 레트로바이러스를 표적 세포에 지향시킨다. 레트로바이러스를 표적화시키는 이러한 방법은 레트로바이러스의 표면 상에서 표적화 리간드와 상호작용하는 수용체 또는 결합 분자를 보유하는 세포 또는 조직으로 상기 바이러스의 표적화를 보조하기 위하여 레트로바이러스의 외막 상에서 표적화 리간드의 발현을 이용한다. 바이러스에 의한 세포의 감염후, 상기 바이러스는 자신의 핵산을 세포 내로 주입하고, 그리고 레트로바이러스 유전 물질은 숙주 세포 게놈 내로 통합될 수 있다. 표적화를 위한 두 번째 방법은 하기에 더욱 상세하게 기술된 바와 같이, 조절 요소를 활발하게 이용하는 표적화된 세포 내에서 바이러스 게놈의 발현과 전사를 촉진하기 위하여 세포- 또는 조직-특이적 조절 요소를 이용한다. 전달된 레트로바이러스 유전 물질은 이후, 숙주 세포 내에서 전사되고 단백질로 번역된다. 이러한 표적화 조절 요소는 전형적으로, 5' 및/또는 3' LTR에 연결되고 키메라 LTR을 산출한다. 예로써, 특이적 표적 세포 상에서 수용체에 대한 리간드를 인코딩하는 다른 유전자와 함께, 목적되는 이종성 핵산 서열을 본 발명의 바이러스 벡터 내로 삽입함으로써, 상기 벡터는 표적 특이적이다. 바이러스 벡터는 예로써, 당, 당지질, 또는 단백질을 부착함으로써 표적 특이적으로 만들어질 수 있다. 표적화는 바이러스 벡터를 표적으로 하는 항체를 이용함으로써 달성될 수 있다. 당업자는 목적되는 핵산 서열을 포함하는 바이러스 벡터의 표적 특이적 전달이 가능하도록 바이러스 외피에 부착될 수 있는 바이러스 게놈 또는 단백질 내로 삽입될 수 있는 특이적 폴리뉴클레오티드 서열을 인지하거나, 또는 용이하게 확인할 수 있다.

[0113] 따라서 본 발명은 한 구체예에서, 표적화 폴리펩티드에 작동가능하게 연결된 레트로바이러스 env 단백질을 포함하는 키메라 env 단백질을 제시한다. 표적화 폴리펩티드는 세포 특이적 수용체 분자, 세포 특이적 수용체에 대한 리간드, 세포 특정 항원성 에피토프에 대한 항체 또는 항체 단편, 또는 표적 세포에 결합하거나 표적 세포와 상호작용할 수 있는 당분야에서 용이하게 확인된 임의의 다른 리간드일 수 있다. 표적화 폴리펩티드 또는 분자의 실례에는 비오틴-스트렙타비딘을 링커로서 이용한 이가 항체 (Etienne-Julan *et al.*, J. Of General Virol., 73, 3251-3255 (1992); Roux *et al.*, Proc. Natl. Acad. Sci USA 86, 9079-9083 (1989)), 합텐에 대한 단일-사슬 항체 가변 영역을 인코딩하는 서열을 외피 내에 포함하는 재조합 바이러스 (Russell *et al.*, Nucleic Acids Research, 21, 1081-1085 (1993)), 레트로바이러스 외피 내로 펩티드 호르몬 리간드의 클로닝 (Kasahara *et al.*, Science, 266, 1373-1376 (1994)), 키메라 EPO/env 구조체 (Kasahara *et al.*, 1994), LDL 수용체를 발현하는 HeLa 세포의 특이적 감염을 유발하는, 에코트로픽 (ecotropic) MLV 외피 내에서 저밀도 지단백 (LDL) 수용체에 대한 단일-사슬 항체 (Somia *et al.*, Proc. Natl. Acad. Sci USA, 92, 7570-7574 (1995)) 등이 포함된다. 유사하게, ALV의 숙주 범위는 상기 바이러스가 종을 뛰어넘어 쥐 아교모세포종 세포를 특이적으로 감염시킬 수 있도록 인테그린 리간드의 통합에 의해 변경될 수 있고 (Valsesia-Wittmann *et al.*, J. Virol. 68, 4609-4619 (1994)), 그리고 Dornberg 등 (Chu and Dornburg, J. Virol 69, 2659-2663 (1995))은 종양 마커에 대한 단일-사슬 항체를 포함하는 외피를 이용한, 비장 괴사 바이러스 (spleen necrosis virus, SNV), 조류 레트로바이러스의 조직-특이적 표적화를 보고하였다.

[0114] 본 발명에서는 표적 세포를 감염시킬 수 있는 재조합 레트로바이러스를 생산하는 방법을 제시하고, 상기 방법은 바이러스 gag, 바이러스 pol과 바이러스 env를 인코딩하는 폴리뉴클레오티드 서열을 포함하는 벡터로 적절한 숙주 세포를 형질감염시키는 단계, 여기서 상기 벡터는 조절 핵산 서열에 작동가능하게 연결된, 이종성 유전자의 도입을 위한 클로닝 부위를 포함하고, 그리고 재조합 바이러스를 회수하는 단계를 포함한다.

[0115] 본 발명의 레트로바이러스와 방법은 비리온을 증식시키고 생산하기 위하여 보조 바이러스 또는 부가적인 핵산 서열 또는 단백질을 필요로 하지 않는 복제 적격성 레트로바이러스를 제공한다. 가령, 본 발명의 레트로바이러스의 핵산 서열은 예로써, 군 특이적 항원과 역전사효소 (그리고 성숙과 역전사에 필요한 인테그라아제와 프로

테아제-효소)를 앞서 기술된 바와 같이 각각 인코딩한다. 바이러스 gag와 pol은 렌티바이러스, 예를 들면, HIV, 또는 종양바이러스, 예를 들면, MoMLV로부터 유래될 수 있다. 이에 더하여, 본 발명의 레트로바이러스의 핵산 게놈은 바이러스 외피 (ENV) 단백질을 인코딩하는 서열을 포함한다. env 유전자는 임의의 레트로바이러스로부터 유래될 수 있다. env는 인간 및 기타 종의 세포의 형질도입을 가능하게 하는 양쪽성 외피 단백질이거나, 또는 생쥐와 쥐 세포만을 형질도입할 수 있는 에코트로픽 외피 단백질일 수 있다. 게다가, 특정 세포-유형의 수용체에 표적화를 위한 항체 또는 특정 리간드로 외피 단백질의 연쇄에 의해 재조합 바이러스를 표적으로 하는 것이 바람직하다. 앞서 언급된 바와 같이, 레트로바이러스 벡터는 예로써, 당지질, 또는 단백질을 삽입함으로써 표적 특이적으로 만들어질 수 있다. 표적화는 종종, 특정 세포-유형 (가령, 일정한 조직에서 관찰되는 세포 유형, 또는 암 세포 유형)에서 항원에 레트로바이러스 벡터를 표적화시키는 항체를 이용함으로써 달성된다. 당업자는 과도한 실험 없이, 특이적인 표적에 레트로바이러스 벡터의 전달을 달성하는 특이적 방법을 인지하거나, 또는 용이하게 확인할 수 있다. 한 구체예에서, env 유전자는 비-레트로바이러스 (가령, CMV 또는 VSV)로부터 유래된다. 레트로바이러스-유래된 env 유전자의 실례에는 Moloney 뮤린 백혈병 바이러스 (MoMuLV), Harvey 뮤린 육종 바이러스 (HaMuSV), 뮤린 유방 종양 바이러스 (MuMTV), Gibbon 원숭이 백혈병 바이러스 (GaLV), 인간 면역결핍 바이러스 (HIV)와 Rous 육종 바이러스 (RSV)가 포함되지만 이들에 국한되지 않는다. 다른 env 유전자, 예를 들면, 수포성 구내염 바이러스 (VSV) (단백질 G), 사이토메갈로바이러스 외피 (CMV), 또는 인플루엔자 바이러스 적혈구응집소 (HA) 역시 이용될 수 있다.

[0116] 한 구체예에서, 레트로바이러스 게놈은 온코-레트로바이러스 또는 감마레트로바이러스 (gammaretrovirus), 더욱 구체적으로 포유동물 온코-레트로바이러스 또는 감마 레트로바이러스로부터 유래된다. "유래된"은 부모 폴리뉴클레오티드 서열이 자연 발생 서열의 삽입 또는 제거 (가령, IRES의 삽입, 예로써, 시토신 디아미나아제 활성을 갖는 본 발명의 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드의 삽입 등)에 의해 변형된 야생형 종양바이러스라는 것을 의미한다.

[0117] 다른 구체예에서, 본 발명에서는 조절 서열을 이용하여 표적화된 레트로바이러스 벡터를 제시한다. 세포- 또는 조직-특이적 조절 서열 (가령, 프로모터)은 특정한 세포 집단에서 유전자 서열의 발현을 표적으로 하는데 이용될 수 있다. 본 발명에 적합한 포유동물과 바이러스 프로모터는 본 명세서 전반에서 기술된다. 따라서 한 구체예에서, 본 발명에서는 레트로바이러스 게놈의 5' 말단에서 조직-특이적 프로모터 요소를 보유하는 레트로바이러스를 제시한다. 바람직하게는, 예로써 종양 세포에 대한 세포- 또는 조직-특이적 프로모터와 인핸서 (가령, 종양 세포-특이적 인핸서와 프로모터), 그리고 유도성 프로모터 (가령, 테트라사이클린)를 비롯한 이들 조직-특이적 조절 요소/서열은 레트로바이러스 게놈의 LTR의 U3 영역 내에 위치한다.

[0118] 일부 상황에서, 발현을 조절하는 것이 바람직할 수 있다. 가령, 원하는 발현의 수준에 따라, 다양한 강도의 활성을 갖는 상이한 바이러스 프로모터가 이용될 수 있다. 포유동물 세포에서, CMV 극초기 프로모터는 종종, 강한 전사 활성화를 제공하는데 이용된다. 덜 강력한 CMV 프로모터의 변형된 이형 역시 이식유전자의 감소된 발현 수준이 요구될 때 이용된다. 조혈 세포 내에서 이식유전자의 발현이 요구될 때, 레트로바이러스 프로모터, 예를 들면, MLV 또는 MMTV로부터 LTR이 이용될 수 있다. 이용될 수 있는 다른 바이러스 프로모터에는 SV40, RSV LTR, HIV-1과 HIV-2 LTR, 아데노바이러스 프로모터, 예를 들면, E1A, E2A, 또는 MLP 영역으로부터 아데노바이러스 프로모터, AAV LTR, 콜리플라워 모자이크 바이러스 (cauliflower mosaic virus), HSV-TK, 그리고 조류 육종 바이러스가 포함된다.

[0119] 유사하게, 비-표적화된 조직에 대한 잠재적 독성 또는 바람직하지 않은 효과를 감소시키기 위하여 특정한 조직 또는 세포에서 전사를 실행하는 조직 특이적 또는 선택적 프로모터가 이용될 수 있다. 가령, 프로모터, 예를 들면, PSA, 프로바신 (probasin), 전립성 산성 포스파타아제 (prostatic acid phosphatase) 또는 전립선-특이적 입상 칼리크레인 (glandular kallikrein) (hK2)이 전립선 내에서 유전자 발현을 표적으로 하는데 이용된다. 이용될 수 있는 다른 프로모터/조절 도메인은 표 1에서 열거된다.

[0120] 일정한 징후에서, 유전자 요법 벡터의 투여후 특정 시점에서 전사를 활성화시키는 것이 바람직할 수 있다. 이는 호르몬 또는 사이토킨 조절성 프로모터와 같은 프로모터로 달성될 수 있다. 가령, 이러한 징후가 특정 스테로이드가 생산되거나 배달되는 생식샘 조직인 치료적 적용에서, 안드로겐 또는 에스트로겐 조절된 프로모터의 이용이 유리할 수 있다. 호르몬 조절성 프로모터에는 MMTV, MT-1, 엑디손 (ecdysone)과 RuBisco가 포함된다. 다른 호르몬 조절된 프로모터, 예를 들면, 갑상선, 뇌하수체와 부신 호르몬에 반응성인 것들이 이용될 수 있다. 이용될 수 있는 사이토킨과 염증성 단백질 반응성 프로모터에는 K와 T 키니노젠 (Kininogen) (Kageyama *et al.*, 1987), c-fos, TNF-알파, C-반응성 단백질 (Arcone *et al.*, 1988), 합토글로빈 (haptoglobin) (Oliviero *et al.*, 1987), 혈청 아밀로이드 A2, C/EBP 알파, IL-1, IL-6 (Poli and Cortese, 1989), 보체 C3 (Wilson *et*

al., 1990), IL-8, 알파-1 산성 당단백질 (Prowse and Baumann, 1988), 알파-1 항트립신 (antitrypsin), 지단백 리파아제 (Zechner *et al.*, 1988), 안지오텐시노겐 (angiotensinogen) (Ron *et al.*, 1990), 피브리노겐, c-jun (포르볼 에스테르, TNF-알파, UV 방사, 레티노산, 그리고 과산화수소에 의해 유도), 콜라게나아제 (collagenase) (포르볼 에스테르와 레티노산에 의해 유도), 메탈로티오네인 (metallothionein) (중금속과 글루코코르티코이드 유도), 스트로멜리신 (Stromelysin) (포르볼 에스테르, 인터루킨-1과 EGF에 의해 유도), 알파-2 마크로글로불린 (macroglobulin)과 알파-1 항키모트립신 (antichymotrypsin)이 포함된다. 종양 특이적 프로모터, 예를 들면, 오스테오칼신 (osteocalcin), 저산소증-반응 요소 (hypoxia-responsive element, HRE), MAGE-4, CEA, 알파-태아단백질, GRP78/BiP와 티로시나아제 (tyrosinase) 역시 종양 세포 내에서 유전자 발현을 조절하는데 이용될 수 있다.

[0121] 이에 더하여, 이러한 프로모터 목록은 총체적 또는 제한적인 것으로 간주되지 않으며, 당업자는 본 명세서에서 개시된 프로모터와 방법에 관련하여 이용될 수 있는 다른 프로모터를 인지할 것이다.

[0123] 조직 특이적 프로모터

[0124] 조직 프로모터

[0125] 췌장 인슐린 엘라스틴 아밀라아제

[0126] pdr-1 pdx-1 글루코키나아제

[0127] 간 알부민 PEPCK HBV 인핸서

[0128] α 태아단백질 아포리포단백질 C α-1

[0129] 항트립신 비텔로제닌, NF-AB

[0130] 트랜스사이레틴

[0131] 골격근 미오신 H 사슬 근육 크레아틴 키나아제

[0132] 디스트로핀 칼페인 p94 Skeletal

[0133] 알파-액틴 신속 트로포닌 1

[0134] 피부 케라틴 K6 케라틴 K1

[0135] 폐 CFTR 인간 사이토케라틴 18 (K18)

[0136] 폐 계면활성제 단백질 A, B와 C CC-10 P1

[0137] 평활근 sm22 α SM-알파-액틴

[0138] 내피 세포층 엔도텔린-1 E-셀렉틴

[0139] 폰 빌레브란트 인자 TIE, KDR/flk-1

[0140] 멜라닌세포 티로시나아제

[0141] 지방 조직 지단백 리파아제 (Zechner *et al.*, 1988)

[0142] 아디프신 (Spiegelman *et al.*, 1989)

[0143] 아세틸-CoA 카르복실라아제 (Pape and Kim, 1989)

[0144] 글리세로인산탈수소효소 (Dani *et al.*, 1989)

[0145] 지방세포 P2 (Hunt *et al.*, 1986)

[0146] 유방 유장 산성 단백질 (WAP) (Andres et al. PNAS

[0147] 84:1299-1303 1987

[0148] 혈액 β-글로빈

[0150] "조직-특이적 조절 요소"는 한쪽 조직 내에서 유전자의 전사를 주동할 수 있지만 다른 조직 유형에서 거의 "침묵하는"조절 요소 (가령, 프로모터)이다. 하지만, 조직-특이적 프로모터는 그들이 침묵하는 조직 내에서 검출가

능한 양의 "배경" 또는 "기본" 활성을 가질 수 있는 것으로 이해될 것이다. 프로모터가 표적 조직 내에서 선택적으로 활성화되는 정도는 선택성 비율 (표적 조직에서 활성/대조 조직에서 활성)로서 표시될 수 있다. 이와 관련하여, 본 발명의 실시에 유용한 조직 특이적 프로모터는 전형적으로, 대략 5 이상의 선택성 비율을 갖는다. 바람직하게는, 선택성 비율은 대략 15 이상이다.

[0151] 게다가, 활성에서 단일 조직 유형에 한정됨 없이 일정한 프로모터는 그럼에도 불구하고, 그들이 일군의 조직에서 활발하고 다른 군에서 덜 활발하거나 침묵한다는 점에서 선택성을 나타내는 것으로 이해될 것이다. 이런 프로모터 역시, "조직 특이적"으로 지칭되고, 그리고 본 발명에 이용이 고려된다. 따라서 본 발명에 이용되는 조직-특이적 조절 요소는 이종성 단백질, 예를 들면, 시토신 디아미나아제 활성을 갖는 본 발명의 폴리펩티드의 조절에 적용가능성, 그리고 레트로바이러스 벡터 내에서 표적화 폴리뉴클레오티드 서열로서 적용가능성을 갖는다.

[0152] 본 발명의 레트로바이러스 벡터와 CD 폴리뉴클레오티드와 폴리펩티드는 다수의 세포 증식성 질환과 질병 (참조: U.S. Pat. No. 4,405,712와 4,650,764; Friedmann, 1989, Science, 244:1275-1281; Mulligan, 1993, Science, 260:926-932, R. Crystal, 1995, Science 270:404-410, 이들 각각은 본 발명에 참조로서 편입됨; 또한, The Development of Human Gene Therapy, Theodore Friedmann, Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1999. ISBN 0-87969-528-5, 이는 본 발명에 참조로서 편입됨)을 비롯한 넓은 범위의 질환과 질병을 치료하는데 이용될 수 있다.

[0153] 본 발명에서는 또한, 세포 증식성 질환의 치료를 위한 유전자 요법을 제시한다. 이런 요법은 증식성 질환을 앓는 개체의 세포 내로 적절한 치료 폴리뉴클레오티드 서열 (가령, 시토신 디아미나아제 활성을 갖는 본 발명의 폴리펩티드)의 도입에 의해 치료 효과를 달성할 것이다. 폴리뉴클레오티드 구조체의 전달은 본 발명의 재조합 레트로바이러스 벡터를 이용함으로써 달성될 수 있다.

[0154] 이에 더하여, 본 명세서에 기술된 바와 같은 치료 방법 (가령, 유전자 요법 또는 유전자 전달 방법)은 생체내에서 또는 탈체에서 수행될 수 있다. 가령, 세포 증식성 질환 또는 질병의 치료를 위한 방법에서, 예로써 외과적으로 또는 방사에 의해, 유전자 요법에 앞서 대부분의 종양을 제거하는 것이 유용하다. 일부 구체예에서, 레트로바이러스 요법은 화학요법이 선행하거나 후행한다.

[0155] 따라서 본 발명에서는 비-분열 세포, 분열 세포 또는 종양 세포를 감염시킬 수 있는 재조합 레트로바이러스를 제시하고, 여기서 상기 재조합 레트로바이러스는 바이러스 GAG; 바이러스 POL; 바이러스 ENV; IRES에 작동가능하게 연결된 이종성 핵산 (가령, 시토신 디아미나아제 활성을 갖는 본 발명의 폴리펩티드 포함); 그리고 포장, 역전사와 통합에 필요한 cis-작용 핵산 서열을 포함한다. 이러한 재조합 레트로바이러스는 렌티바이러스, 예를 들면, HIV, 또는 종양바이러스일 수 있다. 재조합 레트로바이러스를 생산하는 방법에 대하여 앞서 기술된 바와 같이, 본 발명의 재조합 레트로바이러스는 VPR, VIF, NEF, VPX, TAT, REV, 그리고 VPU 단백질 중에서 최소한 한 가지를 더욱 포함할 수 있다. 특정 이론에 한정됨 없이, 이들 유전자/단백질 산물 중에서 한 가지 이상이 생산된 재조합 레트로바이러스 (가령, NEF)의 바이러스 역가를 증가시키는데 중요하거나, 또는 비리온의 감염과 포장에 필요할 것으로 생각된다.

[0156] 본 발명에서는 또한, 특정 핵산 (가령, 이종성 서열, 예를 들면, 시토신 디아미나아제 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오티드)의 발현을 제공하기 위하여 표적 세포로의 핵산 전달의 방법을 제시한다. 이런 이유로, 다른 구체예에서, 본 발명에서는 표적 세포에서 이종성 핵산의 도입과 발현을 위한 방법을 제시하고, 상기 방법은 본 발명의 재조합 바이러스로 표적 세포를 형질감염시키는 단계, 그리고 표적 세포 내에서 이종성 핵산을 발현시키는 단계를 포함한다. 앞서 언급된 바와 같이, 표적 세포는 그들이 레트로바이러스에 의해 감염될 수 있으면, 분열, 비-분열, 종양, 영속화, 변형된, 그리고 당업자에 의해 인식되는 기타 세포 유형을 비롯한 임의의 세포 유형일 수 있다.

[0157] 본 발명에서는 또한, 세포 증식성 또는 면역원성 질환의 치료를 위한 유전자 요법을 제시한다. 한 구체예에서, 세포 증식성 질환은 본 발명의 CD 폴리뉴클레오티드를 도입하고, 상기 폴리뉴클레오티드를 발현시켜 시토신 디아미나아제 활성을 포함하는 폴리펩티드를 생산하고, 그리고 세포독성 양의 5-FU를 생산할 만큼 충분한 기간 동안 충분한 양으로 세포를 5-플루시토신과 접촉시킴으로써 치료된다.

[0158] 이에 더하여, 본 발명에서는 본 발명의 재조합 레트로바이러스 벡터를 인코딩하는 폴리뉴클레오티드 서열을 제시한다. 상기 폴리뉴클레오티드 서열은 다양한 바이러스 입자 내로 통합될 수 있다. 가령, 유전자 요법에 이용될 수 있는 다양한 바이러스 벡터에는 아데노바이러스, 포진 바이러스, 우두, 또는 바람직하게는, RNA

바이러스, 예를 들면, 레트로바이러스, 더욱 구체적으로 포유동물 종양바이러스가 포함된다. 레트로바이러스 벡터는 뮤린, 유인원 또는 인간 레트로바이러스의 유도체일 수 있다. 외래 유전자 (가령, 이종성 폴리뉴클레오티드 서열)가 삽입될 수 있는 레트로바이러스 벡터의 실례에는 Moloney 뮤린 백혈병 바이러스 (MoMuLV), Harvey 뮤린 육종 바이러스 (HaMuSV), 뮤린 유방 종양 바이러스 (MuMTV), 그리고 Rous 육종 바이러스 (RSV)의 유도체가 포함되지만 이들에 국한되지 않는다. 이들 벡터 모두 형질도입된 세포가 확인되고 산출될 수 있도록 하기 위하여, 선별가능 마커에 대한 유전자를 전달하거나 통합할 수 있다. 또 다른 구체예에서, 본 발명에서는 재조합 레트로바이러스 유래된 구조체를 포함하는 플라스미드를 제시한다. 상기 플라스미드는 표적 세포 또는 세포 배양액, 예를 들면, NIH 3T3, HT1080 (인간), CF2 (개) 또는 다른 조직 배양 세포 내로 직접적으로 도입될 수 있다. 결과의 세포는 레트로바이러스 벡터를 배양 배지 내로 방출한다.

[0159] 본 발명에서는 5'에서 3'로, 전사를 개시하는데 유용한 프로모터 또는 조절 영역; psi 포장 신호; *gag* 인코딩 핵산 서열, *pol* 인코딩 핵산 서열; *env* 인코딩 핵산 서열; 내부 리보솜 결합 부위 핵산 서열; 마커, 치료 (가령, 시토신 디아미나아제 활성을 갖는 폴리펩티드) 또는 진단 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드; 그리고 LTR 핵산 서열을 포함하는 폴리뉴클레오티드 구조체를 제시한다. 특정 구체예에서, 시토신 디아미나아제 활성을 갖는 폴리펩티드를 인코딩하는 이종성 폴리뉴클레오티드는 UPRT 또는 OPRT 활성을 포함하는 폴리펩티드를 인코딩하는 도메인을 더욱 포함할 수 있다.

[0160] 본 명세서 전반에서, 그리고 하기에 기술된 바와 같이, 본 발명의 폴리뉴클레오티드 구조체 (가령, 재조합 복제 적격성 레트로바이러스 폴리뉴클레오티드)의 다양한 조각은 부분적으로, 원하는 숙주 세포, 발현 시간 또는 양, 그리고 이종성 폴리뉴클레오티드에 따라 조작된다. 본 발명의 복제 적격성 레트로바이러스 구조체는 당업자에 의해 개별적으로 변형될 수 있는 다수의 도메인으로 분할될 수 있다.

[0161] 가령, 프로모터는 1번 뉴클레오티드에서부터 대략 582번 뉴클레오티드까지 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열을 갖는 CMV 프로모터를 포함할 수 있고, 그리고 변형된 프로모터가 전사를 주동하고 개시할 수 있으면, 하나 이상 (가령, 2-5, 5-10, 10-20, 20-30, 30-50, 50-100개 또는 그 이상의 핵산 염기)에 변형을 보유할 수도 있다. 한 구체예에서, 프로모터 또는 조절 영역은 CMV-R-U5 도메인 폴리뉴클레오티드를 포함한다. 상기 CMV-R-U5 도메인은 MLV R-U5 영역에 인간 사이토메갈로바이러스로부터 극초기 프로모터를 포함한다. 한 구체예에서, 상기 CMV-R-U5 도메인 폴리뉴클레오티드는 대략 1번 뉴클레오티드에서부터 대략 1202번 뉴클레오티드까지 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열, 또는 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열에 최소한 95% 동일한 서열을 포함하고, 여기서 상기 폴리뉴클레오티드는 그것에 작동가능하게 연결된 핵산 분자의 전사를 촉진한다. 상기 폴리뉴클레오티드의 gag 도메인은 다수의 레트로바이러스로부터 유래될 수 있지만, 전형적으로 온코레트로바이러스, 더욱 구체적으로 포유동물 온코레트로바이러스로부터 유래될 것이다. 한 구체예에서, gag 도메인은 대략 1203번 뉴클레오티드에서부터 대략 2819번 뉴클레오티드까지 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.8% (소수점 첫째자리까지 반올림) 동일성을 갖는 서열을 포함한다. 상기 폴리뉴클레오티드의 pol 도메인은 다수의 레트로바이러스로부터 유래될 수 있지만, 전형적으로 온코레트로바이러스, 더욱 구체적으로 포유동물 온코레트로바이러스로부터 유래될 것이다. 한 구체예에서, pol 도메인은 대략 2820번 뉴클레오티드에서부터 대략 6358번 뉴클레오티드까지 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.9% (소수점 첫째자리까지 반올림) 동일성을 갖는 서열을 포함한다. 상기 폴리뉴클레오티드의 env 도메인은 다수의 레트로바이러스로부터 유래될 수 있지만, 전형적으로 온코레트로바이러스, 더욱 구체적으로 포유동물 온코레트로바이러스로부터 유래될 것이다. 일부 구체예에서, env 코딩 도메인은 양쪽성 env 도메인을 포함한다. 한 구체예에서, env 도메인은 대략 6359번 뉴클레오티드에서부터 대략 8323번 뉴클레오티드까지 서열, 또는 그것에 최소한 95%, 98%, 99% 또는 99.8% (소수점 첫째자리까지 반올림) 동일성을 갖는 서열을 포함한다. 상기 폴리뉴클레오티드의 IRES 도메인은 다수의 내부 리보솜 결합 부위로부터 획득될 수 있다. 한 구체예에서, IRES는 뇌심근염 바이러스로부터 유래된다. 한 구체예에서, IRES 도메인은 대략 8327번 뉴클레오티드에서부터 대략 8876번 뉴클레오티드까지 서열, 또는 상기 도메인이 리보솜의 결합을 허용하면, 그것에 최소한 95%, 98%, 또는 99% (소수점 첫째자리까지 반올림) 동일성을 갖는 서열을 포함한다. 이종성 도메인은 본 발명의 시토신 디아미나아제를 포함할 수 있다. 한 구체예에서, CD 폴리뉴클레오티드는 인간 코돈 최적화된 서열을 포함한다. 또 다른 구체예에서, CD 폴리뉴클레오티드는 시토신 디아미나아제를 갖는 돌연변이 폴리펩티드를 인코딩하고, 여기서 이들 돌연변이는 용융 온도 (melting temperature, Tm)를 10℃ 정도 증가시키는 증가된 열 안정화를 제공하고, 더욱 넓은 온도 범위에서 지속된 동역학적 활성 및 단백질의 증가된 축적된 수준을 가능하게 한다. 한 구체예에서, 시토신 디아미나아제는 대략 8877번 뉴클레오티드에서부터 대략 9353번 뉴클레오티드까지 SEQ ID NO: 19에서 열거된 바와 같은 서열을 포함한다. 이종성 도메인에 뒤이어, 폴리퓨린 (polypurine) 풍부한 도메인이 위치한다. 3' LTR은 다수의 레트로바이러스, 전형적으로 온코레트로바이러스, 바람직하게는 포유동물 온코레트

로바이러스로부터 유래될 수 있다. 한 구체예에서, 3' LTR은 U3-R-U5 도메인을 포함한다. 또 다른 구체예에서, LTR은 대략 9405번 뉴클레오티드에서부터 대략 9998번 뉴클레오티드까지 SEQ ID NO: 19에서 열거된 바와 같은 서열, 또는 그것에 최소한 95%, 98% 또는 99.5% (소수점 첫째자리까지 반올림) 동일한 서열을 포함한다.

[0162] 본 발명에서는 또한, 5'에서 3'로, CMV-R-U5, MLV R-U5 영역에 인간 사이토메갈로바이러스로부터 극초기 프로모터의 융합; PBS, 역전사효소에 대한 프라이머 결합 부위; 5' 접합 부위; Ψ 포장 신호; gag, MLV 군 특정 항원에 대한 ORF; pol, MLV 중합효소 다중단백질에 대한 ORF; 3' 접합 부위; 4070A env, MLV 균주 4070A의 외피 단백질에 대한 ORF; IRES, 뇌심근염 바이러스의 내부 리보솜 결합 부위; 변형된 시토신 디아미나아제 (열안정화된, 그리고 코돈 최적화된); PPT, 폴리퓨린 다발; 그리고 U3-R-U5, MLV 긴 말단 반복을 포함하는 재조합 레트로바이러스 벡터를 제시한다. 이러한 구조는 도 1에서 더욱 묘사된다.

[0163] 본 발명에서는 SEQ ID NO: 19, 20, 또는 22에서 열거된 바와 같은 서열을 포함하는 레트로바이러스 벡터를 제시한다.

[0164] 다른 구체예에서, 본 발명에서는 세포 증식성 질환을 앓는 개체를 치료하는 방법을 제시한다. 개체는 임의의 포유동물, 바람직하게는 인간일 수 있다. 개체는 본 발명의 재조합 복제 적격성 레트로바이러스 벡터와 접촉된다. 이러한 접촉은 생체내에서 또는 탈체에서 이루어질 수 있다. 본 발명의 레트로바이러스 벡터를 투여하는 방법은 당분야에 공지되어 있고, 여기에는 예로써, 전신 투여, 국소 투여, 복막내 투여, 근육내 투여, 두개내 투여, 뇌척수 투여, 그리고 종양 또는 세포-증식성 질환의 부위에 직접 투여가 포함될 수 있다. 다른 투여 경로는 당분야에 공지되어 있다.

[0165] 따라서 본 발명에서는 세포 증식성 질환을 치료하는데 유용한 다양한 제약학적 조성물을 제시한다. 본 발명에 따른 제약학적 조성물은 담체, 부형제와 첨가제, 또는 보조제를 이용하여, 본 발명에 따른 세포 증식성 질환을 치료하거나 조절하는데 유용한 이종성 폴리뉴클레오티드 서열을 포함하는 레트로바이러스 벡터를 개체에 투여하기 적합한 형태로 만듦으로써 제조된다. 빈번하게 이용되는 담체 또는 보조제에는 탄산마그네슘, 이산화티타늄, 락토오스, 만니톨과 기타 당류, 활석, 우유 단백질, 젤라틴, 전분, 비타민, 셀룰로오스와 이의 유도체, 동물성과 식물성 오일, 폴리에틸렌 글리콜, 그리고 용매, 예를 들면, 무균수, 알코올, 글리세롤과 다가알코올이 포함된다. 정맥내 운반제에는 유체와 영양 보충물이 포함된다. 보존제에는 항균제, 항산화제, 킬레이트화제와 불활성 가스가 포함된다. 다른 제약학적으로 허용되는 담체에는 예로써, Remington's pharmaceutical Sciences, 15th ed. Easton: Mack Publishing Co., 1405-1412, 1461-1487 (1975), 그리고 The National Formulary XIV., 14th ed. Washington: American pharmaceutical Association (1975)에서 기술된 바와 같이, 수성 용액, 염을 비롯한 비-독성 부형제, 보존제, 완충제 등이 포함되고, 이들 문헌의 내용은 본 발명에 참조로서 편입된다. 제약학적 조성물의 다양한 성분의 pH와 정확한 농도는 당분야의 일과적인 기술에 따라 조정된다. Goodman and Gilman's The Pharmacological Basis for Therapeutics (7th ed.)를 참조한다.

[0166] 예로써, 그리고 제한 없이, 세포 증식성 질환을 치료하는데 유용한 레트로바이러스 벡터는 양쪽성 ENV 단백질, GAG, 그리고 POL 단백질, U3 영역 레트로바이러스 게놈 내에서 프로모터 서열, 그리고 레트로바이러스 게놈의 표적 세포 내로의 복제, 포장과 통합에 필요한 모든 cis-작용 서열을 포함할 것이다.

[0167] 아래의 실시예는 본 발명을 예시하는 것으로 의도되고 본 발명을 한정하지 않는다. 이들 실시예가 이용될 수 있는 것들의 전형이긴 하지만, 당업자에게 공지된 다른 절차가 대안으로 이용될 수도 있다.

도면의 간단한 설명

[0168] **도 1a-d**에서는 (a) 본 발명의 재조합 레트로바이러스 벡터의 개요; (b) 본 발명의 폴리뉴클레오티드의 플라스미드 지도; (c와 d) 본 발명의 폴리뉴클레오티드의 서열 (SEQ ID NO: 19)을 도시한다.

도 2a-d에서는 CD, OPRT와 UPRT 활성을 갖는 폴리펩티드를 포함하는 본 발명의 다양한 구체예의 산출을 위한 계획을 도시한다.

도 3에서는 감염된 U-87 세포에서 야생형 yCD 단백질과 비교하여 더욱 높은 수준의 yCD2 단백질이 관찰된다는 것을 보여준다.

도 4에서는 본 발명의 CD 폴리펩티드를 포함하는 벡터의 안정성, 그리고 본 발명의 다른 벡터와의 비교를 도시한다.

도 5a-b에서는 감염된 세포가 증가하는 수준의 5-FC에 노출될 때, 본 발명의 시토신 디아미나아제 활성과 벡터

가 야생형 yCD 활성 (T5.0007)과 비교하여 필적하는 또는 더욱 우수한 발현, 그리고 감염된 쥐 RG2(5A) 또는 U87 세포 (5B)의 사멸을 제공한다는 것을 보여준다.

도 6에서는 T5.0002 (yCD2)의 감염된 U87 세포에서 비활성이 T5.0007 (wt yCD)보다 큰 T5.0001 (부분적으로 최적화된 yCD)보다 크다는 것을 보여준다.

다양한 도면에서 유사한 참고 기호 (reference symbol)는 유사한 요소를 지시한다.

발명을 실시하기 위한 구체적인 내용

[0169] **실시예**

[0170] 실시예 1 변형된 CD 유전자의 작제 및 플라스미드 벡터 내로 삽입

[0171] 야생형 효모 시토신 디아미나아제 유전자에 유전적 강화는 (1) 효모 시토신 디아미나아제 단백질의 열 안정성을 증가시키기 위하여 3개의 아미노산 (A23L, I140L과 V108I)을 변화시키는 3가지 위치 돌연변이, 그리고 (2) 아미노산 서열에 대한 추가의 변화 없이, 인간 세포 내에서 단백질 번역 효율을 향상시키기 위하여 인간 코돈 사용 빈도 서열을 강화시키는 부가적인 유전자 서열 변형을 포함하도록 만들어졌다.

[0172] CD에 대한 서열 설계는 CD-최적화된, CD-UPRT (+/- 링커) 및 CD-OPRTase (+/- 링커)를 포함하였다. 최종 시토신 디아미나아제 코딩 서열은 PSI1/Not1 카세트 기초된 전략을 위한 5' 말단에서 PSI1 부위 (전장), 그리고 3' 말단에서 Not1 부위 + 폴리 A 꼬리를 포함할 수 있다.

[0173] 효모 시토신 디아미나아제를 포함하는 아래의 서열이 클로닝, 최적화 및 돌연변이에 이용되었다 (상자로 표시된 핵산은 클로닝을 위한 차후 방법에 이용된 제한 부위를 포함한다:

[0174]
```
AACACGATTATAAATGGTGACAGGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACA
TTGCCTATGAGGAGGCGGCCTTAGGTTACAAAGAGGGTGGTGTTCCTATTGGCGGATGTCTT
ATCAATAACAAAGACGGAAGTGTTCTCGGTCGTGGTCACAACATGAGATTTCAAAAGGGATC
CGCCACACTACATGGTGAGATCTCCACTTTGGAAAACTGTGGGAGATTAGAGGGCAAAGTGT
ACAAAGATACCACTTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATC
ATGTATGGTATTCCACGCTGTGTTGTCGGTGAGAACGTTAATTTCAAAAGTAAGGGCGAGAA
ATATTTACAAACTAGAGGTCACGAGGTTGTTGTTGTTGACGATGAGAGGTGTAAAAAGATCA
TGAAACAATTTATCGATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAGGCGGCC
GCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID
NO:30)
```

[0175] 아래의 표에서는 만들어진 유전자와 플라스미드 벡터, 그리고 이들의 명칭을 요약한다.

표 2

[0176] 벡터 구조체와 명칭

Tocagen Code	참고 명칭	본래 명칭	5'LTR Prom	외피	벡터	IRES	이식 유전자	3'LTR
T5.0000	pACE-yCD	pACE-CD (Tai et al. 2005)	CMV	Ampho (4070A)	pACE	EMCV	Wt 효모 CD	MLV U3
T5.0001	pAC3-yCD1	CDopt 서열	CMV	Ampho (4070A)	pAC3	EMCV	변형된 CD	MLV U3
T5.0002	pAC3-yCD2	CDopt+3pt	CMV	Ampho (4070A)	pAC3	EMCV	Modif-ied CD	MLV U3
T5.0003	pAC3-yCD2-U	Cdopt+3pt-UP RT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-UPRT	MLV U3
T5.0004	pAC3-yCD2-O	CDopt+3pt-OP RT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-OPRT	MLV U3
T5.0005	pAC3-yCD2-L O	CDopt+3pt-LI NK-OPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-L-OPR T	MLV U3
T5.0006	pAC3-eGFP	pAC3-emd, pAC3GFP	CMV	Ampho (4070A)	pAC3	EMCV	Emerald GFP	MLV U3
T5.0007	pAC3-yCD	pAC3-yCD	CMV	Ampho (4070A)	pAC3	EMCV	Wt 효모 CD	MLV U3

[0177] Kasahara *et al.* pACE-CD (U.S. Patent No. 6,899,871, 이의 내용은 본 발명에 참조로서 편입됨)에서 기술된 복제 적격성 레트로바이러스 벡터가 부가적인 변형을 위한 기초로서 이용되었다. 벡터 (pAC3-yCD)는 본 명세서에서 기술된 바와 같이 변형된 효모 시토신 디아미나아제 유전자를 발현하도록 변형되고 이들 구조체에 이용되었다. 최초 형질감염된 복제-적격성 레트로바이러스에 대한 벡터 구조체의 도면과 관련하여 하기 도 1A를 참조한다. CMV는 인간 CMV 극초기 프로모터이고, U3, R과 U5는 바이러스 긴 말단 반복 (LTR)의 상응하는 영역이다. *Gag, pol*과 *env*는 바이러스 단백질 코딩 영역이다. 도 1B와 1D에서는 본 발명의 플라스미드 구조와 서열을 도시한다. 이들 유전자는 계약업체 (Bio Basic Inc., Markham, Ontario, Canada)에서 합성된 이후, pAC3 벡터 골격 (도 1)의 Psi1-Not1 부위 내로 삽입되었다. 상기 플라스미드 골격은 Psi1과 Not1로 플라스미드 pAC3-eGFP를 절단하고 상기 플라스미드와 레트로바이러스 골격을 인코딩하는 대형 (대략 11kb) 단편을 정제함으로써 통상적으로 산출되었다.

[0178] A. 인간화된 코돈 최적화된 CD 유전자 (CD-opt, aka CD1, T5.0001).

[0179] Conrad *et al.* 및 PCT WO 99/60008의 인간 코돈 최적화된 시토신 디아미나아제의 비교에서, 91개의 전체 코돈이 둘 모두에서 최적화되고, 36개의 코돈이 동일하고, 47개의 코돈이 세 번째 염기쌍 변화를 보유하고 (이들 모두 동일한 아미노산을 인코딩한다), 그리고 9개의 코돈이 상이하였다 (하지만, 이들은 동일한 아미노산을 인코딩한다). 상이한 9개의 코돈:

[0180] AGC (Ser) 대(對) TCC (Ser)

[0181] CGT (Arg) 대(對) AGG (Arg)

[0182] CCA (Pro) 대(對) CCT (Pro)

[0183] 이들 모두 동등한 GC 함량을 갖고 동일한 아미노산을 인코딩한다. 상기 고유 효모 유전자 서열은 아래의 CD 유전자 (CD1)를 제공하기 위하여 별도로 코돈 최적화되고, 그리고 IRES와 함께 복제 적격성 레트로바이러스 (RCR)를 인코딩하는 플라스미드 벡터 pAC3 내로 삽입될 때 T5.0001로 명명되었다.

[0184] TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTA**
CGAGGAGGCCGCCCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACA
ACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACC
CTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGA
CACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACG
GCATCCCTAGGTGTGTGGTGGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTG
CAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGATCATGAAGCA
GTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTGATAAGCGGCCGCAG
ATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG. (SEQ ID NO:31)

[0185] B. 열 안정화된 CD 유전자

[0186] 시토신 디아미나아제의 안정성을 강화시키기 위하여 부가적인 변형이 만들어졌다. 야생형 효모 시토신 디아미나아제 유전자에 유전적 강화는 효모 시토신 디아미나아제 단백질의 열 안정성을 증가시키기 위하여 3개의 아미노산 (A23L, I140L과 V108I)을 변화시키는 3가지 위치 돌연변이를 포함하도록 만들어졌다.

[0187] 아래의 프라이머 쌍이 본 발명의 시토신 디아미나아제 폴리펩티드에 대한 유전자의 산출에 이용되었다:

[0188] 센스: 5'-tcgaggatatcggcgagtgaaacccgttattctttttggc-3' (SEQ ID NO: 32)

[0189] 안티센스: 5'-gccaaaaagaataacgggtttcactcgccgatatcctcga-3' (SEQ ID NO: 33)

[0190] 센스: 5'tcggcgagtgatccggcggcggcgcctccggcggcggcgcctccggcggcggcgcc tccggcggcggcgccaacccgttatt-3' (SEQ ID NO: 34)

[0191] 안티센스: 5'-aataacgggttggcgccgccgccggaggcgccgccgccggaggcgccgccgc cggaggcgccgccgccggatcactcgccga-3' (SEQ ID NO: 35)

[0193] 고유 효모 CD의 안정성을 증가시키기 위하여, 3가지 아미노산 치환이 상기 단백질 내로 조작되었다. 이들 치환은 단독으로, 또는 인간 코돈 최적화와 공동으로 수행되었다.

[0194] 3가지 아미노산 치환은 A23L, V108I, I140L이었다. 이들 치환을 인코딩하는 서열은 하기에 도시된다.

[0195]
```
ATGGTGACAGGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACATTGCCTATGAGGA
GGCGTTATTAGGTTACAAAGAGGGTGGTGTTCCTATTGGCGGATGTCTTATCAATAACAAAG
ACGGAAGTGTTCTCGGTCGTGGTCACAACATGAGATTTCAAAAGGGATCCGCCACACTACAT
GGTGAGATCTCCACTTTGGAAAACTGTGGGAGATTAGAGGGCAAAGTGTACAAAGATACCAC
TTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATCATGTATGGTATTC
CACGCTGTGTCATCGGTGAGAACGTTAATTTCAAAAGTAAGGGCGAGAAATATTTACAAACT
AGAGGTCACGAGGTTGTTGTTGTTGACGATGAGAGGTGTAAAAAGTTAATGAAACAATTTAT
CGATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAGGCGGCCGCGCCATAGATAA
AATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:36)
```

[0196] 인코딩된 폴리펩티드는 아래의 서열을 포함한다 (치환된 아미노산은 굵은 글씨-밑줄로 표시됨):

[0197] 1 MVTGGMASKWDQKGMDIAYEEALLGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSAT

[0199] * 61 LHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVIGENVNFKSKGEK

[0200] 121 YLQTRGHEVVVVDDERCKKLMKQFIDERPQDWFEDIGE-

[0202] 최종 구조체 설계는 선호되는 코돈을 이용하여 3가지 아미노산 치환 A23L/V108I/I140L을 통합하고, 그리고 전체 서열에 대한 선호되는 인간 코돈 사용빈도를 이용한다 (상기 유전자는 CDopt+3pt [aka CD2], 그리고 IRES와 함께 RCR을 인코딩하는 플라스미드 벡터 pAC3 내로 삽입될 때 T5.0002로 불린다) (SEQ ID NO: 37):

[0203] 1 ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTACGAG

[0204] 61 GAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAAC

[0205] 121 AAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACC

[0206] 181 CTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAG

[0207] 241 GACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATG

[0208] 301 TACGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAG

[0209] 361 TACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTG

[0210] 421 ATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTGATAA

[0211] 밑줄로 표시된 코돈은 아미노산 치환을 위한 선호되는 코돈을 나타낸다.

[0213] CD-최적화된 서열 설계 (인간 코돈 선호 + 3가지 아미노산 치환)

[0214]
```
AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATA
TCGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTG
ATCAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTC
CGCCACCCTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGT
ACAAGGACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATC
ATGTACGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAA
GTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGA
TGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTAAGCGGCC
GCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID
NO:38)
```

[0215] C. CD-UPRT 융합 유전자 (CDopt+3pt-UPRT, [aka CD-opt-UPRT와 CD2-UPRT], pAC3 플라스미드 RCR 벡터 내에서 T5.0003)의 작제.

[0216] 도 2A에 도시된 계획 I를 이용하여 CD-최적화된-UPRT 서열을 산출하기 위하여, UPRT 폴리펩티드에 연결된 앞서 기술된 바와 같은 CD 폴리펩티드를 포함하는 융합 구조체가 개발되었다. 아래의 프라이머가 CD와 UPRT 사이에 스톱-스타트 (stop-start)를 결실시키는데 이용되었다.

표 3

[0217] 프라이머 서열:

프라이머 명칭	프라이머 서열 (5'에서 3') (SEQ ID NO:)
del118-123	5'-tcgaggatatcggcgagtgaaacccgttattctttttggc-3'(39)
del118-123-안티센스	5'-gccaaaaagaataacgggtttcactcgccgatatcctcga-3'(40)

프라이머 명칭	**길이 (nt.)**	**Tm**	**68℃에서 이중나선 에너지**	**미스매치의 에너지 비용**
del118-123	40	79.06℃	-44.37 kcal/mole	21.1%
del118-123-안티센스	40	79.06℃	-47.95 kcal/mole	20.3%
프라이머 명칭	**프라이머-주형 이중나선**			
del118-123(각각, SEQ ID NO: 39와 41)	5'-tcgaggatatcggcgagtga------aacccgttattctttttggc-3'||||||||||||||||||||||||||||||||ccaagctcctatagccgctcactatctacttgggcaataagaaaaaccgaag			
del118-123-안티센스(각각, SEQ ID NO: 40와 42)	ggttcgaggatatcggcgagtgatagatgaacccgttattctttttggcttc||||||||||||||||||||||||||||||||3'-agctcctatagccgctcact------ttgggcaataagaaaaaccg			

[0218] 결과의 융합 폴리뉴클레오티드는 1296 bp 및 하기에 열거된 서열을 포함한다:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGAACCCGTTATTCTTTTTGGCTTCTCCATTCTTGTACCTTACATATCTTATATATTAT
CCAAACAAAGGGTCTTTCGTTAGCAAACCTAGAAATCTGCAAAAAATGTCTTCGGAACCATTTAAGAA
CGTCTACTTGCTACCTCAAACAAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAA
CTAGACCTGATTTCATTTTCTACTCCGATAGAATCATCAGATTGTTGGTTGAAGAAGGTTTGAACCAT
CTACCTGTGCAAAAGCAAATTGTGGAAACTGACACCAACGAAAACTTCGAAGGTGTCTCATTCATGGG
TAAAATCTGTGGTGTTTCCATTGTCAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTGTA
GGTCTGTGCGTATCGGTAAAATTTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTTATTCTAC
GAAAAATTACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCTGGCCACCGGTGG
TAGTGCTATCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGAGAGAATTTACTTCTTAA
ACCTAATCTGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTTCCCAGAGGTCAGAATTGTTACT
GGTGCCCTCGACAGAGGTCTAGATGAAAACAAGTATCTAGTTCCAGGGTTGGGTGACTTTGGTGACAG
ATACTACTGTGTTTAA**GCGGCCGC**GCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGG
[0219] GGGG (SEQ ID NO:43)

[0220] D. CD-링커 UPRT 융합 유전자 (CDopt+3pt-LINK-UPRT [aka CD-opt-링커-UPRT와 CD2-L-UPRT])의 작제.

[0221] 도 2B에 도시된 계획 II를 이용하여 CD-최적화된-링커-UPRT 서열을 산출하기 위하여, CD 폴리펩티드와 UPRT 폴리펩티드 사이에, 그리고 이들과 인-프레임 (in frame)으로 링커 $(Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_4$ 도메인을 클로닝함으로써 융합 구조체가 개발되었다. 아래의 프라이머가 상기 링커를 삽입하는데 이용되었다.

표 4

[0222]

프라이머 명칭	**프라이머 서열 (5'에서 3') (SEQ ID NO:)**			
ins_60nt_after_477	5'-tcggcgagtgatccggcggcggcgcctccggcggcggcgcctccggcg gcggcgcctccggcggcggcgccaacccgttatt-3' (44)			
ins_60nt_after_477-안티센스	5'-aataacgggttggcgccgccgccggaggcgccgccgccggaggcgcc gccgccggaggcgccgccgccggatcactcgccga-3' (45)			
프라이머 명칭	**길이 (nt.)**	**Tm**	**68℃에서 이중나선 에너지**	**미스매치의 에너지 비용**
ins_60nt_after_477	82	79.77℃	-30.19 kcal/mole	83.3%
ins_60nt_after_477-안티센스	82	79.77℃	-32.31 kcal/mole	82.2%

[0223] 결과의 구조체는 1356 bp 크기 및 하기에 열거된 서열을 포함한다:

AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGC
GCCAACCCGTTATTCTTTTTGGCTTCTCCATTCTTGTACCTTACATATCTTATATATTATCCAAACAA
AGGGTCTTTCGTTAGCAAACCTAGAAATCTGCAAAAAATGTCTTCGGAACCATTTAAGAACGTCTACT
TGCTACCTCAAACAAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAACTAGACCT
GATTTCATTTTCTACTCCGATAGAATCATCAGATTGTTGGTTGAAGAAGGTTTGAACCATCTACCTGT
GCAAAAGCAAATTGTGGAAACTGACACCAACGAAAACTTCGAAGGTGTCTCATTCATGGGTAAAATCT
GTGGTGTTTCCATTGTCAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTGTAGGTCTGTG
CGTATCGGTAAAATTTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTTATTCTACGAAAAATT
ACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCTGGCCACCGGTGGTAGTGCTA
TCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGAGAGAATTTACTTCTTAAACCTAATC
TGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTTCCCAGAGGTCAGAATTGTTACTGGTGCCCT
CGACAGAGGTCTAGATGAAAACAAGTATCTAGTTCCAGGGTTGGGTGACTTTGGTGACAGATACTACT
GTGTTTAAGCGGCCGCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG
[0224] (SEQ ID NO:46)

[0225] E. CD-OPRT 융합 유전자 (CDopt+3pt-OPRT [aka CDopt-OPRT와 CD2-OPRT], pAC3 플라스미드 RCR 벡터 내로 삽입될 때 T5.0004)의 작제

[0226] 도 2C에 도시된 계획 III을 이용하여 CD-최적화된-OPRTase (CD 인간화된 + 3 pt 돌연변이 + OPRTase 기능적 도메인 인간)을 산출하기 위하여, 앞서 기술된 바와 같이 OPRT 폴리펩티드에 연결된 CD 폴리펩티드를 포함하는 융합 구조체가 개발되었다.

[0227] 결과의 구조체는 1269 bp 크기 및 하기에 열거된 서열을 포함한다:

AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAG**GCGGTCGCTCGTG**cagctttggggccattggtgacgggtctgtacgacgtgcaggct
[0228] ttcaagtttggggacttcgtgctgaagagcgggctttcctccccccatctacatcgatctgcggggcat

cgtgtctcgaccgcgtcttctgagtcaggttgcagatattttattccaaactgcccaaaatgcaggca
tcagttttgacaccgtgtgtggagtgccttatacagctttgccattggctacagttatctgttcaacc
aatcaaattccaatgcttattagaaggaaagaaacaaaggattatggaactaagcgtcttgtagaagg
aactattaatccaggagaaacctgtttaatcattgaagatgttgtcaccagtggatctagtgttttgg
aaactgttgaggttcttcagaaggagggcttgaaggtcactgatgccatagtgctgttggacagagag
cagggaggcaaggacaagttgcaggcgcacgggatccgcctccactcagtgtgtacattgtccaaaat
gctggagattctcgagcagcagaaaaaagttgatgctgagacagttgggagagtgaagaggtttattc
aggagaatgtctttgtggcagcgaatcataatggttctccccttttctataaaggaagcacccaaagaa
ctcaGCTTCGGTGCACGTGCAGAGCTGCCCAGGATCCACCCAGTTGCATCGAAGTAAGCGGCCGCGCC
[0229] ATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:47)

[0230] F. CD-링커-OPRT 융합 유전자 (CDopt+3pt-LINK-OPRT, [aka CDopt-링커-OPRT와 CD2-L-OPRT], pAC3 플라스미드 RCR 벡터 내에서 T5.0005)의 작제

[0231] 도 2D에 도시된 계획 IV를 이용하여 CD-최적화된-링커-OPRT 서열을 산출하기 위하여, CD 폴리펩티드와 OPRT 폴리펩티드 사이에, 그리고 이들과 인-프레임 (in frame)으로 링커 $(Ser\text{-}Gly\text{-}Gly\text{-}Gly)_4$ 도메인을 클로닝함으로써 융합 구조체가 개발되었다.

[0232] 결과의 구조체는 1329 bp 크기 및 하기에 열거된 서열을 포함한다:

AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGC
GCCGCGGTCGCTCGTGcagctttggggccattggtgacgggtctgtacgacgtgcaggctttcaagtt
tggggacttcgtgctgaagagcgggctttcctcccccatctacatcgatctgcggggcatcgtgtctc
gaccgcgtcttctgagtcaggttgcagatattttattccaaactgcccaaaatgcaggcatcagtttt
gacaccgtgtgtggagtgccttatacagctttgccattggctacagttatctgttcaaccaatcaaat
tccaatgcttattagaaggaaagaaacaaaggattatggaactaagcgtcttgtagaaggaactatta
atccaggagaaacctgtttaatcattgaagatgttgtcaccagtggatctagtgttttggaaactgtt
gaggttcttcagaaggagggcttgaaggtcactgatgccatagtgctgttggacagagagcagggagg
caaggacaagttgcaggcgcacgggatccgcctccactcagtgtgtacattgtccaaaatgctggaga
ttctcgagcagcagaaaaaagttgatgctgagacagttgggagagtgaagaggtttattcaggagaat
gtctttgtggcagcgaatcataatggttctccccttttctataaaggaagcacccaaagaactcaGCTT
CGGTGCACGTGCAGAGCTGCCCAGGATCCACCCAGTTGCATCGAAGTAAGCGGCCGCGCCATAGATAA
[0233] AATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG. (SEQ ID NO:48)

[0234] **실시예 2 감염성 벡터 생산**

[0235] 벡터는 다양한 방법으로 생산될 수 있지만, 첫 번째 단계는 차후에, 세포 상층액으로부터 수확될 수 있는 감염성 입자의 생산을 가능하게 하기 위하여 플라스미드 DNA 벡터를 세포 내로 도입하는 것이다. 일단 감염성 입자가 산출되면, 당업자에 의해 다른 생산 방법이 수행될 수 있다. 벡터 입자는 293T 세포의 일시적인 형질감염에 의해 산출되었다 (Pear *et al*. Proc Natl Acad Sci U S A. 90:8392-8396 1993). 이들 293T 세포는 해동되고 배양액으로 집어넣어지고, 이후 DMEM 높은 글루코오스 배지 (Hyclone# 30081, 500 mℓ)를 FBS (Hyclone# SH30070, 50 mℓ), L-글루타민 (Cellgro# 25-005-CI, 5 mℓ), NEAA (Hyclone #SH30238, 5 mℓ), 그리고 페니실린-스트렙 (Cellgro# 30-002-CI, 5 mℓ)과 혼합함으로써 제조된 15 mℓ의 DMEM 배지를 내포하는 T-75 플라스크에서 2회 계대되었다. 이들 플라스크는 37℃와 5% CO2에서 배양되었다. 세 번째 계대후, 세포는 각각 5 mℓ의 배지를 내포하는 6개의 T-25 플라스크에서, 1.8 x 10^6개 세포/T-25 (또는 7.2 x 10^4개 세포/cm^2)의 세포 밀도로 접종되었다. T-25 플라스크에 접종하고 하루후, 이들 세포는 인산칼슘 형질감염 키트 (Promega, Cat# E1200)를 이용하여, 상기 바이러스 벡터를 발현하는 T5.0002 플라스미드로 형질감염되었다. 형질감염후 18시간 시점에, 한 세트의 플라스크 (각 세트마다 3개의 플라스크)에서 배지가 10 mM NaB를 포함하는 새로운 배지로 대체되었다. 두 번째 세트의 플라스크에서 배지는 대체되지 않고, 대조 (제로 NaB)로서 기능하였다. NaB 처리후 8시간 시점에, 모든 플라스크에서 배지는 NaB를 포함하지 않는 새로운 배지로 대체되었다. 양쪽 세트의 플라스크에 대하여, 익일 (22시간 지속)까지 발현이 지속되었다. 양쪽 세트의 플라스크로부터 상층액은 수확되고 형질도입 단위 (TU)/mℓ로 표시된 qPCR에 의해 그들의 역가에 대하여 분석되었다 (실시예 3 참조).

[0237] 역가 결과는 하기 표에서 제시된다.

표 5

[0238]

조건	첫 번째 역가	두 번째 역가 (-80℃에서 68일 동안 보관후)
NaB 없음	1.5 (±0.05) x 10^6 TU/mℓ	1.2 (±0.2) x 10^6 TU/mℓ
10 mM NaB	1.4 (±0.3) x 10^6 TU/mℓ	7.0 (±0.14) x 10^5 TU/mℓ

[0239] 차후 벡터 제조물은 부티르산나트륨 없이, 이러한 방식으로 생산되었다. 다른 벡터 플라스미드는 10e5 TU/mℓ 내지 10e7 TU/mℓ의 역가를 갖는 벡터 제조물을 산출하기 위하여 동일한 방식으로 이용되었다. 이런 물질은 원하는 경우에, 하기에 기술된 바와 같이 더욱 정제되고 농축될 수 있다 (참조: US 5792643; T. Rodriguez et al. J Gene Med 9:233 2007; US Application 61218063). 본 발명의 일정한 구체예에서, 용량은 뇌 무게의 그램으로 계산되었다. 이런 구체예에서, 치료 방법에 유용한 본 발명의 복제 적격성 레트로바이러스 벡터의 용량은 뇌 무게 그람당 10^3 내지 10^7 TU의 범위일 수 있다.

[0241] **실시예 3 정량적 PCR 적정 분석**

[0242] 기능적 벡터 농도, 또는 역가는 정량적 PCR-기초된 (qPCR) 방법을 이용하여 결정된다. 이러한 방법에서, 벡터는 형질도입가능 숙주 세포주 (가령, PC-3 인간 전립선 암종 세포, ATCC Cat# CRL-1435)를 표준 부피의 벡터로 감염시키고, 그리고 형질도입후 숙주 세포 내에 존재하는 프로바이러스의 생산량을 측정함으로써 적정된다. 이들 세포와 벡터는 벡터 복제를 중단시키기 위한 항-레트로바이러스 AZT의 첨가에 앞서 완전한 감염이 가능하도록 표준 배양 조건 (37℃, 5% CO2) 하에 24 시간 동안 배양된다. 그 다음, 이들 세포는 배양 접시로부터 수확되고, 그리고 게놈 DNA (gDNA)는 Invitrogen Purelink gDNA 정제 키트를 이용하여 정제되고 무균 RNase-/DNase-없는 물로 정제 칼럼으로부터 용출된다. A260/A280 흡광도 비율 (absorbance ratio)은 샘플의 농도와 상대적 순도를 결정하기 위하여 분광광도계에서 측정된다. gDNA 농도는 qPCR에 대한 입력 DNA가 분석된 모든 샘플에서 일정하도록 부가적인 RNase-/DNase-없는 물로 최소 농도의 소정의 gDNA 제조물 세트에 표준화되었다. 게놈 DNA 순도는 에티듐 브롬화물 염색된 0.8% 아가로즈 겔 상에서 각 샘플의 분취량 (aliquot)의 전기영동에 의해 더욱 평가된다. 샘플이 1.8-2.0의 A260/A280 흡광도 범위를 추월하고 단일 띠의 gDNA를 보이면, 상기 샘플은 벡터의 프로바이러스 사본수 (copy number)의 qPCR 분석이 가능하다. 프로바이러스의 LTR 영역을 통합하는 프라이머 (역-전사된 벡터 DNA 및 숙주 gDNA 내로 통합되는 벡터 DNA)를 이용하여, 공지된 숫자의 세포를 형질도입하기 위하여 공지된 부피의 벡터가 이용될 때 발생하는 형질도입 사건의 총수를 산정하기 위한 qPCR이 수행된다. 반응당 형질도입 사건의 총수는 1E7에서 1E1 사본으로 연속 희석되고 상기 샘플과 동일한 qPCR 조건 하에 측정되는 공지된 사본수의 표적-운반 플라스미드를 이용하는 표준 곡선으로부터 계산된다. 얼마나 많은 게놈 등가물이 각 qPCR 반응 (앞서 결정된 농도로부터)에 이용되는 지, 그리고 반응당 얼마나 많은 형질도입 사건이 발생하는 지를 확인한 이후, 형질도입의 시점에 존재했던 세포의 총수에 기초하여 발생했던 형질도입 사건의 총수를 결정한다. 이러한 값은 최초 형질도입 동안 이들 세포를 내포하는 배지 내로 희석 이후에 벡터의 역가이다. 정정된 역가 값을 계산하기 위하여, 역가의 부피로 나눗셈된 배양액의 부피와 역가의 부피로 곱셈함으로써 희석도가 정정된다. 이들 실험은 복제 배양액에서 수행되고, 그리고 평균 역가 및 이의 연관된 표준 편차 (standard deviation)와 변동 계수 (coefficient of variance)를 결정하기 위하여 각 조건에 대한 삼중 척도를 이용한 qPCR에 의해 분석된다.

[0244] **실시예 4 웨스턴 블롯에 의해 측정된 발현 수준**

[0245] 도 3에서는 야생형 (ACE-yCD) 또는 완전 최적화된 (AC3-yCD2) 시토신 디아미나아제 유전자를 인코딩하는 바이러스로 감염된 U-87 세포의 웨스턴 블롯 분석에서, 야생형 yCD 단백질과 비교하여, 더욱 높은 안정성을 위한 3가지 돌연변이를 보유하는 인간 코돈 최적화된 단백질이 더욱 높은 수준으로 관찰된다는 것을 증명한다.

[0247] **실시예 5 바이러스 벡터의 유전적 안정성.**

[0248] 역전사 및 처리된 세포 내로 첫 번째 통합 사건 이후, DNA 프로바이러스 및 임의의 후손 레트로바이러스는 어느 한쪽 말단에서 MLV로부터 전통적인 LTR 구조를 보유하는 것으로 인식된다. 이러한 배치는 복합 감염 주기 이후에도 안정한 것으로 밝혀졌다 (하기 도 4 참조).

[0249] 대략 10^6개 고유 U-87 세포가 0.01의 MOI에서 바이러스 벡터로 최초 감염되고, 그리고 단일 감염 주기를 완결하기 위하여 완전 감염 때까지 성장되었다. 상층액은 이후, 감염되지 않은 세포에 다시 통과되고, 그리고 이러한 주기가 반복되었다. yCD2 서열의 게놈 안정성은 이식유전자 삽입 부위에 접하는 MLV 특이적 프라이머를 이용한, 감염된 세포로부터 통합된 프로바이러스의 PCR 증폭에 의해 평가되었다. 각 세트의 감염에서, 겔 상에서 전장 앰플리콘 크기를 추적하기 위하여 벡터 플라스미드 (pAC3-yCD2와 the Kasahara *et al.* 벡터 pACE-CD)의 증폭 역시 수행되었다. 전장 앰플리콘 보다 작은 임의의 띠의 출현은 벡터 불안정성의 지표일 것이다. 이들 실험은 본 발명의 벡터 (T5.0002 - 변형된 벡터 및 CDopt+3pt (CD2) 이종성 폴리뉴클레오티드를 포함)가 야생형 효모를 보유하는 pACE-CD 또는 T5.007보다 더욱 많은 횟수의 계대 동안 안정성을 유지한다는 것을 증명하였다.

[0251] **실시예 6 세포 사멸 실험**

[0252] 시험관내 세포 배양액 실험에서, 이들 실험은 pAC3-yCD2/T5.0002 [AC3-yCD2(V)], pAC3-yCD/T5.0007 [AC3-yCD(V)] 또는 기타 벡터로부터 만들어진 바이러스로 감염된 RG2 쥐 세포 (도 5A) 또는 U-87 세포 (도 5B)에서 5-FC 적정을 수행함으로써, yCD2 단백질을 발현하는 세포에서 시토신 디아미나아제가 야생형 yCD 단백질을 발현하는 세포로부터 시토신 디아미나아제와 최소한 동등한 활성을 갖는다는 것을 증명한다. 간단히 말하면, U-87 세포의 경우에, AC3-yCD (야생형 CD) 벡터 또는 AC3-yCD2 (열안정화된 & 코돈 최적화된) 벡터로 0.1의 감염비 (multiplicity of infection) (즉, 100% 감염)에서 감염후 5일 시점에, 이들 세포는 8일 동안 양성 대조로서 증가하는 양의 5-FC 또는 0.1 mM의 5-FU에 종속되었다. 5-FC 처리의 8일 시점에, 세포 배양액은 MTS 분석

(Promega CellTiter 96 AQUEOUS One Solution 증식 분석)을 이용하여 생존능에 대하여 평가되었다. 데이터는 증가하는 용량의 5-FC 처리에서 이들 두 레트로바이러스 벡터 사이에 필적하는 사멸을 증명한다. RG2 배양액은 유사하게 처리되었고, 그리고 T5.0002CD 발현 분석으로부터, 바이러스에 대한 더욱 낮은 농도의 5-FC를 향하여 사멸 곡선에서 약간의 이동을 보인다. U-87 세포는 0.1의 감염비 (MOI)에서 형질도입되고, 바이러스가 확산하도록 5일 동안 배양되고, 그리고 형질도입후 5일 시점부터 세포가 수확되었다. 이들 세포는 이후, 800 x g에서 5분 동안 원심분리로 수확되었다. 상층액은 세포 펠릿으로부터 흡출되고, 5 ㎖의 인산염 완충액 (PBS)으로 세척되고, 800 x g에서 5분 동안 다시 원심분리되었다. 결과의 세포 펠릿은 1.5 ㎖의 PBS에 집어넣어지고, 피펫 팁을 통한 통과에 의해 재현탁되고, -20℃에서 냉동 장치에 배치되었다. 세포는 냉동/해동 방법으로 용해되었다. 앞서 재현탁된 세포는 실온에서 해동되고, 피펫 팁에 통과되고, 프로테아제 저해물질 칵테일과 혼합되고, -20℃에서 다시 동결되었다. 효소 분석에 앞서, 샘플은 실온에서 다시 해동되고 피펫 팁에 통과되었다. 현탁액은 이후, 14,000 rpm에서 5분 동안 탁상용 원심분리기에서 원심분리되었다. 상층액은 펠릿으로부터 이동되고 새로운 에펜도르프 튜브 (eppendorf tube)에 넣어지고 얼음 위에 배치되었다. yCD 효소 활성은 HPLC 분석을 이용하여 평가되었다.

[0253] HPLC 분석은 광어레이 검출기 (photoarray detector)와 자동주입기 (autoinjector)와 일련으로 연결된 Shimadzu LC20AT 회로에서 수행되었다. 고체상은 5 ㎛ 구 크기와 4.0 x 250 mm 칼럼 치수를 갖는 Hypersil BDS C18 HPLC 칼럼이었다. 이동상은 0.01% tert-부틸암모늄 과염소산염과 5% 메탄올을 내포하는 50 mM 인산암모늄, pH 2.1이었다; 상기 시스템은 22℃에서 평형화되었다. 모든 시약은 ACS 등급이고, 그리고 용매는 HPLC 등급이었다. 반응 혼합물은 PBS에서 0.125 ㎎/㎖ 5-FC (1 mM)의 최종 농도에서 800 ㎕로 구성되도록 만들어지고 1.5 ㎖ 자동샘플주입기 바이알에 넣어졌다. 반응은 이후, 200 ㎕의 각 세포 용해물을 첨가함으로써 개시되었다. 반응물 / 자동샘플주입기 바이알은 자동샘플주입기 내에 배치되고 5 ㎕의 반응 혼합물이 주입되었다. 각 반응 바이알로부터 5 ㎕ 분취량을 회수하고 HPLC 칼럼에서 분석함으로써 시점 (time point)이 주기적으로 획득되었다. 5-FC의 5-FU로의 전환 속도는 5-FU의 앞서 산출된 표준 곡선으로부터 공지된 양과 피크 면적을 비교함으로써 계산되었다. 5-FU로 5-FC 전환의 속도는 상응하는 시간 간격에 대하여 산출된 5-FU의 양 (nmol)을 플롯팅 (plotting)함으로써 추론되었다. 세포 샘플에 대한 단백질 농도는 추론되고, 그리고 세포 용해물 샘플의 비활성은 분석에 이용되는 단백질의 양 (㎎)으로 상기 전환 속도 (nmol/min)를 나눗셈함으로써 계산되었다. 도 6에서는 0.1의 MOI에서 형질도입후 5일 시점에, 다양한 벡터의 비활성을 도시한다. 상기 데이터는 조직 배양 세포 내에서 시토신 디아미나아제의 비활성의 측면에서 pACE-yCD (T5.0000) < pAC3-yCD1 (T5.0001) < pAC3-CD2 (T5.0002)임을 증명한다.

[0254] **실시예 7. 완전히 변형된 CD 유전자 (yCD2)로 치료된 종양은 변형되지 않은** 효모 CD 유전자로 치료된 종양보다 더욱 효과적으로 제거되고 재발하지 않는다 (PR-01-08-001).

[0255] 어떤 벡터 구조체가 피하 생쥐/인간 이종이식편 모형에서 가장 효과적인 벡터를 발생시키는 지를 결정하기 위하여, 3가지 상이한 구조체가 평가되었다: T5.0001 (부분적으로 변형된 CD); T5.0007 (변형되지 않은 효모 CD 유전자); T5.0002 (완전히 변형된 yCD2 유전자). 5-FC 전구약물 투여후 종양 성장, 생존과 종양 퇴보가 면역결핍 생쥐에서 인간 신경교종 (U-87)의 피하 모형에서 평가되었다. 5-FC와 종양-벡터 구조체 사이에서 관찰되는 용량 반응 (dose response)을 결정하기 위하여 2가지 상이한 5-FC 농도가 평가되었다. 군당 9-11마리 암컷 생쥐로 구성되는 총 12개 군이 조사되었다. 모든 생쥐는 0일자에, 감염되지 않은 U-87 종양 세포주의 98% 혼합물 및 3가지 TOCA 511 벡터 구조체, T5.0002, T5.0001, T5.0007 중에서 한 가지로 감염된 U-87 세포주의 2% 혼합물; 또는 감염되지 않은 대조 U-87 세포주 (100%)가 오른쪽 옆구리의 등쪽에 이식되었다. 생쥐는 2x10e6개 세포/생쥐로 접종되었다. 군 1-3에서 생쥐는 감염되지 않은 U-87 세포를 보유하고, 군 4-6은 형질도입된 T5.0002 U-87을 포함하는 U-87 세포 혼합물을 보유하고, 군 7-9는 형질도입된 T5.0001 U-87을 포함하는 U-87 세포 혼합물을 보유하고, 그리고 군 10-12는 형질도입된 T5.0007 U-87을 포함하는 U-87 세포 혼합물을 보유하였다. 종양은 종양 크기가 대략 100㎜^{3}이 될 때까지 6일 동안 성장되었다. 생쥐의 각 벡터 투약 군은 6일자부터 단일 IP 주사로서 2가지 용량의 5-FC (200 또는 500 ㎎/㎏/day) 중에서 한 가지, 또는 5-FC 없음이 무작위로 투여되었다. 5-FC 투여는 28일 연속으로 매일 지속되었다. 29일후 생존하는 생쥐는 종양 크기에 대하여 평가되고, 이후 분석을 위하여 종양 (존재하면)이 채취되는 서로 다른 일자에서 희생되었다. 29일자에, 군 4-12로부터 생쥐는 다시 무작위화되고, 그리고 이들 생쥐는 2가지 하위군 (종양 퇴화를 모니터하기 위하여 5-FC 치료가 지속되거나, 또는 종양 재성장을 모니터하기 위하여 5-FC 치료가 중단됨)으로 세분되었다.

[0256] ***결과:*** T5.0002, T5.0001과 T5.0007로 치료된 종양은 모두 29일자에, 200과 500 ㎎/㎏ 용량에서 검출되지 않은 정도까지 종양 퇴보를 보였다. T5.0002 치료된 동물 모두 5-FC 치료가 중단될 때, 39일자까지 종양이 재발하지

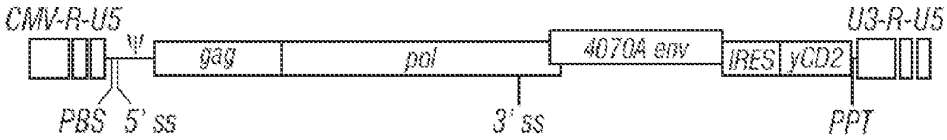
않았다. 대조적으로, T5.0001과 T5.0007로 치료된 일부 종양은 39일 시점에 재발하였다. 실험은 39일자에 종결되었다.

[0257] ***고찰:*** T5.0002 + 5-FC는 T5.0001 또는 T5.0007 plu 5-FC보다 더욱 효과적인 항종양 요법이다.

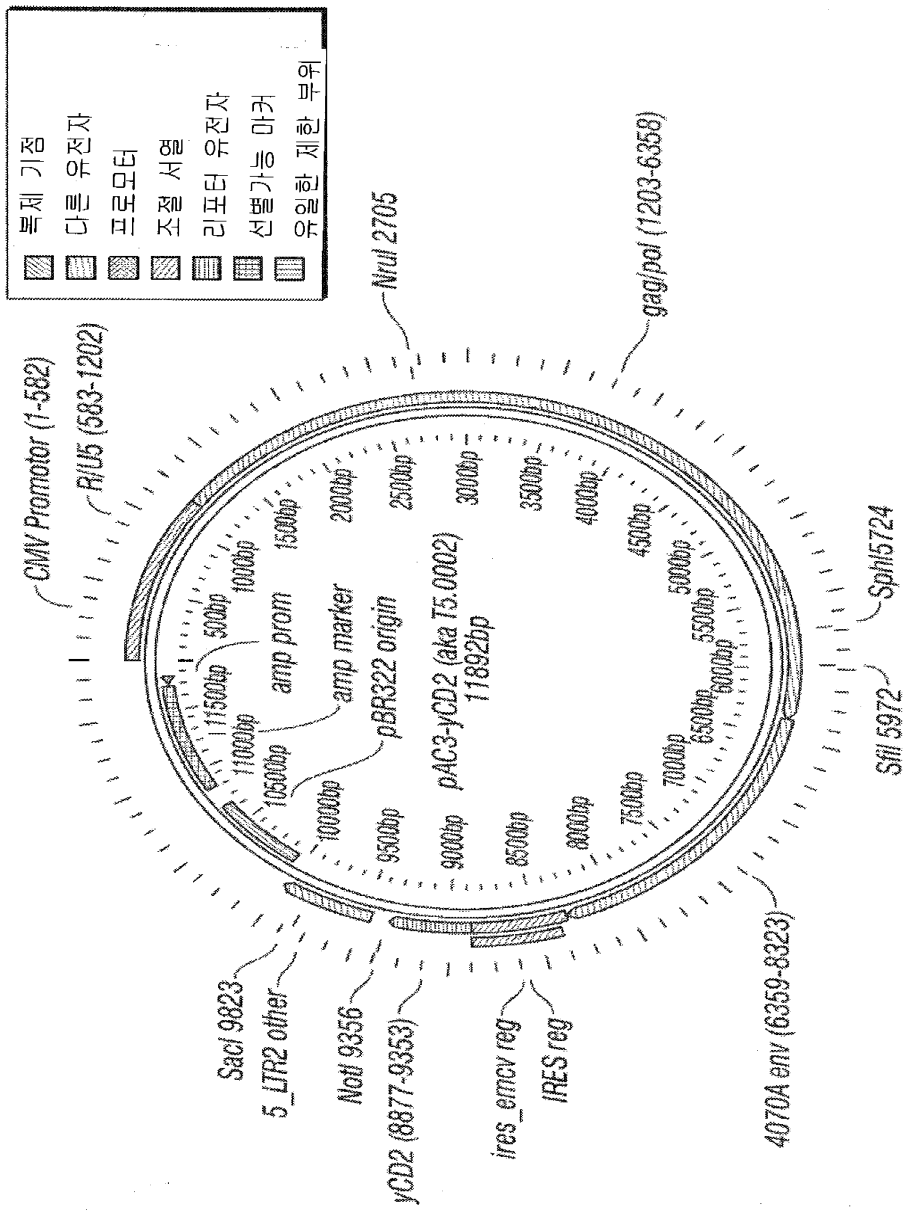
[0258] 본 발명의 다수의 구체예가 기술되었다. 그럼에도 불구하고, 본 발명의 기술적 사상과 범위를 벗어나지 않는 다양한 개변이 가능하다. 따라서 다른 구체예 역시 하기의 특허청구범위 내에 있다.

도면

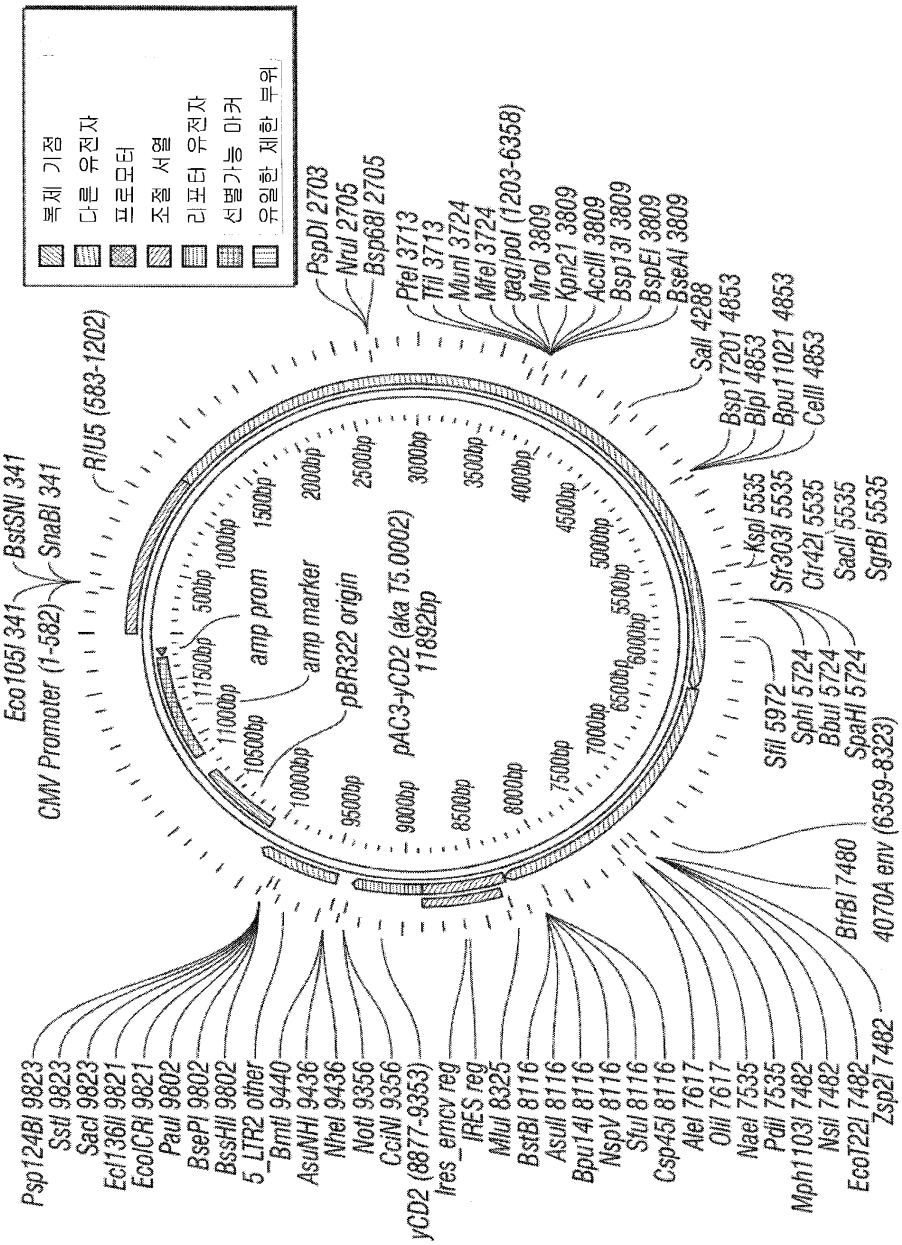
도면1a



도면1b



도면1c


복제 기점
다른 유전자
프로모터
조절 서열
리포터 유전자
선별가능 마커
유일한 제한 부위
pAC3-yCD2 (aka T5.0002)
11892bp
amp prom
amp marker
pBR322 origin
R/U5 (583-1202)
CMV Promoter (1-582)
Eco105I 341
BstSNI 341
SnaBI 341
PspDI 2703
NruI 2705
Bsp68I 2705
PfeI 3713
TfiI 3713
MunI 3724
MfeI 3724
gag/pol (1203-6358)
MroI 3809
Kpn21 3809
AccIII 3809
Bsp13I 3809
BspEI 3809
BseAI 3809
SalI 4288
Bsp17201 4853
BlpI 4853
Bpu11021 4853
CelII 4853
KspI 5535
Sfr303I 5535
Clr42I 5535
SacII 5535
SgrBI 5535
SfiI 5972
SphI 5724
BbuI 5724
SpaHI 5724
BfrBI 7480
4070A env (6359-8323)
Zsp2I 7482
EcoT22I 7482
NsiI 7482
Mph1103I 7482
PdiI 7535
NaeI 7535
OliI 7617
AleI 7617
Csp45I 8116
SfuI 8116
NspV 8116
Bpu14I 8116
AsuII 8116
BstBI 8116
MluI 8325
IRES reg
Ires_emcv reg
yCD2 (8877-9353)
CciNI 9356
NotI 9356
NheI 9436
AsuNHI 9436
BmtI 9440
5_LTR2 other
BssHII 9802
BsePI 9802
PauI 9802
Eco/CRI 9821
Ecl136II 9821
SacI 9823
SstI 9823
Psp124BI 9823

도면1da


TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAAT
GGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGG
CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTA
GTGAACCGGCGCCAGTCCTCCGATTGACTGAGTCGCCCGGGTACCCGTGTATCCAATAAACCCTCTTGCAGTTGCATCCGAC
TTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCATTTGGGGGCTCGT
CCGGGATCGGGAGACCCCTGCCCAGGGACCACCGACCCACCACCGGGAGGTAAGCTGGCCAGCAACTTATCTGTGTCTGTCC
GATTGTCTAGTGTCTATGACTGATTTTATGCGCCTGCGTCGGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGGACCCGTG
GTGGAACTGACGAGTTCGGAACACCCGGCCGCAACCCTGGGAGACGTCCCAGGGACTTCGGGGGCCGTTTTTGTGGCCCGAC
CTGAGTCCAAAAATCCCGATCGTTTTGGACTCTTTGGTGCACCCCCCTTAGAGGAGGGATATGTGGTTCTGGTAGGAGACGA
GAACCTAAAACAGTTCCCGCCTCCGTCTGAATTTTTGCTTTCGGTTTGGGACCGAAGCCGCGCCGCGCGTCTTGTCTGCTGC
AGCATCGTTCTGTGTTGTCTCTGTCTGACTGTGTTTCTGTATTTGTCTGAGAATATGGGCCAGACTGTTACCACTCCCTTAA
GTTTGACCTTAGGTCACTGGAAAGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTAC
CTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTT
AAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTG
ACCCCCCTCCCTGGGTCAAGCCCTTTGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGA
ACCTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGCGCCAAACCTAAACCTCAAGTT
CTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTACAGAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCG
ACAGGGACGGAAATGGTGGAGAAGCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCTACGTGGGAG
ACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAGGAGGAAACGGACAGCTTCAATACTGG
CCGTTCTCCTCTTCTGACCTTTACAACTGGAAAAATAATAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGA
TCGAGTCTGTTCTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCGGAGAAGAAAA
ACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGGCGCCCCACTCAACTGCCCAATGAAGTCGATGCC
GCTTTTCCCCTCGAGCGCCCAGACTGGGATTACACCACCCAGGCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCC
TAGCGGGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAAGGGCCCAATGAGTCTCC
CTCGGCCTTCCTAGAGAGACTTAAGGAAGCCTATCGCAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAAT
GTGTCTATGTCTTTCATTTGGCAGTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGC
TTGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAGAGAGGAACGTATCAGGAGAGA
AACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGATGAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATG
AGCAAGCTATTGGCCACTGTCGTTAGTGGACAGAAACAGGATAGACAGGGAGGAGAACGAAGGAGGTCCCAACTCGATCGCG
ACCAGTGTGCCTACTGCAAAGAAAAGGGGCACTGGGCTAAAGATTGTCCCAAGAAACCACGAGGACCTCGGGGACCAAGACC
CMV promoter
R
U5
GT = 5'ss
gag

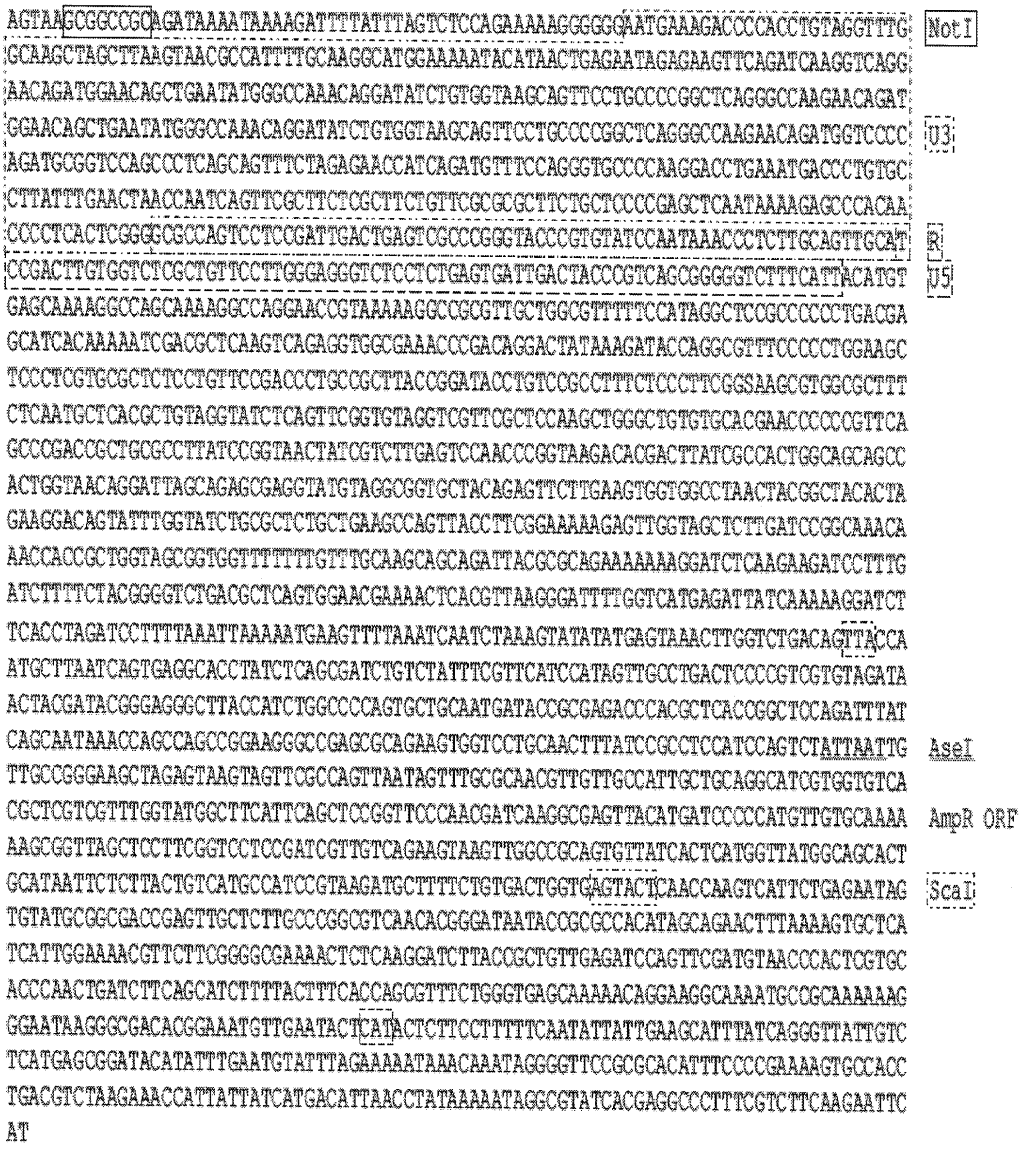
도면1db



도면1dc



도면1dda



도면1ddb

```
      CMV Promotor (1-582)>>>
      |
1     TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCG 60
      ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGC

61    CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT 120
      GCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAA

121   GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA 180
      CTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGT

181   ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCC 240
      TACCCACCTCATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGG

241   AAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA 300
      TTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT

                                            Eco105I
                                            |
                                            SnaBI
                                            |
                                            BstSNI
                                            |
301   CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC 360
      GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATG

361   CATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGG 420
      GTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCC

421   ATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG 480
      TAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGC

481   GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGT 540
      CCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACA

                                             R region (583-650)>>>
                                             |
541   ACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGGCGCCAGTCCTCCGATTG 600
      TGCCACCCTCCAGATATATTCGTCTCGACCAAATCACTTGGCCGCGGTCAGGAGGCTAAC
```

도면1de

```
                                                U5 region (651-1202)>>>
                                                |
601   ACTGAGTCGCCCGGGTACCCGTGTATCCAATAAACCCTCTTGCAGTTGCATCCGACTTGT 660
      TGACTCAGCGGGCCCATGGGCACATAGGTTATTTGGGAGAACGTCAACGTAGGCTGAACA

661   GGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTCAGCGGGGGTC 720
      CCAGAGCGACAAGGAACCCTCCCAGAGGAGACTCACTAACTGATGGGCAGTCGCCCCCAG

721   TTTCATTTGGGGGCTCGTCCGGGATCGGGAGACCCCTGCCCAGGGACCACCGACCCACCA 780
      AAAGTAAACCCCCGAGCAGGCCCTAGCCCTCTGGGGACGGGTCCCTGGTGGCTGGGTGGT

            5'SS (788)
            |
781   CCGGGAGGTAAGCTGGCCAGCAACTTATCTGTGTCTGTCCGATTGTCTAGTGTCTATGAC 840
      GGCCCTCCATTCGACCGGTCGTTGAATAGACACAGACAGGCTAACAGATCACAGATACTG

841   TGATTTTATGCGCCTGCGTCGGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGGACCCG 900
```

도면1df

```
      ACTAAAATACGCGGACGCAGCCATGATCAATCGATTGATCGAGACATAGACCGCCTGGGC

901   TGGTGGAACTGACGAGTTCGGAACACCCGGCCGCAACCCTGGGAGACGTCCCAGGGACTT 960
      ACCACCTTGACTGCTCAAGCCTTGTGGGCCGGCGTTGGGACCCTCTGCAGGGTCCCTGAA

961   CGGGGGCCGTTTTTGTGGCCCGACCTGAGTCCAAAAATCCCGATCGTTTTGGACTCTTTG 1020
      GCCCCCGGCAAAAACACCGGGCTGGACTCAGGTTTTTAGGGCTAGCAAAACCTGAGAAAC

1021  GTGCACCCCCCTTAGAGGAGGGATATGTGGTTCTGGTAGGAGACGAGAACCTAAAACAGT 1080
      CACGTGGGGGGAATCTCCTCCCTATACACCAAGACCATCCTCTGCTCTTGGATTTTGTCA

1081  TCCCGCCTCCGTCTGAATTTTTGCTTTCGGTTTGGGACCGAAGCCGCGCCGCGCGTCTTG 1140
      AGGGCGGAGGCAGACTTAAAAACGAAAGCCAAACCCTGGCTTCGGCGCGGCGCGCAGAAC

1141  TCTGCTGCAGCATCGTTCTGTGTTGTCTCTGTCTGACTGTGTTTCTGTATTTGTCTGAGA 1200
      AGACGACGTCGTAGCAAGACACAACAGAGACAGACTGACACAAAGACATAAACAGACTCT

        gag(1203,2819)>>>
        |
1201  ATATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGTCG 1260
      TATACCCGGTCTGACAATGGTGAGGGAATTCAAACTGGAATCCAGTGACCTTTCTACAGC

1261  AGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGCT 1320
      TCGCCTAGCGAGTGTTGGTCAGCCATCTACAGTTCTTCTCTGCAACCCAATGGAAGACGA

1321  CTGCAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGACC 1380
      GACGTCTTACCGGTTGGAAATTGCAGCCTACCGGCGCTCTGCCGTGGAAATTGGCTCTGG

1381  TCATCACCCAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGG 1440
      AGTAGTGGGTCCAATTCTAGTTCCAGAAAAGTGGACCGGGCGTACCTGTGGGTCTGGTCC

1441  TCCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCT 1500
      AGGGGATGTAGCACTGGACCCTTCGGAACCGAAAACTGGGGGGAGGGACCCAGTTCGGGA

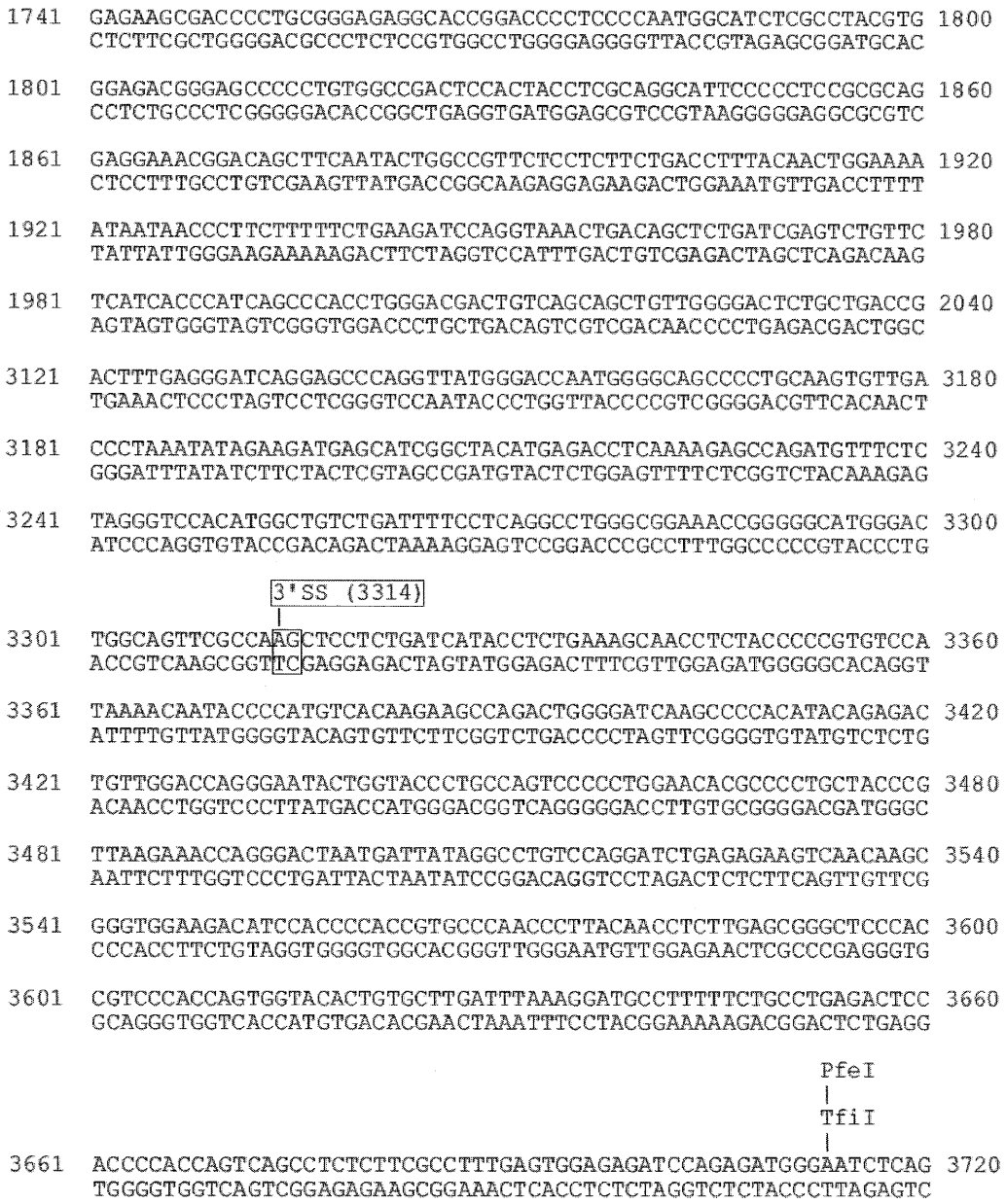
1501  TTGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAAC 1560
      AACATGTGGGATTCGGAGGCGGAGGAGAAGGAGGTAGGCGGGGCAGAGAGGGGGAACTTG

1561  CTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGCG 1620
      GAGGAGCAAGCTGGGGCGGAGCTAGGAGGGAAATAGGTCGGGAGTGAGGAAGAGATCCGC

1621  CCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTACAG 1680
      GGTTTGGATTTGGAGTTCAAGAAAGACTGTCACCCCCCGGCGAGTAGCTGGATGAATGTC

1681  AAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGTG 1740
      TTCTGGGGGGCGGAATATCCCTGGGTTCTGGTGGGGGAAGGCTGTCCCTGCCTTTACCAC
```

도면1dg


1741 GAGAAGCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCTACGTG 1800
CTCTTCGCTGGGGACGCCCTCTCCGTGGCCTGGGGAGGGGTTACCGTAGAGCGGATGCAC
1801 GGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCATTCCCCCTCCGCGCAG 1860
CCTCTGCCCTCGGGGGACACCGGCTGAGGTGATGGAGCGTCCGTAAGGGGGAGGCGCGTC
1861 GAGGAAACGGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAAA 1920
CTCCTTTGCCTGTCGAAGTTATGACCGGCAAGAGGAGAAGACTGGAAATGTTGACCTTTT
1921 ATAATAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGTCTGTTC 1980
TATTATTGGGAAGAAAAAGACTTCTAGGTCCATTTGACTGTCGAGACTAGCTCAGACAAG
1981 TCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACCG 2040
AGTAGTGGGTAGTCGGGTGGACCCTGCTGACAGTCGTCGACAACCCCTGAGACGACTGGC
3121 ACTTTGAGGGATCAGGAGCCCAGGTTATGGGACCAATGGGGCAGCCCCTGCAAGTGTTGA 3180
TGAAACTCCCTAGTCCTCGGGTCCAATACCCTGGTTACCCCGTCGGGGACGTTCACAACT
3181 CCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTC 3240
GGGATTTATATCTTCTACTCGTAGCCGATGTACTCTGGAGTTTTCTCGGTCTACAAAGAG
3241 TAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGGGGCATGGGAC 3300
ATCCCAGGTGTACCGACAGACTAAAAGGAGTCCGGACCCGCCTTTGGCCCCCGTACCCTG
3'SS (3314)
3301 TGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCA 3360
ACCGTCAAGCGGTTCGAGGAGACTAGTATGGAGACTTTCGTTGGAGATGGGGGCACAGGT
3361 TAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGAC 3420
ATTTTGTTATGGGGTACAGTGTTCTTCGGTCTGACCCCTAGTTCGGGGTGTATGTCTCTG
3421 TGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCCG 3480
ACAACCTGGTCCCTTATGACCATGGGACGGTCAGGGGGACCTTGTGCGGGGACGATGGGC
3481 TTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGC 3540
AATTCTTTGGTCCCTGATTACTAATATCCGGACAGGTCCTAGACTCTCTTCAGTTGTTCG
3541 GGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCAC 3600
CCCACCTTCTGTAGGTGGGGTGGCACGGGTTGGGAATGTTGGAGAACTCGCCCGAGGGTG
3601 CGTCCCACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCC 3660
GCAGGGTGGTCACCATGTGACACGAACTAAATTTCCTACGGAAAAAGACGGACTCTGAGG
PfeI
TfiI
3661 ACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCAG 3720
TGGGGTGGTCAGTCGGAGAGAAGCGGAAACTCACCTCTCTAGGTCTCTACCCTTAGAGTC

도면1dh

```
          MfeI
          |
          MunI
          |
3721  GACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTGATG 3780
      CTGTTAACTGGACCTGGTCTGAGGGTGTCCCAAAGTTTTTGTCAGGGTGGGACAAACTAC

                                  MroI
                                  |
                                  BseAI
                                  |
                                  Bsp13I
                                  |
                                  BspEI
                                  |
                                  Kpn2I
                                  |
                                  AccIII
                                  |
3781  AGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGATCCTGCTAC 3840
      TCCGTGACGTGTCTCTGGATCGTCTGAAGGCCTAGGTCGTGGGTCTGAACTAGGACGATG

3841  AGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTC 3900
      TCATGCACCTACTGAATGACGACCGGCGGTGAAGACTCGATCTGACGGTTGTTCCATGAG
3121  ACTTTGAGGGATCAGGAGCCCAGGTTATGGGACCAATGGGGCAGCCCCTGCAAGTGTTGA 3180
      TGAAACTCCCTAGTCCTCGGGTCCAATACCCTGGTTACCCCGTCGGGGACGTTCACAACT

3181  CCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTC 3240
      GGGATTTATATCTTCTACTCGTAGCCGATGTACTCTGGAGTTTTCTCGGTCTACAAAGAG

3241  TAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGGGGCATGGGAC 3300
      ATCCCAGGTGTACCGACAGACTAAAAGGAGTCCGGACCCGCCTTTGGCCCCCGTACCCTG

                  3'SS (3314)
                  |
3301  TGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCA 3360
      ACCGTCAAGCGGTTCGAGGAGACTAGTATGGAGACTTTCGTTGGAGATGGGGGCACAGGT

3361  TAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGAC 3420
      ATTTTGTTATGGGGTACAGTGTTCTTCGGTCTGACCCCTAGTTCGGGGTGTATGTCTCTG

3421  TGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCCG 3480
      ACAACCTGGTCCCTTATGACCATGGGACGGTCAGGGGGACCTTGTGCGGGGACGATGGGC

3481  TTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGC 3540
      AATTCTTTGGTCCCTGATTACTAATATCCGGACAGGTCCTAGACTCTCTTCAGTTGTTCG
```

도면1di

```
3541  GGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCAC 3600
      CCCACCTTCTGTAGGTGGGGTGGCACGGGTTGGGAATGTTGGAGAACTCGCCCGAGGGTG

3601  CGTCCCACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCC 3660
      GCAGGGTGGTCACCATGTGACACGAACTAAATTTCCTACGGAAAAAGACGGACTCTGAGG

                                                          PfeI
                                                          |
                                                          TfiI
                                                          |
3661  ACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCAG 3720
      TGGGGTGGTCAGTCGGAGAGAAGCGGAAACTCACCTCTCTAGGTCTCTACCCTTAGAGTC

         MfeI
         |
         MunI
         |
3721  GACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGTTTGATG 3780
      CTGTTAACTGGACCTGGTCTGAGGGTGTCCCAAAGTTTTTGTCAGGGTGGGACAAACTAC

                                MroI
                                |
                                BseAI
                                |
                                Bsp13I
                                |
                                BspEI
                                |
                                Kpn2I
                                |
                                AccIII
                                |
3781  AGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGATCCTGCTAC 3840
      TCCGTGACGTGTCTCTGGATCGTCTGAAGGCCTAGGTCGTGGGTCTGAACTAGGACGATG

3841  AGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTC 3900
      TCATGCACCTACTGAATGACGACCGGCGGTGAAGACTCGATCTGACGGTTGTTCCATGAG
```

도면1dj

```
3901  GGGCCCTGTTACAAACCCTAGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAA 3960
      CCCGGGACAATGTTTGGGATCCCTTGGAGCCCATAGCCCGGAGCCGGTTCTTTCGGGTTT

3961  TTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTGA 4020
      AAACGGTCTTTGTCCAGTTCATAGACCCCATAGAAGATTTTCTCCCAGTCTCTACCGACT

4021  CTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCCTCGACAACTAA 4080
      GACTCCGGTCTTTTCTCTGACACTACCCCGTCGGATGAGGCTTCTGGGGAGCTGTTGATT

4081  GGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCCCTGGGTTTGCAGAAATGG 4140
      CCCTCAAGGATCCCTGCCGTCCGAAGACAGCGGAGACCTAGGGACCCAAACGTCTTTACC

4141  CAGCCCCCTTGTACCCTCTCACCAAAACGGGGACTCTGTTTAATTGGGGCCCAGACCAAC 4200
      GTCGGGGGAACATGGGAGAGTGGTTTTGCCCCTGAGACAAATTAACCCCGGGTCTGGTTG

4201  AAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAG 4260
      TTTTCCGGATAGTTCTTTAGTTCGTTCGAGAAGATTGACGGGGTCGGGACCCCAACGGTC

                                  SalI
                                   |
4261  ATTTGACTAAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTCC 4320
      TAAACTGATTCGGGAAACTTGAGAAACAGCTGCTCTTCGTCCCGATGCGGTTTCCACAGG

4321  TAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCTAGACC 4380
      ATTGCGTTTTTGACCCTGGAACCGCAGCCGGCCACCGGATGGACAGGTTTTTCGATCTGG

4381  CAGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATTGCCGTACTGACAA 4440
      GTCATCGTCGACCCACCGGGGGAACGGATGCCTACCATCGTCGGTAACGGCATGACTGTT

4441  AGGATGCAGGCAAGCTAACCATGGGACAGCCACTAGTCATTCTGGCCCCCCATGCAGTAG 4500
      TCCTACGTCCGTTCGATTGGTACCCTGTCGGTGATCAGTAAGACCGGGGGGTACGTCATC

4501  AGGCACTAGTCAAACAACCCCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATC 4560
      TCCGTGATCAGTTTGTTGGGGGGCTGGCGACCGAAAGGTTGCGGGCCTACTGAGTGATAG

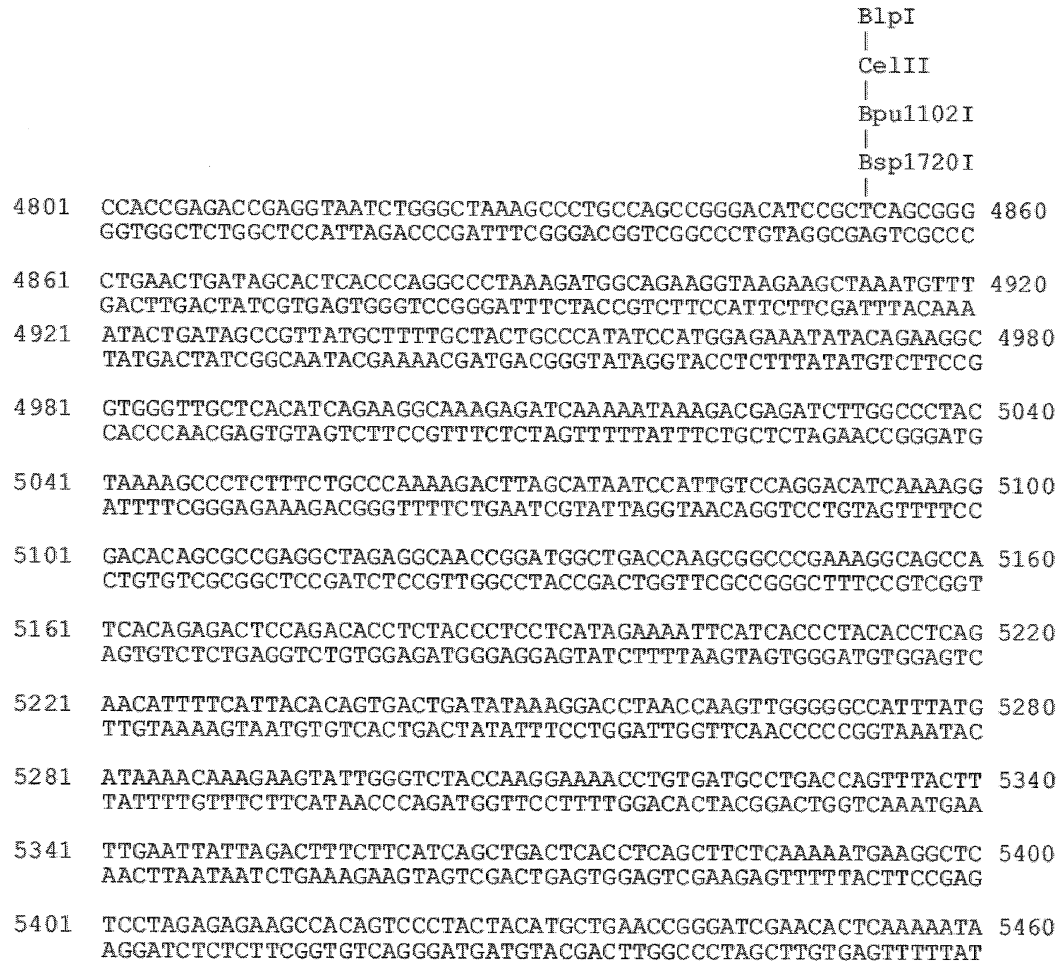
4561  AGGCCTTGCTTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCGG 4620
      TCCGGAACGAAAACCTGTGCCTGGCCCAGGTCAAGCCTGGCCACCATCGGGACTTGGGCC

4621  CTACGCTGCTCCCACTGCCTGAGGAAGGGCTGCAACACAACTGCCTTGATATCCTGGCCG 4680
      GATGCGACGAGGGTGACGGACTCCTTCCCGACGTTGTGTTGACGGAACTATAGGACCGGC

4681  AAGCCCACGGAACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACCACACCT 4740
      TTCGGGTGCCTTGGGCTGGGCTGGATTGCCTGGTCGGCGAGGGTCTGCGGCTGGTGTGGA

4741  GGTACACGGATGGAAGCAGTCTCTTACAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGA 4800
      CCATGTGCCTACCTTCGTCAGAGAATGTTCTCCCTGTCGCATTCCGCCCTCGACGCCACT
```

도면1dk


BlpI
CelII
Bpu1102I
Bsp1720I
4801 CCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGG 4860
GGTGGCTCTGGCTCCATTAGACCCGATTTCGGGACGGTCGGCCCTGTAGGCGAGTCGCCC
4861 CTGAACTGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAATGTTT 4920
GACTTGACTATCGTGAGTGGGTCCGGGATTTCTACCGTCTTCCATTCTTCGATTTACAAA
4921 ATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGAAATATACAGAAGGC 4980
TATGACTATCGGCAATACGAAAACGATGACGGGTATAGGTACCTCTTTATATGTCTTCCG
4981 GTGGGTTGCTCACATCAGAAGGCAAAGAGATCAAAAATAAAGACGAGATCTTGGCCCTAC 5040
CACCCAACGAGTGTAGTCTTCCGTTTCTCTAGTTTTTATTTCTGCTCTAGAACCGGGATG
5041 TAAAAGCCCTCTTTCTGCCCAAAAGACTTAGCATAATCCATTGTCCAGGACATCAAAAGG 5100
ATTTTCGGGAGAAAGACGGGTTTTCTGAATCGTATTAGGTAACAGGTCCTGTAGTTTTCC
5101 GACACAGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGCCA 5160
CTGTGTCGCGGCTCCGATCTCCGTTGGCCTACCGACTGGTTCGCCGGGCTTTCCGTCGGT
5161 TCACAGAGACTCCAGACACCTCTACCCTCCTCATAGAAAATTCATCACCCTACACCTCAG 5220
AGTGTCTCTGAGGTCTGTGGAGATGGGAGGAGTATCTTTTAAGTAGTGGGATGTGGAGTC
5221 AACATTTTCATTACACAGTGACTGATATAAAGGACCTAACCAAGTTGGGGGCCATTTATG 5280
TTGTAAAAGTAATGTGTCACTGACTATATTTCCTGGATTGGTTCAACCCCCGGTAAATAC
5281 ATAAAACAAAGAAGTATTGGGTCTACCAAGGAAAACCTGTGATGCCTGACCAGTTTACTT 5340
TATTTTGTTTCTTCATAACCCAGATGGTTCCTTTTGGACACTACGGACTGGTCAAATGAA
5341 TTGAATTATTAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAAAATGAAGGCTC 5400
AACTTAATAATCTGAAAGAAGTAGTCGACTGAGTGGAGTCGAAGAGTTTTTACTTCCGAG
5401 TCCTAGAGAGAAGCCACAGTCCCTACTACATGCTGAACCGGGATCGAACACTCAAAAATA 5460
AGGATCTCTCTTCGGTGTCAGGGATGATGTACGACTTGGCCCTAGCTTGTGAGTTTTTAT

도면1d1

```
5461  TCACTGAGACCTGCAAAGCTTGTGCACAAGTCAACGCCAGCAAGTCTGCCGTTAAACAGG 5520
      AGTGACTCTGGACGTTTCGAACACGTGTTCAGTTGCGGTCGTTCAGACGGCAATTTGTCC

                    SacII
                    |
                    SgrBI
                    |
                    Cfr42I
                    |
                    Sfr303I
                    |
                    KspI
                    |
5521  GAACTAGGGTCCGCGGGCATCGGCCCGGCACTCATTGGGAGATCGATTTCACCGAGATAA 5580
      CTTGATCCCAGGCGCCCGTAGCCGGGCCGTGAGTAACCCTCTAGCTAAAGTGGCTCTATT

5581  AGCCCGGATTGTATGGCTATAAATATCTTCTAGTTTTTATAGATACCTTTTCTGGCTGGA 5640
      TCGGGCCTAACATACCGATATTTATAGAAGATCAAAAATATCTATGGAAAAGACCGACCT

5641  TAGAAGCCTTCCCAACCAAGAAAGAAACCGCCAAGGTCGTAACCAAGAAGCTACTAGAGG 5700
      ATCTTCGGAAGGGTTGGTTCTTTCTTTGGCGGTTCCAGCATTGGTTCTTCGATGATCTCC

                              PaeI
                              |
                              BbuI
                              |
                              SpaHI
                              |
                              SphI
                              |
5701  AGATCTTCCCCAGGTTCGGCATGCCTCAGGTATTGGGAACTGACAATGGGCCTGCCTTCG 5760
      TCTAGAAGGGGTCCAAGCCGTACGGAGTCCATAACCCTTGACTGTTACCCGGACGGAAGC

6901  CTTCCAAGGGGCTACTCGAGGGGGCAGATGCAACCCTCTAGTCCTAGAATTCACTGATGC 6960
      GAAGGTTCCCCGATGAGCTCCCCCGTCTACGTTGGGAGATCAGGATCTTAAGTGACTACG

6961  AGGAAAAAAGGCTAACTGGGACGGGCCCAAATCGTGGGGACTGAGACTGTACCGGACAGG 7020
      TCCTTTTTTCCGATTGACCCTGCCCGGGTTTAGCACCCCTGACTCTGACATGGCCTGTCC

7021  AACAGATCCTATTACCATGTTCTCCCTGACCCGGCAGGTCCTTAATGTGGGACCCCGAGT 7080
      TTGTCTAGGATAATGGTACAAGAGGGACTGGGCCGTCCAGGAATTACACCCTGGGGCTCA

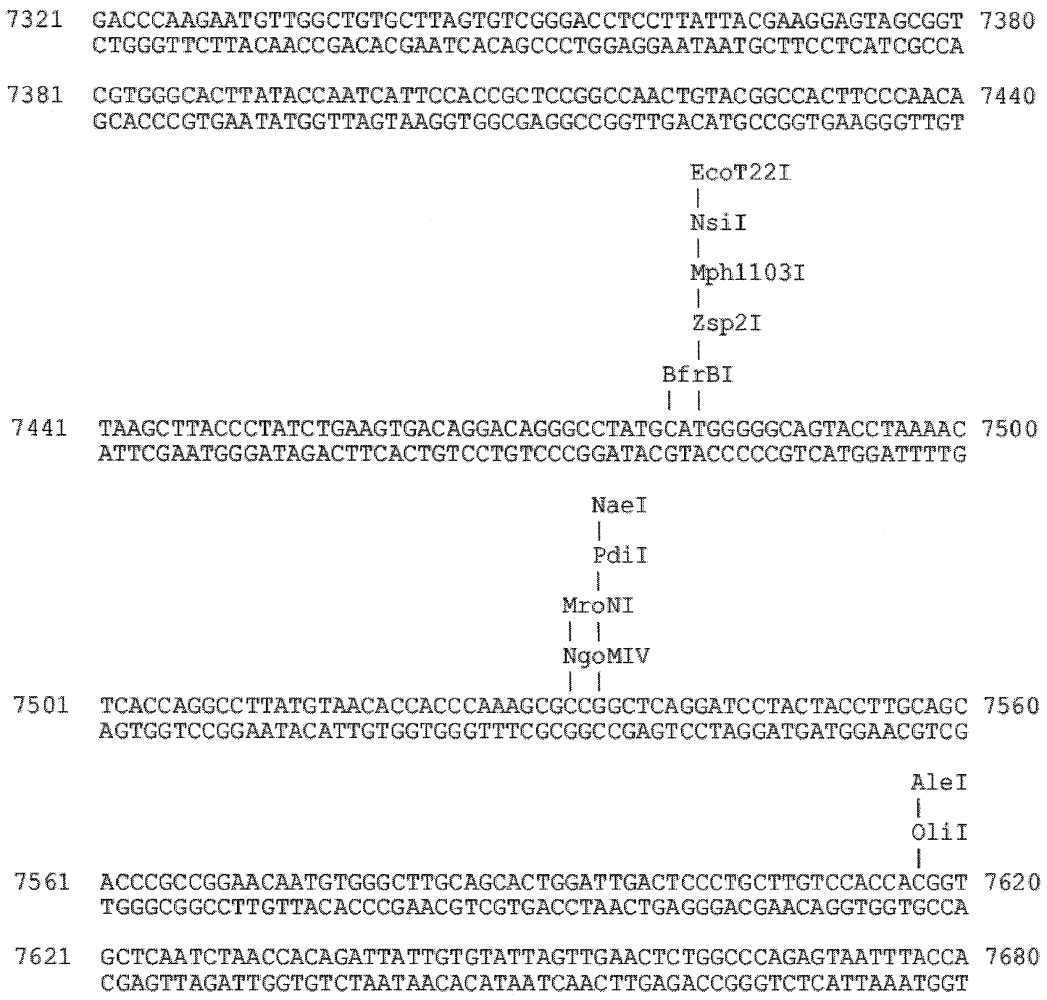
7081  CCCCATAGGGCCCAACCCAGTATTACCCGACCAAAGACTCCCTTCCTCACCAATAGAGAT 7140
      GGGGTATCCCGGGTTGGGTCATAATGGGCTGGTTTCTGAGGGAAGGAGTGGTTATCTCTA

7141  TGTACCGGCTCCACAGCCACCTAGCCCCCTCAATACCAGTTACCCCCCTTCCACTACCAG 7200
      ACATGGCCGAGGTGTCGGTGGATCGGGGGGAGTTATGGTCAATGGGGGGAAGGTGATGGTC

7201  TACACCCTCAACCTCCCCTACAAGTCCAAGTGTCCCACAGCCACCCCCAGGAACTGGAGA 7260
      ATGTGGGAGTTGGAGGGGATGTTCAGGTTCACAGGGTGTCGGTGGGGGTCCTTGACCTCT

7261  TAGACTACTAGCTCTAGTCAAAGGAGCCTATCAGGCGCTTAACCTCACCAATCCCGACAA 7320
      ATCTGATGATCGAGATCAGTTTCCTCGGATAGTCCGCGAATTGGAGTGGTTAGGGCTGTT
```

도면1dm


7321 GACCCAAGAATGTTGGCTGTGCTTAGTGTCGGGACCTCCTTATTACGAAGGAGTAGCGGT 7380
CTGGGTTCTTACAACCGACACGAATCACAGCCCTGGAGGAATAATGCTTCCTCATCGCCA
7381 CGTGGGCACTTATACCAATCATTCCACCGCTCCGGCCAACTGTACGGCCACTTCCCAACA 7440
GCACCCGTGAATATGGTTAGTAAGGTGGCGAGGCCGGTTGACATGCCGGTGAAGGGTTGT
EcoT22I
NsiI
Mph1103I
Zsp2I
BfrBI
7441 TAAGCTTACCCTATCTGAAGTGACAGGACAGGGCCTATGCATGGGGGCAGTACCTAAAAC 7500
ATTCGAATGGGATAGACTTCACTGTCCTGTCCCGGATACGTACCCCCGTCATGGATTTTG
NaeI
PdiI
MroNI
NgoMIV
7501 TCACCAGGCCTTATGTAACACCACCCAAAGCGCCGGCTCAGGATCCTACTACCTTGCAGC 7560
AGTGGTCCGGAATACATTGTGGTGGGTTTCGCGGCCGAGTCCTAGGATGATGGAACGTCG
AleI
OliI
7561 ACCCGCCGGAACAATGTGGGCTTGCAGCACTGGATTGACTCCCTGCTTGTCCACCACGGT 7620
TGGGCGGCCTTGTTACACCCGAACGTCGTGACCTAACTGAGGGACGAACAGGTGGTGCCA
7621 GCTCAATCTAACCACAGATTATTGTGTATTAGTTGAACTCTGGCCCAGAGTAATTTACCA 7680
CGAGTTAGATTGGTGTCTAATAACACATAATCAACTTGAGACCGGGTCTCATTAAATGGT

도면1dn

```
6901  CTTCCAAGGGGCTACTCGAGGGGGCAGATGCAACCCTCTAGTCCTAGAATTCACTGATGC 6960
      GAAGGTTCCCCGATGAGCTCCCCCGTCTACGTTGGGAGATCAGGATCTTAAGTGACTACG

6961  AGGAAAAAAGGCTAACTGGGACGGGCCCAAATCGTGGGGACTGAGACTGTACCGGACAGG 7020
      TCCTTTTTTCCGATTGACCCTGCCCGGGTTTAGCACCCCTGACTCTGACATGGCCTGTCC

7021  AACAGATCCTATTACCATGTTCTCCCTGACCCGGCAGGTCCTTAATGTGGGACCCCGAGT 7080
      TTGTCTAGGATAATGGTACAAGAGGGACTGGGCCGTCCAGGAATTACACCCTGGGGCTCA

7081  CCCCATAGGGCCCAACCCAGTATTACCCGACCAAAGACTCCCTTCCTCACCAATAGAGAT 7140
      GGGGTATCCCGGGTTGGGTCATAATGGGCTGGTTTCTGAGGGAAGGAGTGGTTATCTCTA

7141  TGTACCGGCTCCACAGCCACCTAGCCCCCTCAATACCAGTTACCCCCCTTCCACTACCAG 7200
      ACATGGCCGAGGTGTCGGTGGATCGGGGGAGTTATGGTCAATGGGGGGAAGGTGATGGTC

7201  TACACCCTCAACCTCCCCTACAAGTCCAAGTGTCCCACAGCCACCCCCAGGAACTGGAGA 7260
      ATGTGGGAGTTGGAGGGGATGTTCAGGTTCACAGGGTGTCGGTGGGGGTCCTTGACCTCT

7261  TAGACTACTAGCTCTAGTCAAAGGAGCCTATCAGGCGCTTAACCTCACCAATCCCGACAA 7320
      ATCTGATGATCGAGATCAGTTTCCTCGGATAGTCCGCGAATTGGAGTGGTTAGGGCTGTT

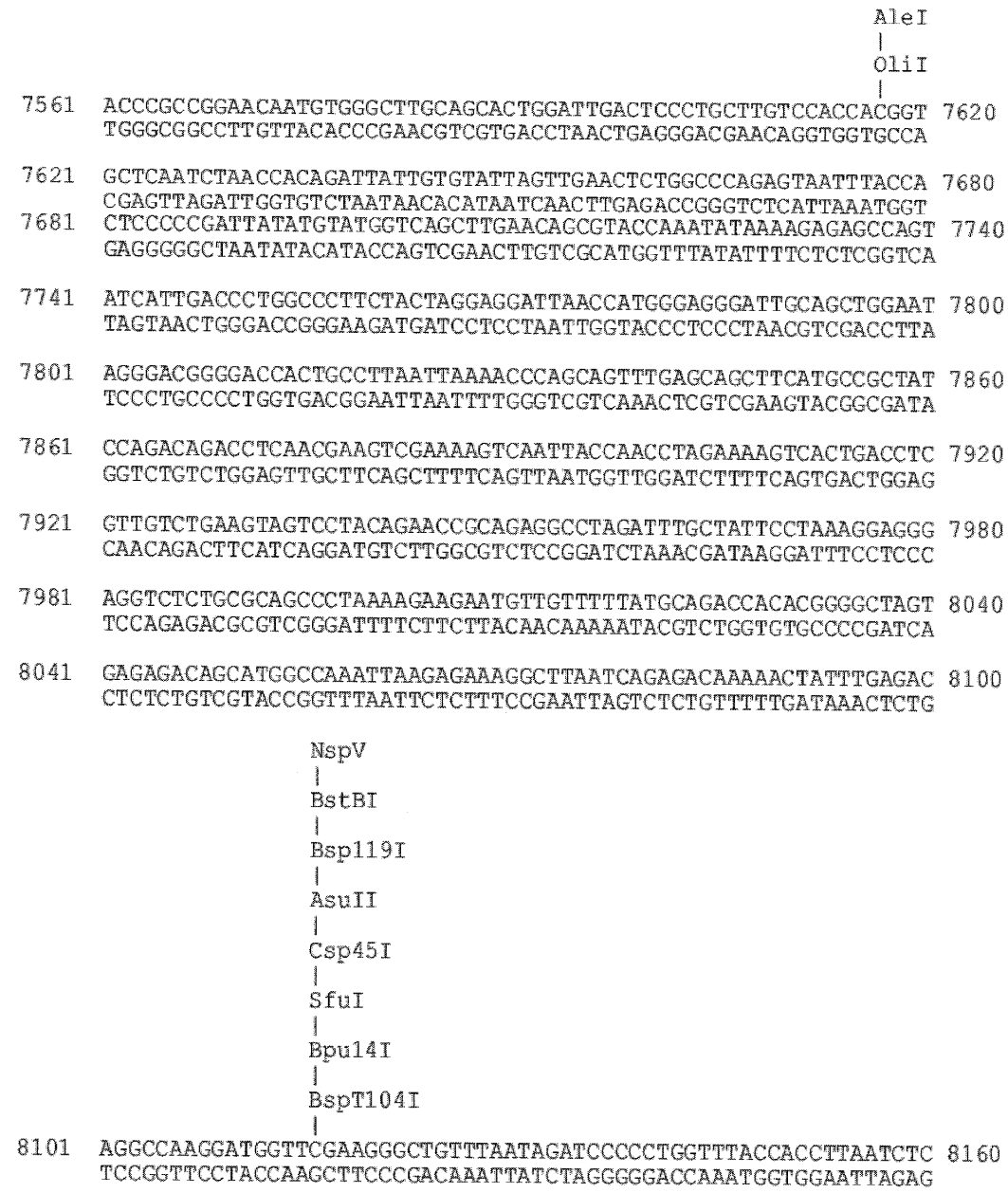
7321  GACCCAAGAATGTTGGCTGTGCTTAGTGTCGGGACCTCCTTATTACGAAGGAGTAGCGGT 7380
      CTGGGTTCTTACAACCGACACGAATCACAGCCCTGGAGGAATAATGCTTCCTCATCGCCA

7381  CGTGGGCACTTATACCAATCATTCCACCGCTCCGGCCAACTGTACGGCCACTTCCCAACA 7440
      GCACCCGTGAATATGGTTAGTAAGGTGGCGAGGCCGGTTGACATGCCGGTGAAGGGTTGT

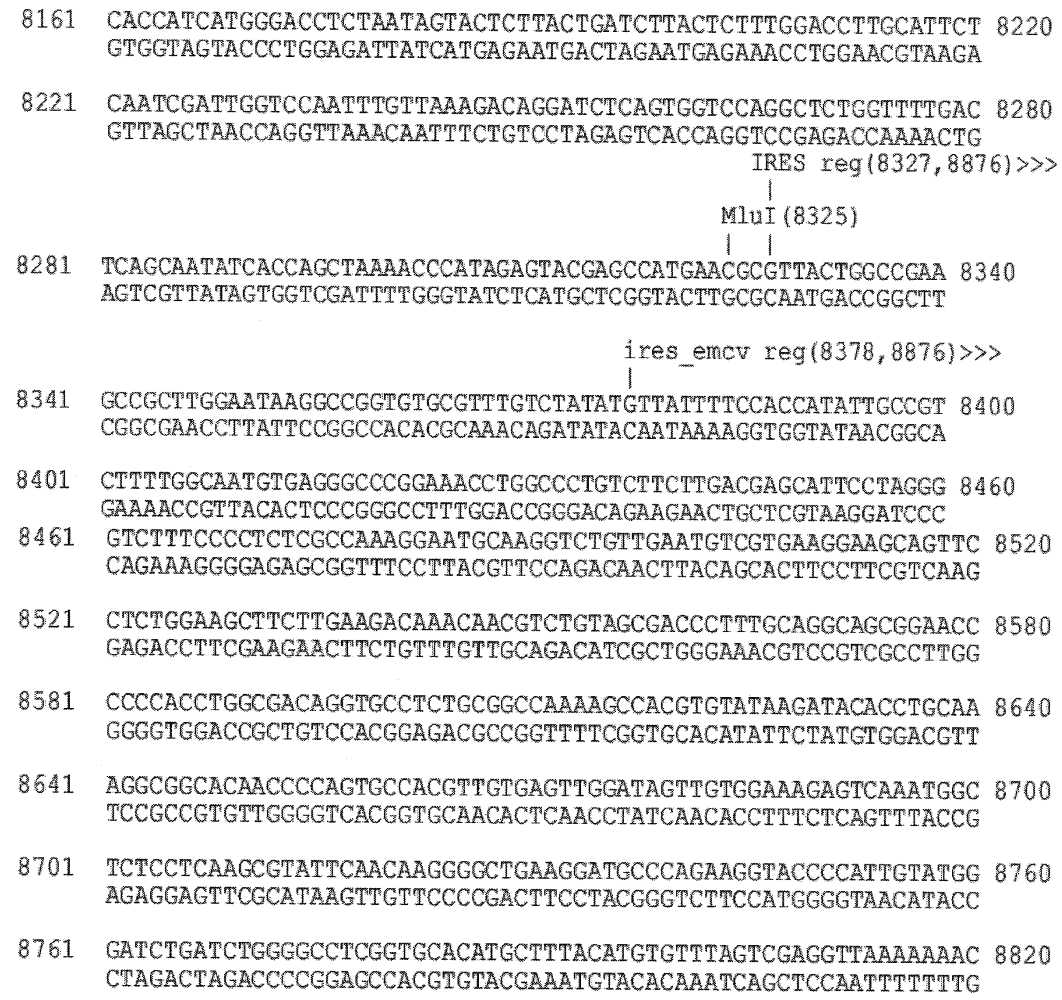
                                                EcoT22I
                                                |
                                                NsiI
                                                |
                                                Mph1103I
                                                |
                                                Zsp2I
                                                |
                                             BfrBI
                                             | |
7441  TAAGCTTACCCTATCTGAAGTGACAGGACAGGGCCTATGCATGGGGGCAGTACCTAAAAC 7500
      ATTCGAATGGGATAGACTTCACTGTCCTGTCCCGGATACGTACCCCCGTCATGGATTTTG

                                          NaeI
                                          |
                                          PdiI
                                          |
                                       MroNI
                                       | |
                                       NgoMIV
                                       | |
7501  TCACCAGGCCTTATGTAACACCACCCAAAGCGCCGGCTCAGGATCCTACTACCTTGCAGC 7560
      AGTGGTCCGGAATACATTGTGGTGGGTTTCGCGGCCGAGTCCTAGGATGATGGAACGTCG
```

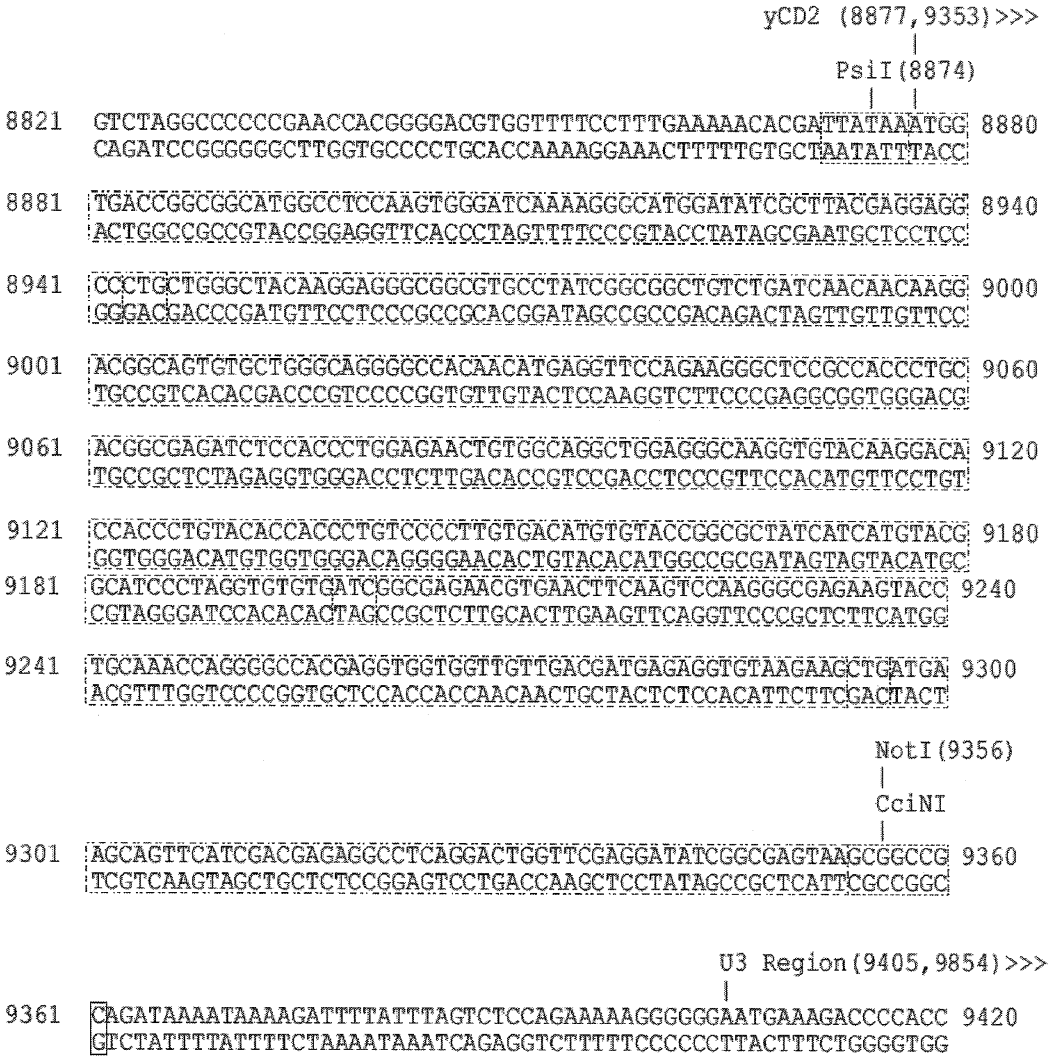
도면1do


AleI
OliI
NspV
BstBI
Bsp119I
AsuII
Csp45I
SfuI
Bpu14I
BspT104I
7561 ACCCGCCGGAACAATGTGGGCTTGCAGCACTGGATTGACTCCCTGCTTGTCCACCACGGT 7620
TGGGCGGCCTTGTTACACCCGAACGTCGTGACCTAACTGAGGGACGAACAGGTGGTGCCA
7621 GCTCAATCTAACCACAGATTATTGTGTATTAGTTGAACTCTGGCCCAGAGTAATTTACCA 7680
CGAGTTAGATTGGTGTCTAATAACACATAATCAACTTGAGACCGGGTCTCATTAAATGGT
7681 CTCCCCCGATTATATGTATGGTCAGCTTGAACAGCGTACCAAATATAAAAGAGAGCCAGT 7740
GAGGGGGCTAATATACATACCAGTCGAACTTGTCGCATGGTTTATATTTTCTCTCGGTCA
7741 ATCATTGACCCTGGCCCTTCTACTAGGAGGATTAACCATGGGAGGGATTGCAGCTGGAAT 7800
TAGTAACTGGGACCGGGAAGATGATCCTCCTAATTGGTACCCTCCCTAACGTCGACCTTA
7801 AGGGACGGGGACCACTGCCTTAATTAAAACCCAGCAGTTTGAGCAGCTTCATGCCGCTAT 7860
TCCCTGCCCCTGGTGACGGAATTAATTTTGGGTCGTCAAACTCGTCGAAGTACGGCGATA
7861 CCAGACAGACCTCAACGAAGTCGAAAAGTCAATTACCAACCTAGAAAAGTCACTGACCTC 7920
GGTCTGTCTGGAGTTGCTTCAGCTTTTCAGTTAATGGTTGGATCTTTTCAGTGACTGGAG
7921 GTTGTCTGAAGTAGTCCTACAGAACCGCAGAGGCCTAGATTTGCTATTCCTAAAGGAGGG 7980
CAACAGACTTCATCAGGATGTCTTGGCGTCTCCGGATCTAAACGATAAGGATTTCCTCCC
7981 AGGTCTCTGCGCAGCCCTAAAAGAAGAATGTTGTTTTTATGCAGACCACACGGGGCTAGT 8040
TCCAGAGACGCGTCGGGATTTTCTTCTTACAACAAAAATACGTCTGGTGTGCCCCGATCA
8041 GAGAGACAGCATGGCCAAATTAAGAGAAAGGCTTAATCAGAGACAAAAACTATTTGAGAC 8100
CTCTCTGTCGTACCGGTTTAATTCTCTTTCCGAATTAGTCTCTGTTTTTGATAAACTCTG
8101 AGGCCAAGGATGGTTCGAAGGGCTGTTTAATAGATCCCCCTGGTTTACCACCTTAATCTC 8160
TCCGGTTCCTACCAAGCTTCCCGACAAATTATCTAGGGGGACCAAATGGTGGAATTAGAG

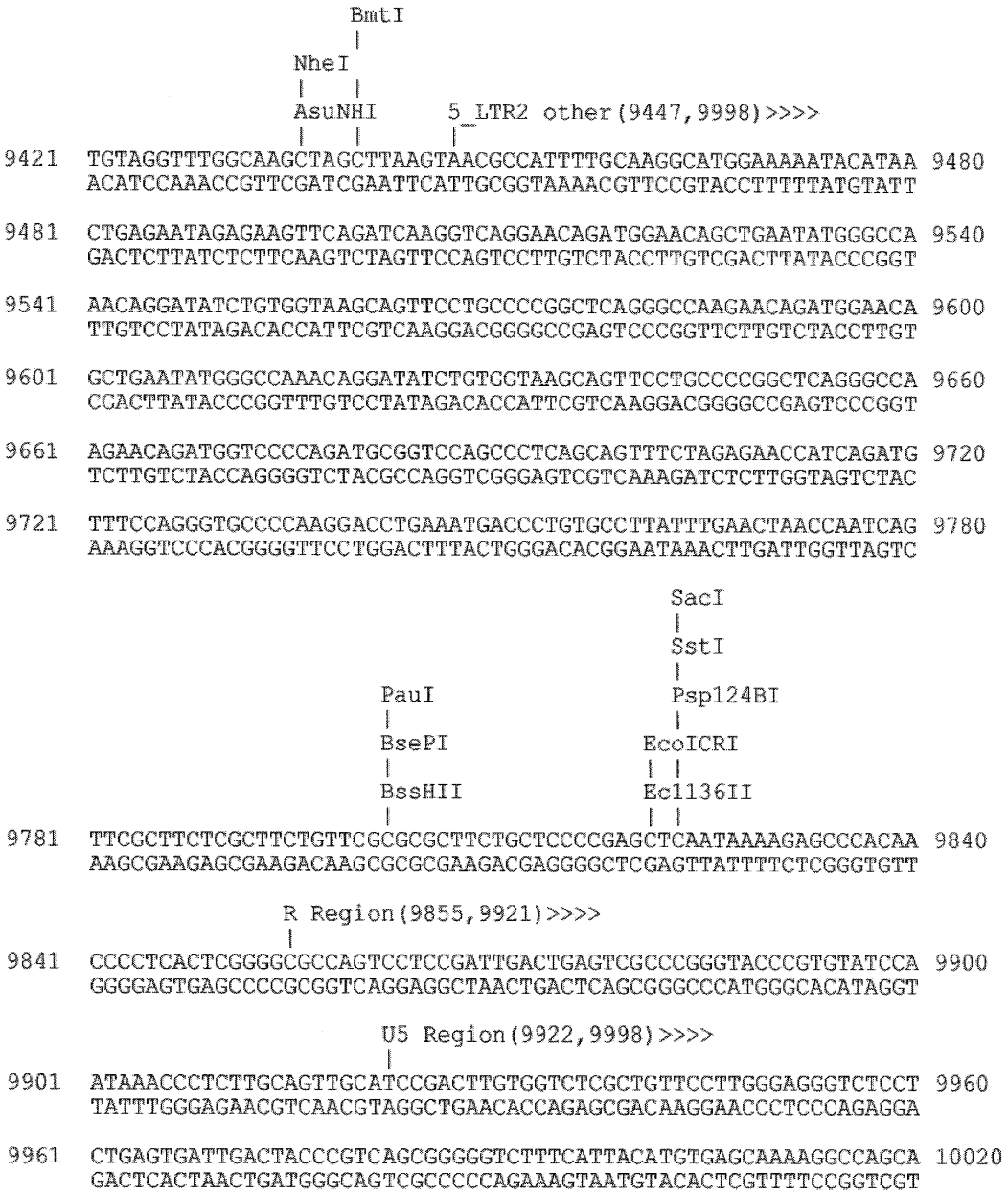
도면1dp


8161 CACCATCATGGGACCTCTAATAGTACTCTTACTGATCTTACTCTTTGGACCTTGCATTCT 8220
GTGGTAGTACCCTGGAGATTATCATGAGAATGACTAGAATGAGAAACCTGGAACGTAAGA
8221 CAATCGATTGGTCCAATTTGTTAAAGACAGGATCTCAGTGGTCCAGGCTCTGGTTTTGAC 8280
GTTAGCTAACCAGGTTAAACAATTTCTGTCCTAGAGTCACCAGGTCCGAGACCAAAACTG
IRES reg(8327,8876)>>>
MluI(8325)
8281 TCAGCAATATCACCAGCTAAAACCCATAGAGTACGAGCCATGAACGCGTTACTGGCCGAA 8340
AGTCGTTATAGTGGTCGATTTTGGGTATCTCATGCTCGGTACTTGCGCAATGACCGGCTT
ires_emcv reg(8378,8876)>>>
8341 GCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGT 8400
CGGCGAACCTTATTCCGGCCACACGCAAACAGATATACAATAAAAGGTGGTATAACGGCA
8401 CTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGG 8460
GAAAACCGTTACACTCCCGGGCCTTTGGACCGGGACAGAAGAACTGCTCGTAAGGATCCC
8461 GTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTC 8520
CAGAAAGGGGAGAGCGGTTTCCTTACGTTCCAGACAACTTACAGCACTTCCTTCGTCAAG
8521 CTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACC 8580
GAGACCTTCGAAGAACTTCTGTTTGTTGCAGACATCGCTGGGAAACGTCCGTCGCCTTGG
8581 CCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAA 8640
GGGGTGGACCGCTGTCCACGGAGACGCCGGTTTTCGGTGCACATATTCTATGTGGACGTT
8641 AGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGC 8700
TCCGCCGTGTTGGGGTCACGGTGCAACACTCAACCTATCAACACCTTTCTCAGTTTACCG
8701 TCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGG 8760
AGAGGAGTTCGCATAAGTTGTTCCCCGACTTCCTACGGGTCTTCCATGGGGTAACATACC
8761 GATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAAC 8820
CTAGACTAGACCCCGGAGCCACGTGTACGAAATGTACACAAATCAGCTCCAATTTTTTTG

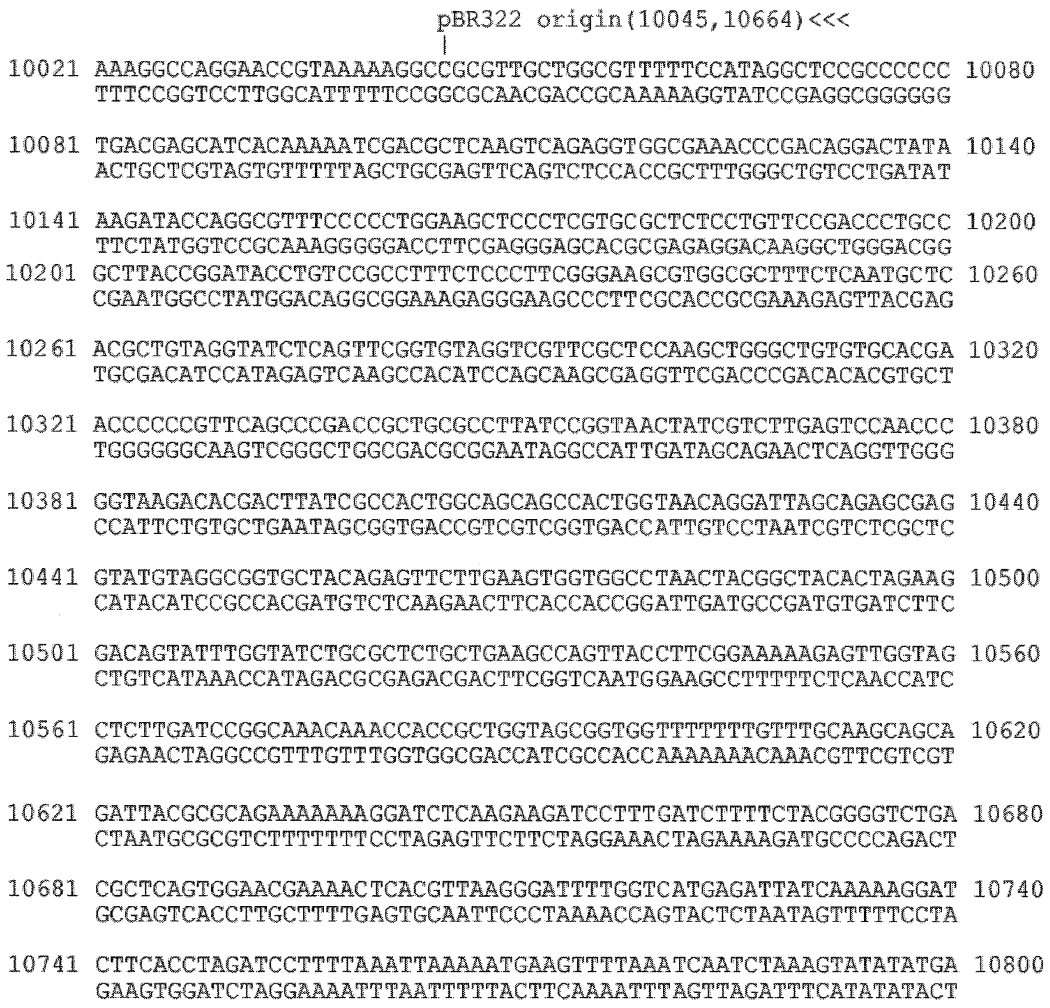
도면1dq


yCD2 (8877,9353)>>>
PsiI(8874)
8821 GTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATTATAAATGG 8880
CAGATCCGGGGGGCTTGGTGCCCCTGCACCAAAAGGAAACTTTTTGTGCTAATATTTACC
8881 TGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTACGAGGAGG 8940
ACTGGCCGCCGTACCGGAGGTTCACCCTAGTTTTCCCGTACCTATAGCGAATGCTCCTCC
8941 CCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAGG 9000
GGGACGACCCGATGTTCCTCCCGCCGCACGGATAGCCGCCGACAGACTAGTTGTTGTTCC
9001 ACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGC 9060
TGCCGTCACACGACCCGTCCCCGGTGTTGTACTCCAAGGTCTTCCCGAGGCGGTGGGACG
9061 ACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACA 9120
TGCCGCTCTAGAGGTGGGACCTCTTGACACCGTCCGACCTCCCGTTCCACATGTTCCTGT
9121 CCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACG 9180
GGTGGGACATGTGGTGGGACAGGGGAACACTGTACACATGGCCGCGATAGTAGTACATGC
9181 GCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACC 9240
CGTAGGGATCCACACACTAGCCGCTCTTGCACTTGAAGTTCAGGTTCCCGCTCTTCATGG
9241 TGCAAACCACGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGATGA 9300
ACGTTTGGTCCCCGGTGCTCCACCACCAACAACTGCTACTCTCCACATTCTTCGACTACT
NotI(9356)
CciNI
9301 AGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTAAGCGGCCG 9360
TCGTCAAGTAGCTGCTCTCCGGAGTCCTGACCAAGCTCCTATAGCCGCTCATTCGCCGGC
U3 Region(9405,9854)>>>
9361 CAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAATGAAAGACCCCACC 9420
GTCTATTTTATTTTCTAAAATAAATCAGAGGTCTTTTTCCCCCCTTACTTTCTGGGGTGG

도면1dr


BmtI
NheI
AsuNHI
5_LTR2 other(9447,9998)>>>>
9421 TGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAAATACATAA 9480
ACATCCAAACCGTTCGATCGAATTCATTGCGGTAAAACGTTCCGTACCTTTTTATGTATT
9481 CTGAGAATAGAGAAGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCA 9540
GACTCTTATCTCTTCAAGTCTAGTTCCAGTCCTTGTCTACCTTGTCGACTTATACCCGGT
9541 AACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACA 9600
TTGTCCTATAGACACCATTCGTCAAGGACGGGGCCGAGTCCCGGTTCTTGTCTACCTTGT
9601 GCTGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCA 9660
CGACTTATACCCGGTTTGTCCTATAGACACCATTCGTCAAGGACGGGGCCGAGTCCCGGT
9661 AGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGAGAACCATCAGATG 9720
TCTTGTCTACCAGGGGTCTACGCCAGGTCGGGAGTCGTCAAAGATCTCTTGGTAGTCTAC
9721 TTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTAACCAATCAG 9780
AAAGGTCCCACGGGGTTCCTGGACTTTACTGGGACACGGAATAAACTTGATTGGTTAGTC
SacI
SstI
Psp124BI
PauI
BsePI
BssHII
EcoICRI
Ec1136II
9781 TTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCACAA 9840
AAGCGAAGAGCGAAGACAAGCGCGCGAAGACGAGGGGCTCGAGTTATTTTCTCGGGTGTT
R Region(9855,9921)>>>>
9841 CCCCTCACTCGGGGCGCCAGTCCTCCGATTGACTGAGTCGCCCGGGTACCCGTGTATCCA 9900
GGGGAGTGAGCCCCGCGGTCAGGAGGCTAACTGACTCAGCGGGCCCATGGGCACATAGGT
U5 Region(9922,9998)>>>>
9901 ATAAACCCTCTTGCAGTTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCT 9960
TATTTGGGAGAACGTCAACGTAGGCTGAACACCAGAGCGACAAGGAACCCTCCCAGAGGA
9961 CTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCATTACATGTGAGCAAAAGGCCAGCA 10020
GACTCACTAACTGATGGGCAGTCGCCCCCAGAAAGTAATGTACACTCGTTTTCCGGTCGT

도면1ds


pBR322 origin(10045,10664)<<<
10021 AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCC 10080
TTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGG
10081 TGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA 10140
ACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATAT
10141 AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC 10200
TTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGG
10201 GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTC 10260
CGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAG
10261 ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA 10320
TGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCT
10321 ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC 10380
TGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGG
10381 GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG 10440
CCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTC
10441 GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG 10500
CATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTC
10501 GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG 10560
CTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATC
10561 CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCA 10620
GAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGT
10621 GATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA 10680
CTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACT
10681 CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT 10740
GCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTA
10741 CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGA 10800
GAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACT

도면1dt

```
                          amp marker(10819,11679)<<<
                                   |
10801 GTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG 10860
      CATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGAC

10861 TCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGA 10920
      AGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCT

10921 GGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC 10980
      CCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

10981 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAAC 11040
      TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTG

11041 TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCC 11100
      AAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGG

11101 AGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTCACGCTCGTC 11160
      TCAATTATCAAACGCGTTGCAACAACGGTAACGACGTCCGTAGCACCACAGTGCGAGCAG

11161 GTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC 11220
      CAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGG

11221 CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTT 11280
      GTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAA

11281 GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC 11340
      CCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGG

11341 ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG 11400
      TAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCAC

11401 TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAG 11460
      ATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATC

11461 CAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGAT 11520
      GTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTA

11521 CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC 11580
      GAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCG

11581 ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA 11640
      TAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTT

11641 AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTA 11700
      TTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAAT
```

도면1du

```
                          amp prom(11721,11749)<<<
                                   |
11701 TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA 11760
      AACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTT

11761 AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGA 11820
      TTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCT

11821 AACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCT 11880
      TTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGA

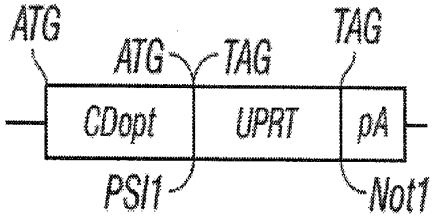
11881 TCAAGAATTCAT 11892
      AGTTCTTAAGTA
```

도면2a

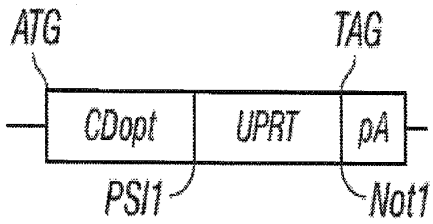
1. PSI1/Not1 부위로 CDopt와 UPRT 유전자의 합성을 분리한다.



2. PSI1로 절단하고 평활 말단 결찰을 수행한다.



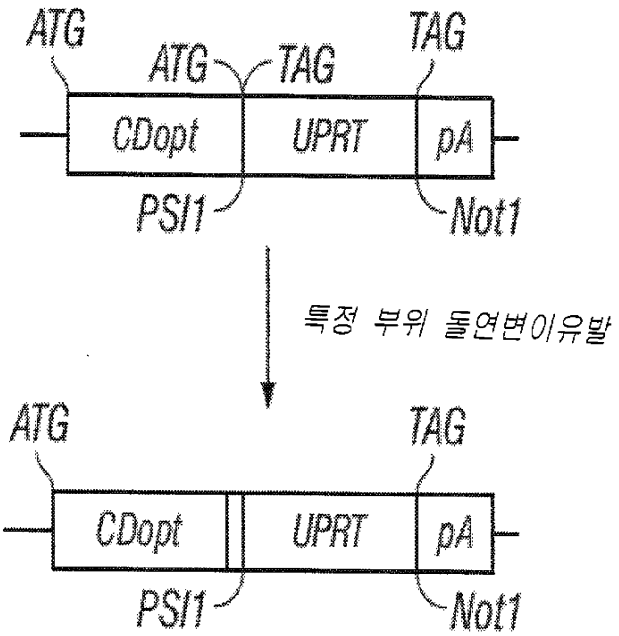
3. 특정 부위 돌연변이유발로 CDopt 유전자 종결 코돈과 UPRT 개시 코돈의 결실



4. CDopt-UPRT 유전자 크기: 1.3 kb

도면2b

1. CDopt-링커-UPRT의 합성: 링커의 특정 부위 돌연변이유발 삽입



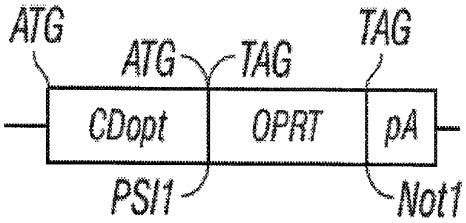
2. CDopt-링커-UPRT 유전자 크기: 1.36 kb

도면2c

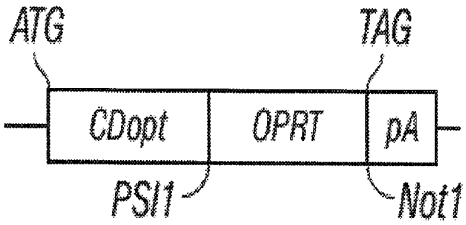
1. PSI1/Not1 부위로 CDopt 유전자와 OPRT 도메인의 합성을 분리한다.



2. PSI1로 절단하고 평활 말단 결찰을 수행한다.



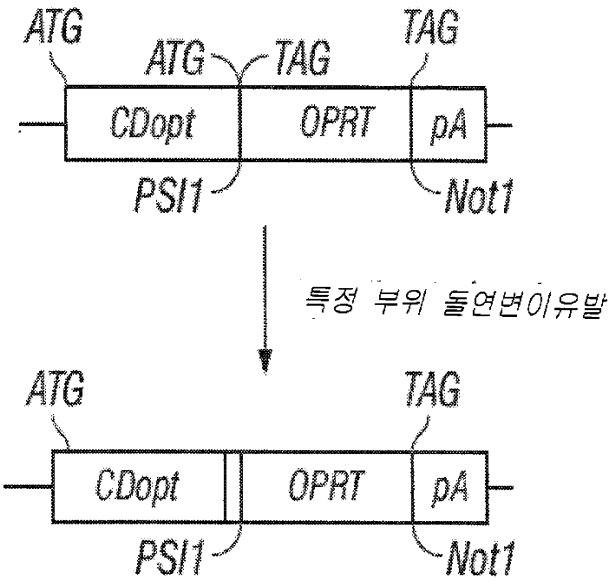
3. 특정 부위 돌연변이유발로 CDopt 유전자 종결 코돈과 OPRT 개시 코돈의 결실



4. CDopt-OPRT 유전자 크기: 1.3 kb

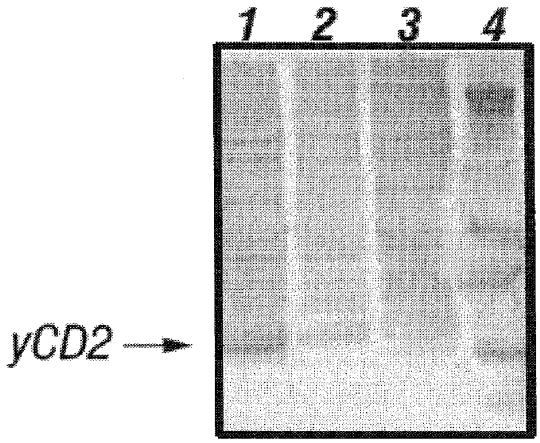
도면2d

1. CDopt-링커-OPRT의 합성: 링커의 특정 부위 돌연변이유발 삽입



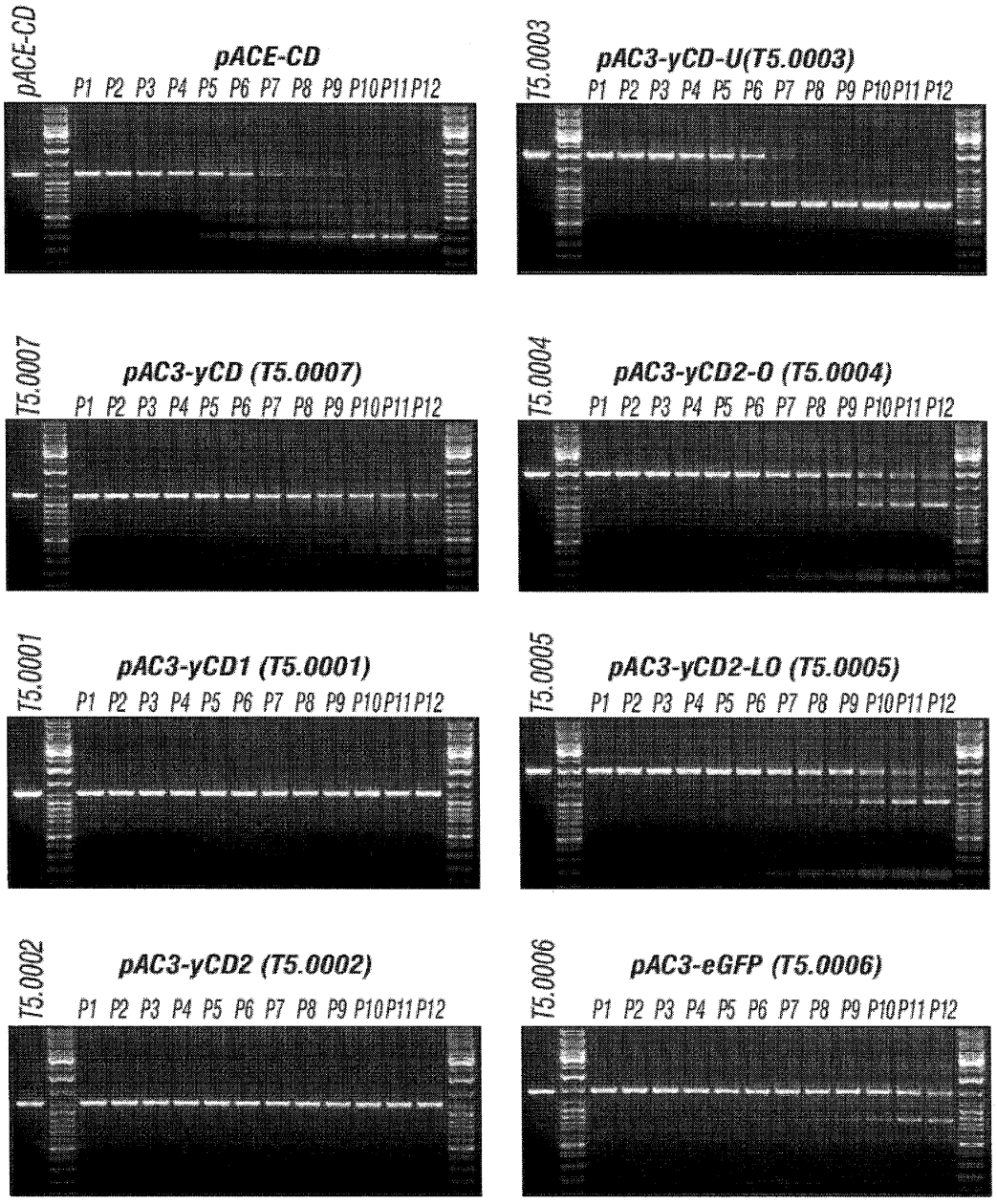
2. CD-OPRT 유전자 크기: 1.33 kb

도면3

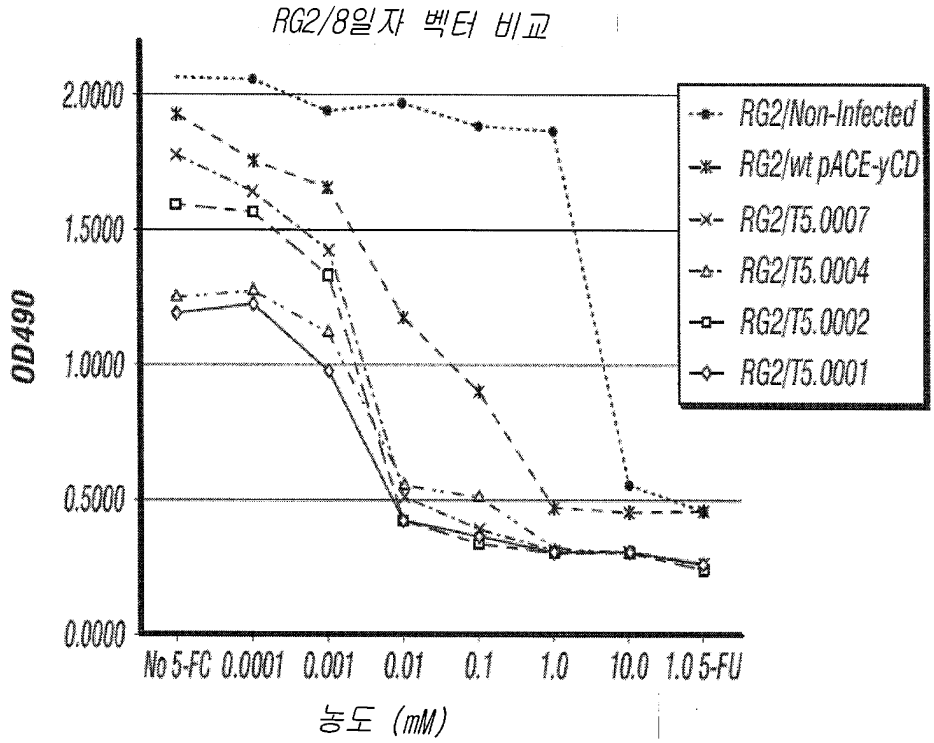


1 - U87 + AC3-yCD2 (V)
2 - U87 + ACE-yCD (V)
3 - U87 (감염되지 않음)
4 - 분자량 기준

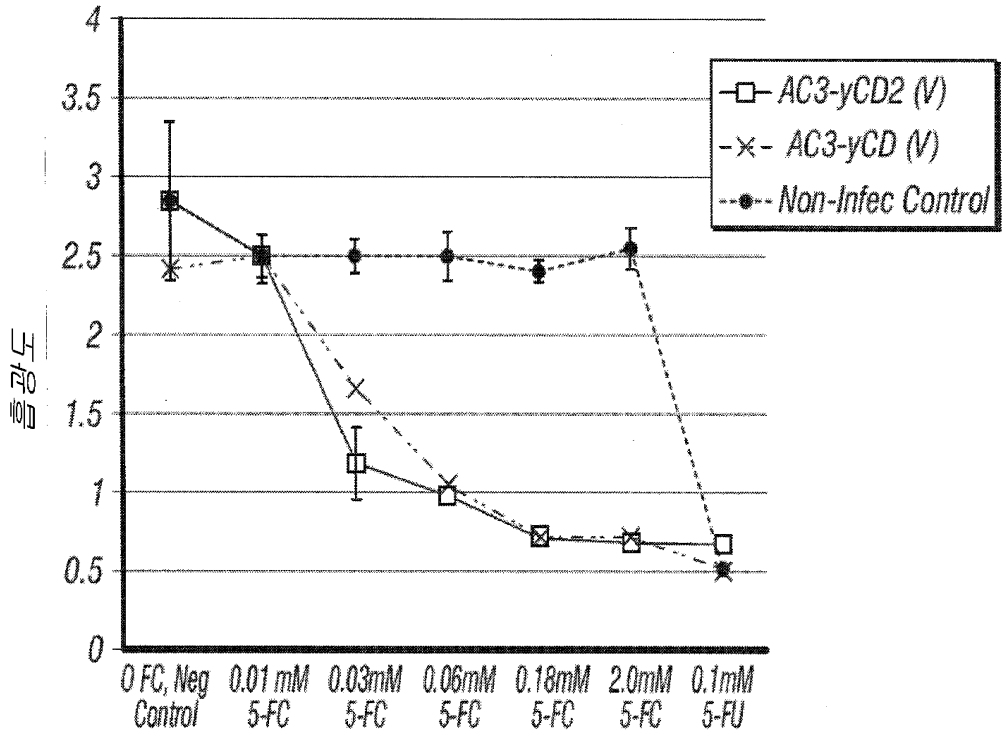
도면4


pACE-CD
pACE-CD
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0003
pAC3-yCD-U(T5.0003)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0007
pAC3-yCD (T5.0007)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0004
pAC3-yCD2-O (T5.0004)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0001
pAC3-yCD1 (T5.0001)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0005
pAC3-yCD2-LO (T5.0005)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0002
pAC3-yCD2 (T5.0002)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12
T5.0006
pAC3-eGFP (T5.0006)
P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12

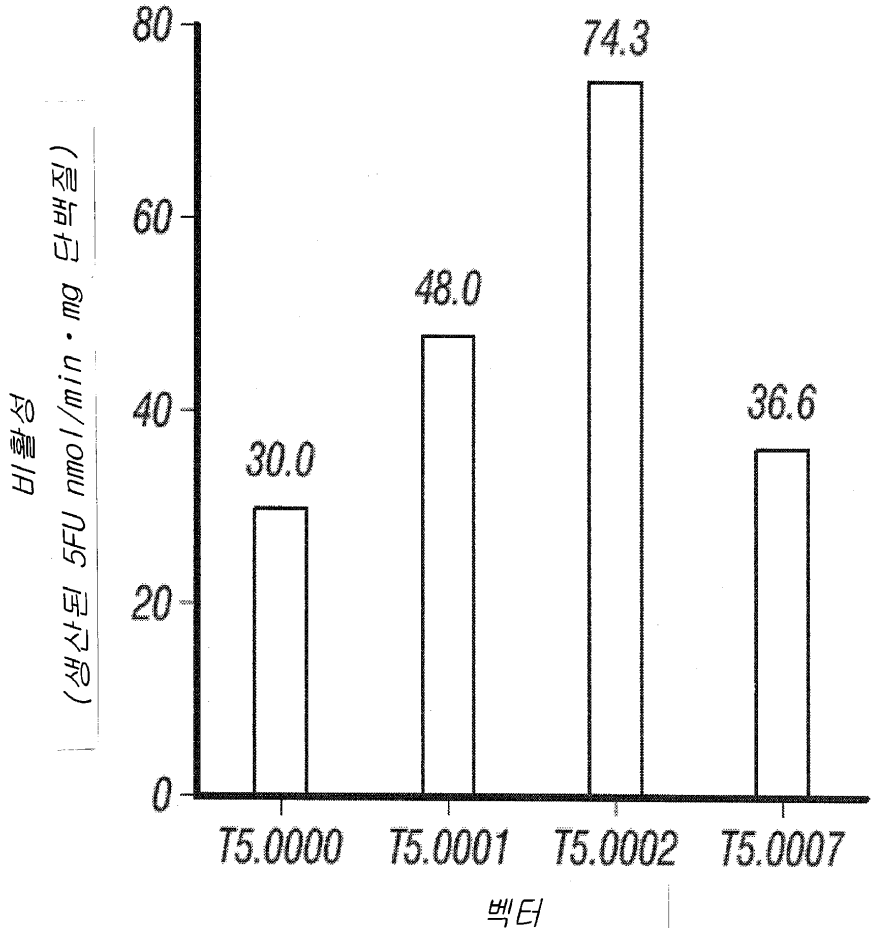
도면5a


RG2/8일자 벡터 비교
OD490
2.0000
1.5000
1.0000
0.5000
0.0000
No 5-FC
0.0001
0.001
0.01
0.1
1.0
10.0
1.0 5-FU
농도 (mM)
RG2/Non-Infected
RG2/wt pACE-yCD
RG2/T5.0007
RG2/T5.0004
RG2/T5.0002
RG2/T5.0001

도면5b


흡광도
4
3.5
3
2.5
2
1.5
1
0.5
0
0 FC, Neg Control
0.01 mM 5-FC
0.03mM 5-FC
0.06mM 5-FC
0.18mM 5-FC
2.0mM 5-FC
0.1mM 5-FU
AC3-yCD2 (V)
AC3-yCD (V)
Non-Infec Control

도면6



서 열 목 록

SEQUENCE LISTING

<110> Tocagen Inc.

Gruber, Harry E.

Logg, Christopher R.

Jolly, Douglas

Perez, Omar

<120> GENE THERAPY VECTORS AND CYTOSINE DEAMINASES

<130> 00014-007AU1

<140> Not yet assinged

<141> 2011-03-24

<150> US 61/100,666

<151> 2008-09-26

<150> US 61/318,728

<151> 2010-03-29

<150> PCT/US09/58510

<151> 2009-09-26

<150> US 61/120,618

<151> 2008-12-08

<150> US 61/186,823

<151> 2009-06-13

<160

> 48

<170> PatentIn version 3.5

<210> 1

<211> 477

<212> DNA

<213> Saccharomyces cerevisiae

<220><221> CDS

<222> (1)..(477)

<400> 1

atg gtg aca ggg gga atg gca agc aag tgg gat cag aag ggt atg gac 48

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

att gcc tat gag gag gcg gcc tta ggt tac aaa gag ggt ggt gtt cct 96

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

20 25 30

att ggc gga tgt ctt atc aat aac aaa gac gga agt gtt ctc ggt cgt 144

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

35 40 45

ggt cac aac atg aga ttt caa aag gga tcc gcc aca cta cat ggt gag 192

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

atc tcc act ttg gaa aac tgt ggg aga tta gag ggc aaa gtg tac aaa 240

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

gat acc act ttg tat acg acg ctg tct cca tgc gac atg tgt aca ggt 288

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

gcc atc atc atg tat ggt att cca cgc tgt gtt gtc ggt gag aac gtt 336

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

aat ttc aaa agt aag ggc gag aaa tat tta caa act aga ggt cac gag 384

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

gtt gtt gtt gtt gac gat gag agg tgt aaa aag atc atg aaa caa ttt 432

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe

130 135 140

atc gat gaa aga cct cag gat tgg ttt gaa gat att ggt gag tag 477

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 2

<211> 158

<212> PRT

<213> Saccharomyces cerevisiae

<400> 2

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

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

20 25 30

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

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 3

<211> 477

<212> DNA

<213> Artificial Sequence

<220><223> Engineered cytosine deaminase

<220><221> CDS

<222> (1)..(477)

<400> 3

atg gtg aca ggg gga atg gca agc aag tgg gat cag aag ggt atg gac 48

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

att gcc tat gag gag gcg tta tta ggt tac aaa gag ggt ggt gtt cct 96

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

20 25 30

att ggc gga tgt ctt atc aat aac aaa gac gga agt gtt ctc ggt cgt 144

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

35 40 45

ggt cac aac atg aga ttt caa aag gga tcc gcc aca cta cat ggt gag 192

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

atc tcc act ttg gaa aac tgt ggg aga tta gag ggc aaa gtg tac aaa 240

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

gat acc act ttg tat acg acg ctg tct cca tgc gac atg tgt aca ggt 288

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

gcc atc atc atg tat ggt att cca cgc tgt gtc atc ggt gag aac gtt 336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
100 105 110
aat ttc aaa agt aag ggc gag aaa tat tta caa act aga ggt cac gag 384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
115 120 125
gtt gtt gtt gtt gac gat gag agg tgt aaa aag tta atg aaa caa ttt 432

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
130 135 140
atc gat gaa aga cct cag gat tgg ttt gaa gat att ggt gag tag 477
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
145 150 155

<210> 4
<211> 158
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Construct
<400> 4

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

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

65 70 75 80
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 5

<211> 480

<212> DNA

<213> Artificial Sequence

<220><223> Human codon optimized cytosine deaminase

<220><221> CDS

<222> (1)..(480)

<400> 5

atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

atc gct tac gag gag gcc gca ctg ggc tac aag gag ggc ggc gtg cct 96

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

20 25 30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144

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

35 40 45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag 240

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc 288

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

gct atc atc atg tac ggc atc cct agg tgt gtg gtc ggc gag aac gtg 336

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag 384

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

gtg gtg gtt gtt gac gat gag agg tgt aag aag atc atg aag cag ttc 432

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe

130 135 140

atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tga taa 480

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 6

<211> 158

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 6

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

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

20 25 30

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

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 7

<211> 756

<212> DNA

<213> Saccharomyces cerevisiae

<220><221> CDS

<222> (1)..(756)

<400> 7

atg aac ccg tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca tat 48

Met Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr

1 5 10 15

ctt ata tat tat cca aac aaa ggg tct ttc gtt agc aaa cct aga aat 96

Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn

20 25 30

ctg caa aaa atg tct tcg gaa cca ttt aag aac gtc tac ttg cta cct 144

Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro

35 40 45

caa aca aac caa ttg ctg ggt ttg tac acc atc atc aga aat aag aat 192

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

50 55 60

aca act aga cct gat ttc att ttc tac tcc gat aga atc atc aga ttg 240

Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu

65 70 75 80

```
ttg gtt gaa gaa ggt ttg aac cat cta cct gtg caa aag caa att gtg        288

Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val
                85                  90                  95
gaa act gac acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt aaa        336
Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys
            100                 105                 110
atc tgt ggt gtt tcc att gtc aga gct ggt gaa tcg atg gag caa gga        384
Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly

        115                 120                 125
tta aga gac tgt tgt agg tct gtg cgt atc ggt aaa att tta att caa        432
Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln
    130                 135                 140
agg gac gag gag act gct tta cca aag tta ttc tac gaa aaa tta cca        480
Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro
145                 150                 155                 160

gag gat ata tct gaa agg tat gtc ttc cta tta gac cca atg ctg gcc        528
Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala
                165                 170                 175
acc ggt ggt agt gct atc atg gct aca gaa gtc ttg att aag aga ggt        576
Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly
            180                 185                 190
gtt aag cca gag aga att tac ttc tta aac cta atc tgt agt aag gaa        624

Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu
        195                 200                 205
ggg att gaa aaa tac cat gcc gcc ttc cca gag gtc aga att gtt act        672
Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr
    210                 215                 220
ggt gcc ctc gac aga ggt cta gat gaa aac aag tat cta gtt cca ggg        720
Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly
```

225 230 235 240

ttg ggt gac ttt ggt gac aga tac tac tgt gtt taa 756

Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val

245 250

<210> 8

<211> 251

<212> PRT

<213> Saccharomyces cerevisiae

<400> 8

Met Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr

1 5 10 15

Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn

20 25 30

Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro

35 40 45

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

50 55 60

Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu

65 70 75 80

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

85 90 95

Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys

100 105 110

Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly

115 120 125

Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln

130 135 140

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

145 150 155 160

Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala

165 170 175

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

180 185 190

Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu

195 200 205

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

210 215 220

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

225 230 235 240

Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val

245 250

<210> 9

<211> 1443

<212> DNA

<213> homo sapiens

<220><221> CDS

<222> (1)..(1443)

<400> 9

atg gct gtt gct cgt gct gct ctt ggt cct ctt gtt act ggt ctt tat 48

Met Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr

1 5 10 15

gat gtt caa gct ttt aaa ttt ggt gat ttt gtt ctt aaa tct ggt ctt 96

Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu

20 25 30

tct tct cct att tat att gat ctt cgt ggt att gtt tct cgt cct cgt 144

Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg

35 40 45

ctt ctt tct caa gtt gct gat att ctt ttt caa act gct caa aat gct 192

Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala

50 55 60

ggt att tct ttt gat act gtt tgt ggt gtt cct tat act gct ctt cct 240

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

65 70 75 80

```
ctt gct act gtt att tgt tct act aat caa att cct atg ctt att cgt        288
Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg
                85                  90                  95
cgt aaa gaa act aaa gat tat ggt act aaa cgt ctt gtt gaa ggt act        336
Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr
            100                 105                 110
att aat cct ggt gaa act tgt ctt att att gaa gat gtt gtt act tct        384

Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser
        115                 120                 125
ggt tct tct gtt ctt gaa act gtt gaa gtt ctt caa aaa gaa ggt ctt        432
Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu
    130                 135                 140
aaa gtt act gat gct att gtt ctt ctt gat cgt gaa caa ggt ggt aaa        480
Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly Lys

145                 150                 155                 160
gat aaa ctt caa gct cat ggt att cgt ctt cat tct gtt tgt act ctt        528
Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu
                165                 170                 175
tct aaa atg ctt gaa att ctt gaa caa caa aaa aaa gtt gat gct gaa        576
Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu
            180                 185                 190

act gtt ggt cgt gtt aaa cgt ttt att caa gaa aat gtt ttt gtt gct        624
Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala
        195                 200                 205
gct aat cat aat ggt tct cct ctt tct att aaa gaa gct cct aaa gaa        672
Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu
    210                 215                 220
ctt tct ttt ggt gct cgt gct gaa ctt cct cgt att cat cct gtt gct        720

Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala
225                 230                 235                 240
tct aaa ctt ctt cgt ctt atg caa aaa aaa gaa act aat ctt tgt ctt        768
```

```
Ser Lys Leu Leu Arg Leu Met Gln Lys Lys Glu Thr Asn Leu Cys Leu
                245                 250                 255
tct gct gat gtt tct ctt gct cgt gaa ctt ctt caa ctt gct gat gct        816
Ser Ala Asp Val Ser Leu Ala Arg Glu Leu Leu Gln Leu Ala Asp Ala

            260                 265                 270
ctt ggt cct tct att tgt atg ctt aaa act cat gtt gat att ctt aat        864
Leu Gly Pro Ser Ile Cys Met Leu Lys Thr His Val Asp Ile Leu Asn
        275                 280                 285
gat ttt act ctt gat gtt atg aaa gaa ctt att act ctt gct aaa tgt        912
Asp Phe Thr Leu Asp Val Met Lys Glu Leu Ile Thr Leu Ala Lys Cys
    290                 295                 300

cat gaa ttt ctt att ttt gaa gat cgt aaa ttt gct gat att ggt aat        960
His Glu Phe Leu Ile Phe Glu Asp Arg Lys Phe Ala Asp Ile Gly Asn
305                 310                 315                 320
act gtt aaa aaa caa tat gaa ggt ggt att ttt aaa att gct tct tgg       1008
Thr Val Lys Lys Gln Tyr Glu Gly Gly Ile Phe Lys Ile Ala Ser Trp
                325                 330                 335
gct gat ctt gtt aat gct cat gtt gtt cct ggt tct ggt gtt gtt aaa       1056

Ala Asp Leu Val Asn Ala His Val Val Pro Gly Ser Gly Val Val Lys
            340                 345                 350
ggt ctt caa gaa gtt ggt ctt cct ctt cat cgt ggt tgt ctt ctt att       1104
Gly Leu Gln Glu Val Gly Leu Pro Leu His Arg Gly Cys Leu Leu Ile
        355                 360                 365
gct gaa atg tct tct act ggt tct ctt gct act ggt gat tat act cgt       1152
Ala Glu Met Ser Ser Thr Gly Ser Leu Ala Thr Gly Asp Tyr Thr Arg

    370                 375                 380
gct gct gtt cgt atg gct gaa gaa cat tct gaa ttt gtt gtt ggt ttt       1200
Ala Ala Val Arg Met Ala Glu Glu His Ser Glu Phe Val Val Gly Phe
385                 390                 395                 400
att tct ggt tct cgt gtt tct atg aaa cct gaa ttt ctt cat ctt act       1248
Ile Ser Gly Ser Arg Val Ser Met Lys Pro Glu Phe Leu His Leu Thr
```

405 410 415

cct ggt gtt caa ctt gaa gct ggt ggt gat aat ctt ggt caa caa tat 1296
Pro Gly Val Gln Leu Glu Ala Gly Gly Asp Asn Leu Gly Gln Gln Tyr
420 425 430
aat tct cct caa gaa gtt att ggt aaa cgt ggt tct gat att att att 1344
Asn Ser Pro Gln Glu Val Ile Gly Lys Arg Gly Ser Asp Ile Ile Ile
435 440 445
gtt ggt cgt ggt att att tct gct gct gat cgt ctt gaa gct gct gaa 1392

Val Gly Arg Gly Ile Ile Ser Ala Ala Asp Arg Leu Glu Ala Ala Glu
450 455 460
atg tat cgt aaa gct gct tgg gaa gct tat ctt tct cgt ctt ggt gtt 1440
Met Tyr Arg Lys Ala Ala Trp Glu Ala Tyr Leu Ser Arg Leu Gly Val
465 470 475 480
taa 1443
<210> 10
<211> 480
<212> PRT
<213> homo sapiens

<400> 10
Met Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr
1 5 10 15
Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu
20 25 30
Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg
35 40 45
Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala
50 55 60

Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro
65 70 75 80
Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg
85 90 95

Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr
100 105 110
Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr Ser
115 120 125

Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu
130 135 140
Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly Lys
145 150 155 160
Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu
165 170 175
Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala Glu
180 185 190

Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala
195 200 205
Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu
210 215 220
Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val Ala
225 230 235 240
Ser Lys Leu Leu Arg Leu Met Gln Lys Lys Glu Thr Asn Leu Cys Leu
245 250 255

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

Thr Val Lys Lys Gln Tyr Glu Gly Gly Ile Phe Lys Ile Ala Ser Trp
325 330 335
Ala Asp Leu Val Asn Ala His Val Val Pro Gly Ser Gly Val Val Lys

340 345 350

Gly Leu Gln Glu Val Gly Leu Pro Leu His Arg Gly Cys Leu Leu Ile

355 360 365

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

370 375 380

Ala Ala Val Arg Met Ala Glu Glu His Ser Glu Phe Val Val Gly Phe

385 390 395 400

Ile Ser Gly Ser Arg Val Ser Met Lys Pro Glu Phe Leu His Leu Thr

405 410 415

Pro Gly Val Gln Leu Glu Ala Gly Gly Asp Asn Leu Gly Gln Gln Tyr

420 425 430

Asn Ser Pro Gln Glu Val Ile Gly Lys Arg Gly Ser Asp Ile Ile Ile

435 440 445

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

450 455 460

Met Tyr Arg Lys Ala Ala Trp Glu Ala Tyr Leu Ser Arg Leu Gly Val

465 470 475 480

<210> 11

<211> 1227

<212> DNA

<213> Artificial Sequence

<220><223> Fusion construct CDopt-UPRT

<220><221> CDS

<222> (1)..(1227)

<400> 11

atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96

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

20 25 30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144

```
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
        35                  40                  45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag        192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
    50                  55                  60
atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag        240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65                  70                  75                  80
gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc        288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
                85                  90                  95
gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg        336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
            100                 105                 110
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag        384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

        115                 120                 125
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc        432
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
    130                 135                 140
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag aac ccg        480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Asn Pro
145                 150                 155                 160

tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca tat ctt ata tat        528
Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr Tyr Leu Ile Tyr
                165                 170                 175
tat cca aac aaa ggg tct ttc gtt agc aaa cct aga aat ctg caa aaa        576
Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn Leu Gln Lys
            180                 185                 190
```

atg tct tcg gaa cca ttt aag aac gtc tac ttg cta cct caa aca aac 624

Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro Gln Thr Asn
195 200 205

caa ttg ctg ggt ttg tac acc atc atc aga aat aag aat aca act aga 672
Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys Asn Thr Thr Arg
210 215 220

cct gat ttc att ttc tac tcc gat aga atc atc aga ttg ttg gtt gaa 720
Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu Leu Val Glu

225 230 235 240

gaa ggt ttg aac cat cta cct gtg caa aag caa att gtg gaa act gac 768
Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile Val Glu Thr Asp
245 250 255

acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt aaa atc tgt ggt 816
Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys Ile Cys Gly
260 265 270

gtt tcc att gtc aga gct ggt gaa tcg atg gag caa gga tta aga gac 864
Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln Gly Leu Arg Asp
275 280 285

tgt tgt agg tct gtg cgt atc ggt aaa att tta att caa agg gac gag 912
Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile Gln Arg Asp Glu
290 295 300

gag act gct tta cca aag tta ttc tac gaa aaa tta cca gag gat ata 960

Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro Glu Asp Ile
305 310 315 320

tct gaa agg tat gtc ttc cta tta gac cca atg ctg gcc acc ggt ggt 1008
Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala Thr Gly Gly
325 330 335

agt gct atc atg gct aca gaa gtc ttg att aag aga ggt gtt aag cca 1056
Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg Gly Val Lys Pro

```
            340                 345                 350

gag aga att tac ttc tta aac cta atc tgt agt aag gaa ggg att gaa       1104
Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu Gly Ile Glu
        355                 360                 365

aaa tac cat gcc gcc ttc cca gag gtc aga att gtt act ggt gcc ctc       1152
Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr Gly Ala Leu
    370                 375                 380

gac aga ggt cta gat gaa aac aag tat cta gtt cca ggg ttg ggt gac       1200
Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly Leu Gly Asp
385                 390                 395                 400

ttt ggt gac aga tac tac tgt gtt taa                                  1227
Phe Gly Asp Arg Tyr Tyr Cys Val
                405
```

<210> 12

<211> 408

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 12

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

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65                  70                  75                  80
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
                85                  90                  95
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
```

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Asn Pro

145 150 155 160

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

165 170 175

Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg Asn Leu Gln Lys

180 185 190

Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu Pro Gln Thr Asn

195 200 205

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

210 215 220

Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg Leu Leu Val Glu

225 230 235 240

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

245 250 255

Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly Lys Ile Cys Gly

260 265 270

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

275 280 285

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

290 295 300

Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu Pro Glu Asp Ile

305 310 315 320

Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu Ala Thr Gly Gly

325 330 335

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

340 345 350

Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys Glu Gly Ile Glu
355 360 365
Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val Thr Gly Ala Leu
370 375 380

Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro Gly Leu Gly Asp
385 390 395 400
Phe Gly Asp Arg Tyr Tyr Cys Val
405

<210> 13
<211> 1287
<212> DNA
<213> Artificial Sequence
<220><223> Fusion construction - CDopt - linker - UPRT
<220><221> CDS
<222> (1)..(1287)
<400> 13

atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15

atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96
Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35 40 45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
50 55 60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag 240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80

```
gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc        288

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
                85                  90                  95
gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg        336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
            100                 105                 110
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag        384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

        115                 120                 125
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc        432
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
    130                 135                 140
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tcc ggc        480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly
145                 150                 155                 160

ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc        528
Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly
                165                 170                 175
ggc gcc aac ccg tta ttc ttt ttg gct tct cca ttc ttg tac ctt aca        576
Gly Ala Asn Pro Leu Phe Phe Leu Ala Ser Pro Phe Leu Tyr Leu Thr
            180                 185                 190
tat ctt ata tat tat cca aac aaa ggg tct ttc gtt agc aaa cct aga        624

Tyr Leu Ile Tyr Tyr Pro Asn Lys Gly Ser Phe Val Ser Lys Pro Arg
        195                 200                 205
aat ctg caa aaa atg tct tcg gaa cca ttt aag aac gtc tac ttg cta        672
Asn Leu Gln Lys Met Ser Ser Glu Pro Phe Lys Asn Val Tyr Leu Leu
    210                 215                 220
cct caa aca aac caa ttg ctg ggt ttg tac acc atc atc aga aat aag        720
Pro Gln Thr Asn Gln Leu Leu Gly Leu Tyr Thr Ile Ile Arg Asn Lys
```

```
225                 230                 235                 240

aat aca act aga cct gat ttc att ttc tac tcc gat aga atc atc aga        768
Asn Thr Thr Arg Pro Asp Phe Ile Phe Tyr Ser Asp Arg Ile Ile Arg

                245                 250                 255

ttg ttg gtt gaa gaa ggt ttg aac cat cta cct gtg caa aag caa att        816
Leu Leu Val Glu Glu Gly Leu Asn His Leu Pro Val Gln Lys Gln Ile

            260                 265                 270

gtg gaa act gac acc aac gaa aac ttc gaa ggt gtc tca ttc atg ggt        864
Val Glu Thr Asp Thr Asn Glu Asn Phe Glu Gly Val Ser Phe Met Gly

        275                 280                 285

aaa atc tgt ggt gtt tcc att gtc aga gct ggt gaa tcg atg gag caa        912
Lys Ile Cys Gly Val Ser Ile Val Arg Ala Gly Glu Ser Met Glu Gln

    290                 295                 300

gga tta aga gac tgt tgt agg tct gtg cgt atc ggt aaa att tta att        960
Gly Leu Arg Asp Cys Cys Arg Ser Val Arg Ile Gly Lys Ile Leu Ile

305                 310                 315                 320

caa agg gac gag gag act gct tta cca aag tta ttc tac gaa aaa tta       1008
Gln Arg Asp Glu Glu Thr Ala Leu Pro Lys Leu Phe Tyr Glu Lys Leu

                325                 330                 335

cca gag gat ata tct gaa agg tat gtc ttc cta tta gac cca atg ctg       1056
Pro Glu Asp Ile Ser Glu Arg Tyr Val Phe Leu Leu Asp Pro Met Leu

            340                 345                 350

gcc acc ggt ggt agt gct atc atg gct aca gaa gtc ttg att aag aga       1104
Ala Thr Gly Gly Ser Ala Ile Met Ala Thr Glu Val Leu Ile Lys Arg

        355                 360                 365

ggt gtt aag cca gag aga att tac ttc tta aac cta atc tgt agt aag       1152
Gly Val Lys Pro Glu Arg Ile Tyr Phe Leu Asn Leu Ile Cys Ser Lys

    370                 375                 380

gaa ggg att gaa aaa tac cat gcc gcc ttc cca gag gtc aga att gtt       1200
Glu Gly Ile Glu Lys Tyr His Ala Ala Phe Pro Glu Val Arg Ile Val

385                 390                 395                 400
```

act ggt gcc ctc gac aga ggt cta gat gaa aac aag tat cta gtt cca 1248

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

405 410 415

ggg ttg ggt gac ttt ggt gac aga tac tac tgt gtt taa 1287

Gly Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val

420 425

<210> 14

<211> 428

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 14

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

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

20 25 30

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

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly

145 150 155 160

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

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

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

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

Thr Gly Ala Leu Asp Arg Gly Leu Asp Glu Asn Lys Tyr Leu Val Pro
405 410 415

Gly Leu Gly Asp Phe Gly Asp Arg Tyr Tyr Cys Val
420 425

<210> 15

<211> 1200

<212> DNA

<213> Artificial Sequence

<220><223> Fusion Construct - CDopt3 - OPRT

<220><221> CDS

<222> (1)..(1200)

<400> 15

atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48
Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15

atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96

Ile Ala Tyr Glu Glu Ala Leu Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg 144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35 40 45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag 192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag 240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80

gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc 288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95

gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg 336

```
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
            100                 105                 110
aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag        384
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
        115                 120                 125
gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc        432

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
    130                 135                 140
atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag gcg gtc        480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ala Val
145                 150                 155                 160
gct cgt gca gct ttg ggg cca ttg gtg acg ggt ctg tac gac gtg cag        528
Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr Asp Val Gln

                165                 170                 175
gct ttc aag ttt ggg gac ttc gtg ctg aag agc ggg ctt tcc tcc ccc        576
Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly Leu Ser Ser Pro
            180                 185                 190
atc tac atc gat ctg cgg ggc atc gtg tct cga ccg cgt ctt ctg agt        624
Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro Arg Leu Leu Ser
        195                 200                 205

cag gtt gca gat att tta ttc caa act gcc caa aat gca ggc atc agt        672
Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn Ala Gly Ile Ser
    210                 215                 220
ttt gac acc gtg tgt gga gtg cct tat aca gct ttg cca ttg gct aca        720
Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu Pro Leu Ala Thr
225                 230                 235                 240
gtt atc tgt tca acc aat caa att cca atg ctt att aga agg aaa gaa        768

Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg Arg Lys Glu
                245                 250                 255
aca aag gat tat gga act aag cgt ctt gta gaa gga act att aat cca        816
Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly Thr Ile Asn Pro
```

260 265 270

gga gaa acc tgt tta atc att gaa gat gtt gtc acc agt gga tct agt 864

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

275 280 285

gtt ttg gaa act gtt gag gtt ctt cag aag gag ggc ttg aag gtc act 912

Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu Lys Val Thr

290 295 300

gat gcc ata gtg ctg ttg gac aga gag cag gga ggc aag gac aag ttg 960

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

305 310 315 320

cag gcg cac ggg atc cgc ctc cac tca gtg tgt aca ttg tcc aaa atg 1008

Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu Ser Lys Met

325 330 335

ctg gag att ctc gag cag cag aaa aaa gtt gat gct gag aca gtt ggg 1056

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

340 345 350

aga gtg aag agg ttt att cag gag aat gtc ttt gtg gca gcg aat cat 1104

Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala Ala Asn His

355 360 365

aat ggt tct ccc ctt tct ata aag gaa gca ccc aaa gaa ctc agc ttc 1152

Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu Leu Ser Phe

370 375 380

ggt gca cgt gca gag ctg ccc agg atc cac cca gtt gca tcg aag taa 1200

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

385 390 395

<210> 16

<211> 399

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 16

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

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

            100                 105                 110
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
        115                 120                 125
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
    130                 135                 140
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ala Val
145                 150                 155                 160
Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu Tyr Asp Val Gln

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

225                 230                 235                 240
Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile Arg Arg Lys Glu
```

245 250 255

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

260 265 270

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

275 280 285

Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly Leu Lys Val Thr

290 295 300

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

305 310 315 320

Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr Leu Ser Lys Met

325 330 335

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

340 345 350

Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val Ala Ala Asn His

355 360 365

Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys Glu Leu Ser Phe

370 375 380

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

385 390 395

<210> 17

<211> 1260

<212> DNA

<213> Artificial Sequence

<220><223> Fusion Construct - CDopt3 - linker - OPRT

<220><221> CDS

<222> (1)..(1260)

<400> 17

atg gtg acc ggc ggc atg gcc tcc aag tgg gat caa aag ggc atg gat 48

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

atc gct tac gag gag gcc ctg ctg ggc tac aag gag ggc ggc gtg cct 96

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

```
            20                  25                  30

atc ggc ggc tgt ctg atc aac aac aag gac ggc agt gtg ctg ggc agg        144
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg

        35                  40                  45

ggc cac aac atg agg ttc cag aag ggc tcc gcc acc ctg cac ggc gag        192
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
    50                  55                  60

atc tcc acc ctg gag aac tgt ggc agg ctg gag ggc aag gtg tac aag        240
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65                  70                  75                  80

gac acc acc ctg tac acc acc ctg tcc cct tgt gac atg tgt acc ggc        288
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
                85                  90                  95

gct atc atc atg tac ggc atc cct agg tgt gtg atc ggc gag aac gtg        336
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val
            100                 105                 110

aac ttc aag tcc aag ggc gag aag tac ctg caa acc agg ggc cac gag        384

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
        115                 120                 125

gtg gtg gtt gtt gac gat gag agg tgt aag aag ctg atg aag cag ttc        432
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe
    130                 135                 140

atc gac gag agg cct cag gac tgg ttc gag gat atc ggc gag tcc ggc        480
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly

145                 150                 155                 160

ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc ggc gcc tcc ggc ggc        528
Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly
                165                 170                 175

ggc gcc gcg gtc gct cgt gca gct ttg ggg cca ttg gtg acg ggt ctg        576
Gly Ala Ala Val Ala Arg Ala Ala Leu Gly Pro Leu Val Thr Gly Leu
```

180 185 190

tac gac gtg cag gct ttc aag ttt ggg gac ttc gtg ctg aag agc ggg 624
Tyr Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly
195 200 205

ctt tcc tcc ccc atc tac atc gat ctg cgg ggc atc gtg tct cga ccg 672
Leu Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro
210 215 220

cgt ctt ctg agt cag gtt gca gat att tta ttc caa act gcc caa aat 720

Arg Leu Leu Ser Gln Val Ala Asp Ile Leu Phe Gln Thr Ala Gln Asn
225 230 235 240

gca ggc atc agt ttt gac acc gtg tgt gga gtg cct tat aca gct ttg 768
Ala Gly Ile Ser Phe Asp Thr Val Cys Gly Val Pro Tyr Thr Ala Leu
245 250 255

cca ttg gct aca gtt atc tgt tca acc aat caa att cca atg ctt att 816
Pro Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile

260 265 270

aga agg aaa gaa aca aag gat tat gga act aag cgt ctt gta gaa gga 864
Arg Arg Lys Glu Thr Lys Asp Tyr Gly Thr Lys Arg Leu Val Glu Gly
275 280 285

act att aat cca gga gaa acc tgt tta atc att gaa gat gtt gtc acc 912
Thr Ile Asn Pro Gly Glu Thr Cys Leu Ile Ile Glu Asp Val Val Thr
290 295 300

agt gga tct agt gtt ttg gaa act gtt gag gtt ctt cag aag gag ggc 960
Ser Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly
305 310 315 320

ttg aag gtc act gat gcc ata gtg ctg ttg gac aga gag cag gga ggc 1008
Leu Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly
325 330 335

aag gac aag ttg cag gcg cac ggg atc cgc ctc cac tca gtg tgt aca 1056

Lys Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr

340 345 350

ttg tcc aaa atg ctg gag att ctc gag cag cag aaa aaa gtt gat gct 1104

Leu Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala

355 360 365

gag aca gtt ggg aga gtg aag agg ttt att cag gag aat gtc ttt gtg 1152

Glu Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val

370 375 380

gca gcg aat cat aat ggt tct ccc ctt tct ata aag gaa gca ccc aaa 1200

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

385 390 395 400

gaa ctc agc ttc ggt gca cgt gca gag ctg ccc agg atc cac cca gtt 1248

Glu Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val

405 410 415

gca tcg aag taa 1260

Ala Ser Lys

<210> 18

<211> 419

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 18

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

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

20 25 30

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

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Ile Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Leu Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu Ser Gly

145 150 155 160

Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly Gly Ala Ser Gly Gly

165 170 175

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

180 185 190

Tyr Asp Val Gln Ala Phe Lys Phe Gly Asp Phe Val Leu Lys Ser Gly

195 200 205

Leu Ser Ser Pro Ile Tyr Ile Asp Leu Arg Gly Ile Val Ser Arg Pro

210 215 220

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

225 230 235 240

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

245 250 255

Pro Leu Ala Thr Val Ile Cys Ser Thr Asn Gln Ile Pro Met Leu Ile

260 265 270

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

275 280 285

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

290 295 300

Ser Gly Ser Ser Val Leu Glu Thr Val Glu Val Leu Gln Lys Glu Gly

305 310 315 320

Leu Lys Val Thr Asp Ala Ile Val Leu Leu Asp Arg Glu Gln Gly Gly
325 330 335
Lys Asp Lys Leu Gln Ala His Gly Ile Arg Leu His Ser Val Cys Thr

340 345 350
Leu Ser Lys Met Leu Glu Ile Leu Glu Gln Gln Lys Lys Val Asp Ala
355 360 365
Glu Thr Val Gly Arg Val Lys Arg Phe Ile Gln Glu Asn Val Phe Val
370 375 380
Ala Ala Asn His Asn Gly Ser Pro Leu Ser Ile Lys Glu Ala Pro Lys
385 390 395 400
Glu Leu Ser Phe Gly Ala Arg Ala Glu Leu Pro Arg Ile His Pro Val

405 410 415
Ala Ser Lys

<210> 19
<211> 11892
<212> DNA
<213> Artificial Sequence
<220><223> RCR Vector - pAC3-yCD2
<400> 19
tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300
catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540
acggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600
actgagtcgc ccgggtaccc gtgtatccaa taaaccctct tgcagttgca tccgacttgt 660

ggtctcgctg ttccttggga gggtctcctc tgagtgattg actacccgtc agcgggggtc 720

tttcatttgg gggctcgtcc gggatcggga gacccctgcc cagggaccac cgacccacca 780

ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840

tgattttatg cgcctgcgtc ggtactagtt agctaactag ctctgtatct ggcggacccg 900

tggtggaact gacgagttcg gaacacccgg ccgcaaccct gggagacgtc ccagggactt 960

cgggggccgt ttttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020

gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt 1080

tcccgcctcc gtctgaattt ttgctttcgg tttgggaccg aagccgcgcc gcgcgtcttg 1140

tctgctgcag catcgttctg tgttgtctct gtctgactgt gtttctgtat ttgtctgaga 1200

atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg 1260

agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt accttctgct 1320

ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc 1380

tcatcaccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg 1440

tcccctacat cgtgacctgg gaagccttgg cttttgaccc ccctccctgg gtcaagccct 1500

ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac 1560

ctcctcgttc gaccccgcct cgatcctccc tttatccagc cctcactcct tctctaggcg 1620

ccaaacctaa acctcaagtt ctttctgaca gtgggggggcc gctcatcgac ctacttacag 1680

aagacccccc gccttatagg gacccaagac cacccccttc cgacagggac ggaaatggtg 1740

gagaagcgac ccctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg 1800

ggagacggga gccccctgtg gccgactcca ctacctcgca ggcattcccc ctccgcgcag 1860

gaggaaacgg acagcttcaa tactggccgt tctcctcttc tgacctttac aactggaaaa 1920

ataataaccc ttctttttct gaagatccag gtaaactgac agctctgatc gagtctgttc 1980

tcatcaccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg 2040

gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc 2100

gccccactca actgcccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg 2160

attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcctagcgg 2220

gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag 2280

ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgcaggtaca 2340

ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc 2400

agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc 2460

ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa 2520

gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg 2580

agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca 2640

ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg 2700

atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtcccaaga 2760

aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg 2820

gaggtcaggg tcaggagccc ccccctgaac ccaggataac cctcaaagtc ggggggcaac 2880

ccgtcacctt cctggtagat actggggccc aacactccgt gctgacccaa aatcctggac 2940

ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga 3000

ccacggatcg caaagtacat ctagctaccg gtaaggtcac ccactctttc ctccatgtac 3060

cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc 3120

actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga 3180

ccctaaatat agaagatgag catcggctac atgagacctc aaaagagcca gatgtttctc 3240

tagggtccac atggctgtct gattttcctc aggcctgggc ggaaaccggg ggcatgggac 3300

tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca 3360

taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac 3420

tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctacccg 3480

ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc 3540

gggtggaaga catccacccc accgtgccca acccttacaa cctcttgagc gggctcccac 3600

cgtcccacca gtggtacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc 3660

accccaccag tcagcctctc ttcgcctttg agtggagaga tccagagatg ggaatctcag 3720

gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg 3780

aggcactgca cagagaccta gcagacttcc ggatccagca cccagacttg atcctgctac 3840

agtacgtgga tgacttactg ctggccgcca cttctgagct agactgccaa caaggtactc 3900

gggccctgtt acaaacccta gggaacctcg ggtatcgggc ctcggccaag aaagcccaaa 3960

tttgccagaa acaggtcaag tatctggggt atcttctaaa agagggtcag agatggctga 4020

ctgaggccag aaaagagact gtgatggggc agcctactcc gaagaccccct cgacaactaa 4080

gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggttt gcagaaatgg 4140

cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac 4200

aaaaggccta tcaagaaatc aagcaagctc ttctaactgc cccagccctg gggttgccag 4260

atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc 4320

taacgcaaaa actgggacct tggcgtcggc cggtggccta cctgtccaaa aagctagacc 4380

cagtagcagc tgggtggccc ccttgcctac ggatggtagc agccattgcc gtactgacaa 4440

aggatgcagg caagctaacc atgggacagc cactagtcat tctggccccc catgcagtag 4500
aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccggatg actcactatc 4560
aggccttgct tttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaacccgg 4620
ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg 4680
aagcccacgg aacccgaccc gacctaacgg accagccgct cccagacgcc gaccacacct 4740
ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcggga gctgcggtga 4800
ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg 4860

ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaatgttt 4920
atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc 4980
gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac 5040
taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg 5100
gacacagcgc cgaggctaga ggcaaccgga tggctgacca agcggcccga aaggcagcca 5160
tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcatcaccc tacacctcag 5220
aacattttca ttacacagtg actgatataa aggacctaac caagttgggg gccatttatg 5280

ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt 5340
ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc 5400
tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata 5460
tcactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg 5520
gaactagggt ccgcgggcat cggcccggca ctcattggga gatcgatttc accgagataa 5580
agcccggatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga 5640
tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg 5700

agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg 5760
tctccaaggt gagtcagaca gtggccgatc tgttggggat tgattggaaa ttacattgtg 5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt 5880
taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc 5940
tgtaccgagc ccgcaacacg ccgggccccc atggcctcac cccatatgag atcttatatg 6000
gggcaccccc gcccccttgta aacttccctg accctgacat gacaagagtt actaacagcc 6060
cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc 6120

tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg 6180

gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac 6240

cttacacagt cctgctgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga 6300

tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat 6360

ggcgcgttca acgctctcaa aacccccctca agataagatt aacccgtgga agcccttaat 6420

agtcatggga gtcctgttag gagtagggat ggcagagagc ccccatcagg tctttaatgt 6480

aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg 6540

aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga 6600

gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag 6660

acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcggggtg 6720

tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta 6780

ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca ccccctggga 6840

cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc 6900

cttccaaggg gctactcgag ggggcagatg caaccctcta gtcctagaat tcactgatgc 6960

aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg 7020

aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt 7080

ccccataggg cccaacccag tattacccga ccaaagactc ccttcctcac caatagagat 7140

tgtaccggct ccacagccac ctagcccccct caataccagt tacccccctt ccactaccag 7200

tacaccctca acctccccta caagtccaag tgtcccacag ccacccccag gaactggaga 7260

tagactacta gctctagtca aaggagccta tcaggcgctt aacctcacca atcccgacaa 7320

gacccaagaa tgttggctgt gcttagtgtc gggacctcct tattacgaag gagtagcggt 7380

cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca 7440

taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac 7500

tcaccaggcc ttatgtaaca ccacccaaag cgccggctca ggatcctact accttgcagc 7560

acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt 7620

gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca 7680

ctcccccgat tatatgtatg gtcagcttga acagcgtacc aaatataaaa gagagccagt 7740

atcattgacc ctggcccttc tactaggagg attaaccatg ggagggattg cagctggaat 7800

agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat 7860

ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc 7920

gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg 7980

aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt 8040

gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac 8100

aggccaagga tggttcgaag ggctgtttaa tagatccccc tggtttacca ccttaatctc 8160

caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct 8220

caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac 8280

tcagcaatat caccagctaa aacccataga gtacgagcca tgaacgcgtt actggccgaa 8340

gccgcttgga ataaggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt 8400

cttttggcaa tgtgagggcc cggaaacctg gccctgtctt cttgacgagc attcctaggg 8460

gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc 8520

ctctggaagc ttcttgaaga caaacaacgt ctgtagcgac cctttgcagg cagcggaacc 8580

ccccacctgg cgacaggtgc ctctgcggcc aaaagccacg tgtataagat acacctgcaa 8640

aggcggcaca accccagtgc cacgttgtga gttggatagt tgtggaaaga gtcaaatggc 8700

tctcctcaag cgtattcaac aaggggctga aggatgccca gaaggtaccc cattgtatgg 8760

gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac 8820

gtctaggccc cccgaaccac ggggacgtgg ttttcctttg aaaaacacga ttataaatgg 8880

tgaccggcgg catggcctcc aagtgggatc aaaagggcat ggatatcgct tacgaggagg 8940

ccctgctggg ctacaaggag ggcggcgtgc ctatcggcgg ctgtctgatc aacaacaagg 9000

acggcagtgt gctgggcagg ggccacaaca tgaggttcca gaagggctcc gccaccctgc 9060

acggcgagat ctccaccctg gagaactgtg gcaggctgga gggcaaggtg tacaaggaca 9120

ccaccctgta caccaccctg tccccttgtg acatgtgtac cggcgctatc atcatgtacg 9180

gcatccctag gtgtgtgatc ggcgagaacg tgaacttcaa gtccaagggc gagaagtacc 9240

tgcaaaccag gggccacgag gtggtggttg ttgacgatga gaggtgtaag aagctgatga 9300

agcagttcat cgacgagagg cctcaggact ggttcgagga tatcggcgag taagcggccg 9360

cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agaccccacc 9420

tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcatggaa aaatacataa 9480

ctgagaatag agaagttcag atcaaggtca ggaacagatg gaacagctga atatgggcca 9540

aacaggatat ctgtggtaag cagttcctgc cccggctcag ggccaagaac agatggaaca 9600

gctgaatatg ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctcagggcca 9660

agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccatcagatg 9720

tttccagggt gccccaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag 9780

ttcgcttctc gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa 9840

```
cccctcactc ggggcgccag tcctccgatt gactgagtcg cccgggtacc cgtgtatcca     9900

ataaaccctc ttgcagttgc atccgacttg tggtctcgct gttccttggg agggtctcct     9960
ctgagtgatt gactacccgt cagcgggggt ctttcattac atgtgagcaa aaggccagca    10020
aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc tccgcccccc    10080
tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga caggactata    10140
aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc    10200
gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt ctcaatgctc    10260
acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga    10320

acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc    10380
ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag    10440
gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct acactagaag    10500
gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gagttggtag    10560
ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt gcaagcagca    10620
gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga    10680
cgctcagtgg aacgaaaact cacgttaagg gattttggtc atgagattat caaaaaggat    10740

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga    10800
gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg    10860
tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga    10920
gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc    10980
agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac    11040
tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc    11100
agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt cacgctcgtc    11160

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc    11220
catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt    11280
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc    11340
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg    11400
tatgcggcga ccgagttgct cttgcccggc gtcaacacgg gataataccg cgccacatag    11460
cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat    11520
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc    11580
```

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 11640

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 11700

ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa 11760

aaataaacaa ataggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga 11820

aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtct 11880

tcaagaattc at 11892

<210> 20

<211> 11892

<212> DNA

<213> Artificial Sequence

<220><223> RCR Vector - pAC3-yCD

<400> 20

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600

actgagtcgc ccgggtaccc gtgtatccaa taaaccctct tgcagttgca tccgacttgt 660

ggtctcgctg ttccttggga gggtctcctc tgagtgattg actacccgtc agcgggggtc 720

tttcatttgg gggctcgtcc gggatcggga gacccctgcc cagggaccac cgacccacca 780

ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840

tgattttatg cgcctgcgtc ggtactagtt agctaactag ctctgtatct ggcggacccg 900

tggtggaact gacgagttcg gaacacccgg ccgcaaccct gggagacgtc ccagggactt 960

cgggggccgt ttttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020

gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt 1080

tcccgcctcc gtctgaattt ttgctttcgg tttgggaccg aagccgcgcc gcgcgtcttg 1140

tctgctgcag catcgttctg tgttgtctct gtctgactgt gtttctgtat ttgtctgaga 1200

atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg 1260

agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt accttctgct 1320

ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc 1380

tcatcaccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg 1440

tcccctacat cgtgacctgg gaagccttgg cttttgaccc ccctccctgg gtcaagccct 1500

ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac 1560

ctcctcgttc gaccccgcct cgatcctccc tttatccagc cctcactcct tctctaggcg 1620

ccaaacctaa acctcaagtt ctttctgaca gtgggggggcc gctcatcgac ctacttacag 1680

aagacccccc gccttatagg gacccaagac cacccccttc cgacagggac ggaaatggtg 1740

gagaagcgac ccctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg 1800

ggagacggga gcccccctgtg gccgactcca ctacctcgca ggcattcccc ctccgcgcag 1860

gaggaaacgg acagcttcaa tactggccgt tctcctcttc tgacctttac aactggaaaa 1920

ataataaccc ttctttttct gaagatccag gtaaactgac agctctgatc gagtctgttc 1980

tcatcaccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg 2040

gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc 2100

gccccactca actgcccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg 2160

attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcctagcgg 2220

gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag 2280

ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgcaggtaca 2340

ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc 2400

agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc 2460

ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa 2520

gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg 2580

agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca 2640

ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg 2700

atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtcccaaga 2760

aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg 2820

gaggtcaggg tcaggagccc ccccctgaac ccaggataac cctcaaagtc ggggggcaac 2880

ccgtcacctt cctggtagat actggggccc aacactccgt gctgacccaa aatcctggac 2940

ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga 3000

ccacggatcg caaagtacat ctagctaccg gtaaggtcac ccactctttc ctccatgtac 3060

cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc 3120

actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga 3180

ccctaaatat agaagatgag catcggctac atgagacctc aaaagagcca gatgtttctc 3240

tagggtccac atggctgtct gattttcctc aggcctgggc ggaaaccggg ggcatgggac 3300

tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca 3360

taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac 3420

tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctacccg 3480

ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc 3540

gggtggaaga catccacccc accgtgccca acccttacaa cctcttgagc gggctcccac 3600

cgtcccacca gtggtacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc 3660

accccaccag tcagcctctc ttcgcctttg agtggagaga tccagagatg ggaatctcag 3720

gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg 3780

aggcactgca cagagaccta gcagacttcc ggatccagca cccagacttg atcctgctac 3840

agtacgtgga tgacttactg ctggccgcca cttctgagct agactgccaa caaggtactc 3900

gggccctgtt acaaacccta gggaacctcg ggtatcgggc ctcggccaag aaagcccaaa 3960

tttgccagaa acaggtcaag tatctggggt atcttctaaa agagggtcag agatggctga 4020

ctgaggccag aaaagagact gtgatggggc agcctactcc gaagacccct cgacaactaa 4080

gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggttt gcagaaatgg 4140

cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac 4200

aaaaggccta tcaagaaatc aagcaagctc ttctaactgc cccagccctg gggttgccag 4260

atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc 4320

taacgcaaaa actgggacct tggcgtcggc cggtggccta cctgtccaaa aagctagacc 4380

cagtagcagc tgggtggccc ccttgcctac ggatggtagc agccattgcc gtactgacaa 4440

aggatgcagg caagctaacc atgggacagc cactagtcat tctggccccc catgcagtag 4500

aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccggatg actcactatc 4560

aggccttgct tttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaacccgg 4620

ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg 4680

aagcccacgg aacccgaccc gacctaacgg accagccgct cccagacgcc gaccacacct 4740

ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcggga gctgcggtga 4800

ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg 4860

ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaatgttt 4920

atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc 4980

gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac 5040

taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg 5100

gacacagcgc cgaggctaga ggcaaccgga tggctgacca agcggcccga aaggcagcca 5160

tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcatcaccc tacacctcag 5220

aacattttca ttacacagtg actgatataa aggacctaac caagttgggg gccatttatg 5280

ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt 5340

ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc 5400

tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata 5460

tcactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg 5520

gaactagggt ccgcgggcat cggcccggca ctcattggga gatcgatttc accgagataa 5580

agcccggatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga 5640

tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg 5700

agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg 5760

tctccaaggt gagtcagaca gtggccgatc tgttggggat tgattggaaa ttacattgtg 5820

catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt 5880

taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc 5940

tgtaccgagc ccgcaacacg ccgggccccc atggcctcac cccatatgag atcttatatg 6000

gggcaccccc gcccccttgta aacttccctg accctgacat gacaagagtt actaacagcc 6060

cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc 6120

tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg 6180

gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac 6240

cttacacagt cctgctgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga 6300

tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat 6360

ggcgcgttca acgctctcaa aacccccctca agataagatt aacccgtgga agcccttaat 6420

agtcatggga gtcctgttag gagtagggat ggcagagagc ccccatcagg tctttaatgt 6480

aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg 6540

aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga 6600

gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag 6660

acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcggggtg 6720

tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta 6780
ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca ccccctggga 6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc 6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtcctagaat tcactgatgc 6960
aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg 7020
aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt 7080
ccccataggg cccaacccag tattacccga ccaaagactc ccttcctcac caatagagat 7140

tgtaccggct ccacagccac ctagcccccc caataccagt taccccccct ccactaccag 7200
tacaccctca acctccccta caagtccaag tgtcccacag ccacccccag gaactggaga 7260
tagactacta gctctagtca aaggagccta tcaggcgctt aacctcacca atcccgacaa 7320
gacccaagaa tgttggctgt gcttagtgtc gggacctcct tattacgaag gagtagcggt 7380
cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca 7440
taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac 7500
tcaccaggcc ttatgtaaca ccacccaaag cgccggctca ggatcctact accttgcagc 7560

acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt 7620
gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca 7680
ctcccccgat tatatgtatg gtcagcttga acagcgtacc aaatataaaa gagagccagt 7740
atcattgacc ctggcccttc tactaggagg attaaccatg ggagggattg cagctggaat 7800
agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat 7860
ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc 7920
gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg 7980

aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt 8040
gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac 8100
aggccaagga tggttcgaag ggctgtttaa tagatccccc tggtttacca ccttaatctc 8160
caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct 8220
caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac 8280
tcagcaatat caccagctaa aacccataga gtacgagcca tgaacgcgtt actggccgaa 8340
gccgcttgga ataaggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt 8400

ctttggcaa tgtgagggcc cggaaacctg gccctgtctt cttgacgagc attcctaggg 8460

gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc 8520

ctctggaagc ttcttgaaga caaacaacgt ctgtagcgac cctttgcagg cagcggaacc 8580

ccccacctgg cgacaggtgc ctctgcggcc aaaagccacg tgtataagat acacctgcaa 8640

aggcggcaca accccagtgc cacgttgtga gttggatagt tgtggaaaga gtcaaatggc 8700

tctcctcaag cgtattcaac aaggggctga aggatgccca gaaggtaccc cattgtatgg 8760

gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac 8820

gtctaggccc cccgaaccac ggggacgtgg ttttcctttg aaaaacacga ttataaatgg 8880

tgacaggggg aatggcaagc aagtgggatc agaagggtat ggacattgcc tatgaggagg 8940

cggccttagg ttacaaagag ggtggtgttc ctattggcgg atgtcttatc aataacaaag 9000

acggaagtgt tctcggtcgt ggtcacaaca tgagatttca aaagggatcc gccacactac 9060

atggtgagat ctccactttg gaaaactgtg ggagattaga gggcaaagtg tacaaagata 9120

ccactttgta tacgacgctg tctccatgcg acatgtgtac aggtgccatc atcatgtatg 9180

gtattccacg ctgtgttgtc ggtgagaacg ttaatttcaa aagtaagggc gagaaatatt 9240

tacaaactag aggtcacgag gttgttgttg ttgacgatga gaggtgtaaa aagatcatga 9300

aacaatttat cgatgaaaga cctcaggatt ggtttgaaga tattggtgag taggcggccg 9360

cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agaccccacc 9420

tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcatggaa aaatacataa 9480

ctgagaatag agaagttcag atcaaggtca ggaacagatg gaacagctga atatgggcca 9540

aacaggatat ctgtggtaag cagttcctgc cccggctcag ggccaagaac agatggaaca 9600

gctgaatatg ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctcagggcca 9660

agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccatcagatg 9720

tttccagggt gccccaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag 9780

ttcgcttctc gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa 9840

cccctcactc ggggcgccag tcctccgatt gactgagtcg cccgggtacc cgtgtatcca 9900

ataaaccctc ttgcagttgc atccgacttg tggtctcgct gttccttggg agggtctcct 9960

ctgagtgatt gactacccgt cagcgggggt ctttcattac atgtgagcaa aaggccagca 10020

aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc tccgcccccc 10080

tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga caggactata 10140

aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc 10200

gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt ctcaatgctc 10260

acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga 10320

acccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc 10380

ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag 10440

gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct acactagaag 10500

gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gagttggtag 10560

ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt gcaagcagca 10620

gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga 10680

cgctcagtgg aacgaaaact cacgttaagg gattttggtc atgagattat caaaaaggat 10740

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 10800

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 10860

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 10920

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 10980

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 11040

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 11100

agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt cacgctcgtc 11160

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 11220

catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 11280

ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 11340

atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 11400

tatgcggcga ccgagttgct cttgcccggc gtcaacacgg gataataccg cgccacatag 11460

cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 11520

cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 11580

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 11640

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 11700

ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa 11760

aaataaacaa atagggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga 11820

aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtct 11880

tcaagaattc at 11892

<210> 21

<211> 12007

<212> DNA

<213> Artificial Sequence

<220><223> RCR Vector - pACE-CD

<400> 21

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600

actgagtcgc ccgggtaccc gtgtatccaa taaaccctct tgcagttgca tccgacttgt 660

ggtctcgctg ttccttggga gggtctcctc tgagtgattg actacccgtc agcgggggtc 720

tttcatttgg gggctcgtcc gggatcggga gacccctgcc cagggaccac cgacccacca 780

ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840

tgattttatg cgcctgcgtc ggtactagtt agctaactag ctctgtatct ggcggacccg 900

tggtggaact gacgagttcg gaacacccgg ccgcaaccct gggagacgtc ccagggactt 960

cgggggccgt ttttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020

gtgcacccct cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt 1080

tcccgcctcc gtctgaattt ttgctttcgg tttgggaccg aagccgcgcc gcgcgtcttg 1140

tctgctgcag catcgttctg tgttgtctct gtctgactgt gtttctgtat ttgtctgaga 1200

atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg 1260

agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt accttctgct 1320

ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc 1380

tcatcaccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg 1440

tcccctacat cgtgacctgg gaagccttgg cttttgaccc ccctccctgg gtcaagccct 1500

ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac 1560

ctcctcgttc gaccccgcct cgatcctccc tttatccagc cctcactcct tctctaggcg 1620

ccaaacctaa acctcaagtt ctttctgaca gtgggggggcc gctcatcgac ctacttacag 1680

aagacccccc gccttatagg gacccaagac cacccccttc cgacagggac ggaaatggtg 1740

gagaagcgac ccctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg 1800

ggagacggga gccccctgtg gccgactcca ctacctcgca ggcattcccc ctccgcgcag 1860

gaggaaacgg acagcttcaa tactggccgt tctcctcttc tgacctttac aactggaaaa 1920

ataataaccc ttctttttct gaagatccag gtaaactgac agctctgatc gagtctgttc 1980

tcatcaccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg 2040

gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc 2100

gccccactca actgcccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg 2160

attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcctagcgg 2220

gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag 2280

ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgcaggtaca 2340

ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc 2400

agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc 2460

ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa 2520

gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg 2580

agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca 2640

ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg 2700

atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtcccaaga 2760

aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg 2820

gaggtcaggg tcaggagccc ccccctgaac ccaggataac cctcaaagtc ggggggcaac 2880

ccgtcacctt cctggtagat actggggccc aacactccgt gctgacccaa aatcctggac 2940

ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga 3000

ccacggatcg caaagtacat ctagctaccg gtaaggtcac ccactctttc ctccatgtac 3060

cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc 3120

actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga 3180

ccctaaatat agaagatgag catcggctac atgagacctc aaaagagcca gatgtttctc 3240

tagggtccac atggctgtct gattttcctc aggcctgggc ggaaaccggg ggcatgggac 3300

tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca 3360

taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac 3420

tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctacccg 3480

ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc 3540

gggtggaaga catccacccc accgtgccca acccttacaa cctcttgagc gggctcccac 3600

cgtcccacca gtggtacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc 3660

accccaccag tcagcctctc ttcgcctttg agtggagaga tccagagatg ggaatctcag 3720

gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg 3780

aggcactgca cagagaccta gcagacttcc ggatccagca cccagacttg atcctgctac 3840

agtacgtgga tgacttactg ctggccgcca cttctgagct agactgccaa caaggtactc 3900

gggccctgtt acaaacccta gggaacctcg ggtatcgggc ctcggccaag aaagcccaaa 3960

tttgccagaa acaggtcaag tatctggggt atcttctaaa agagggtcag agatggctga 4020

ctgaggccag aaaagagact gtgatggggc agcctactcc gaagaccccct cgacaactaa 4080

gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggttt gcagaaatgg 4140

cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac 4200

aaaaggccta tcaagaaatc aagcaagctc ttctaactgc cccagccctg gggttgccag 4260

atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc 4320

taacgcaaaa actgggacct tggcgtcggc cggtggccta cctgtccaaa aagctagacc 4380

cagtagcagc tgggtggccc ccttgcctac ggatggtagc agccattgcc gtactgacaa 4440

aggatgcagg caagctaacc atgggacagc cactagtcat tctggccccc catgcagtag 4500

aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccggatg actcactatc 4560

aggccttgct tttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaacccgg 4620

ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg 4680

aagcccacgg aacccgaccc gacctaacgg accagccgct cccagacgcc gaccacacct 4740

ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcggga gctgcggtga 4800

ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg 4860

ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaatgttt 4920

atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc 4980

gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac 5040

taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg 5100

gacacagcgc cgaggctaga ggcaaccgga tggctgacca agcggcccga aaggcagcca 5160

tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcatcaccc tacacctcag 5220

aacattttca ttacacagtg actgatataa aggacctaac caagttgggg gccatttatg 5280

ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt 5340

ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc 5400

tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata 5460

tcactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg 5520

gaactagggt ccgcgggcat cggcccggca ctcattggga gatcgatttc accgagataa 5580

agcccggatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga 5640

tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg 5700

agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg 5760

tctccaaggt gagtcagaca gtggccgatc tgttggggat tgattggaaa ttacattgtg 5820

catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt 5880

taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc 5940

tgtaccgagc ccgcaacacg ccgggccccc atggcctcac cccatatgag atcttatatg 6000

gggcaccccc gcccccttgta aacttccctg accctgacat gacaagagtt actaacagcc 6060

cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc 6120

tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg 6180

gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac 6240

cttacacagt cctgctgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga 6300

tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat 6360

ggcgcgttca acgctctcaa aacccccтca agataagatt aacccgtgga agcccttaat 6420

agtcatggga gtcctgttag gagtagggat ggcagagagc ccccatcagg tctttaatgt 6480

aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg 6540

aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga 6600

gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag 6660

acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcggggtg 6720

tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta 6780

ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca ccccctggga 6840

cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc 6900

cttccaaggg gctactcgag ggggcagatg caaccctcta gtcctagaat tcactgatgc 6960

aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg 7020

aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt 7080

ccccataggg cccaacccag tattacccga ccaaagactc ccttcctcac caatagagat 7140

tgtaccggct ccacagccac ctagcccccc caataccagt taccccccct ccactaccag 7200

tacaccctca acctccccta caagtccaag tgtcccacag ccacccccag gaactggaga 7260

tagactacta gctctagtca aaggagccta tcaggcgctt aacctcacca atcccgacaa 7320

gacccaagaa tgttggctgt gcttagtgtc gggacctcct tattacgaag gagtagcggt 7380

cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca 7440

taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac 7500

tcaccaggcc ttatgtaaca ccacccaaag cgccggctca ggatcctact accttgcagc 7560

acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt 7620

gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca 7680

ctcccccgat tatatgtatg gtcagcttga acagcgtacc aaatataaaa gagagccagt 7740

atcattgacc ctggcccttc tactaggagg attaaccatg ggagggattg cagctggaat 7800

agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat 7860

ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc 7920

gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg 7980

aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt 8040

gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac 8100

aggccaagga tggttcgaag ggctgtttaa tagatccccc tggtttacca ccttaatctc 8160

caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct 8220

caatcgatta gtccaatttg ttaaagacag gatatcagtg gtccaggctc tagttttgac 8280

tcaacaatat caccagctga agcctataga gtacgagcca tgacgtacgt tactggccga 8340

agccgcttgg aataaggccg gtgtgcgttt gtctatatgt tattttccac catattgccg 8400

tcttttggca atgtgagggc ccggaaacct ggccctgtct tcttgacgag cattcctagg 8460

ggtctttccc ctctcgccaa aggaatgcaa ggtctgttga atgtcgtgaa ggaagcagtt 8520

cctctggaag cttcttgaag acaaacaacg tctgtagcga ccctttgcag gcagcggaac 8580

cccccacctg gcgacaggtg cctctgcggc caaaagccac gtgtataaga tacacctgca 8640

aaggcggcac aaccccagtg ccacgttgtg agttggatag ttgtggaaag agtcaaatgg 8700

ctctcctcaa gcgtattcaa caaggggctg aaggatgccc agaaggtacc ccattgtatg 8760

ggatctgatc tggggcctcg gtgcacatgc tttacatgtg tttagtcgag gttaaaaaaa 8820

cgtctaggcc ccccgaacca cggggacgtg gttttccttt gaaaaacacg ataataccat 8880

ggtgacaggg ggaatggcaa gcaagtggga tcagaagggt atggacattg cctatgagga 8940

ggcggcctta ggttacaaag agggtggtgt tcctattggc ggatgtctta tcaataacaa 9000

agacggaagt gttctcggtc gtggtcacaa catgagattt caaaagggat ccgccacact 9060
acatggtgag atctccactt tggaaaactg tgggagatta gagggcaaag tgtacaaaga 9120
taccactttg tatacgacgc tgtctccatg cgacatgtgt acaggtgcca tcatcatgta 9180
tggtattcca cgctgtgttg tcggtgagaa cgttaatttc aaaagtaagg gcgagaaata 9240
tttacaaact agaggtcacg aggttgttgt tgttgacgat gagaggtgta aaaagatcat 9300
gaaacaattt atcgatgaaa gacctcagga ttggtttgaa gatattggtg agtaggcggc 9360
cgcgccatag ataaaataaa agattttatt tagtctccag aaaaaggggg gaatgaaaga 9420

ccccacctgt aggtttggca agctagctta agtaacgcca ttttgcaagg catggaaaaa 9480
tacataactg agaatagaga agttcagatc aaggtcagga acagatggaa cagctgaata 9540
tgggccaaac aggatatctg tggtaagcag ttcctgcccc ggctcagggc caagaacaga 9600
tggaacagct gaatatgggc caaacaggat atctgtggta agcagttcct gccccggctc 9660
agggccaaga acagatggtc cccagatgcg gtccagccct cagcagtttc tagagaacca 9720
tcagatgttt ccagggtgcc ccaaggacct gaaatgaccc tgtgccttgt ttaaactaac 9780
caatcagttc gcttctcgct tctgttcgcg cgcttctgct ccccgagctc aataaaagag 9840

cccacaaccc ctcactcggg gcgccagtcc tccgattgac tgagtcgccc gggtacccgt 9900
gtatccaata aaccctcttg cagttgcatc cgacttgtgg tctcgctgtt ccttgggagg 9960
gtctcctctg agtgattgac tacccgtcag cgggggtctt tcatttgggg gctcgtccgg 10020
gatcgggaga cccctgccca gggaccaccg acccaccacc gggaggtaag ctggctgcct 10080
cgcgcgtttc ggtgatgacg gtgaaaacct ctgacatgtg agcaaaaggc cagcaaaagg 10140
ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg 10200
agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat 10260

accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta 10320
ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct 10380
gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc 10440
ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa 10500
gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg 10560
taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag 10620
tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt 10680

gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta 10740

cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc 10800

agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca 10860

cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa 10920

cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat 10980

ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct 11040

taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt 11100

tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat 11160

ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta 11220

atagtttgcg caacgttgtt gccattgctg caggcatcgt ggtgtcacgc tcgtcgtttg 11280

gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt 11340

tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg 11400

cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg 11460

taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc 11520

ggcgaccgag ttgctcttgc ccggcgtcaa cacgggataa taccgcgcca catagcagaa 11580

ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac 11640

cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt 11700

ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg 11760

gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa 11820

gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata 11880

aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc taagaaacca 11940

ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt cgtcttcaag 12000

aattcat 12007

<210> 22

<211> 11893

<212> DNA

<213> Artificial Sequence

<220><223> RCR Vector - pAC3-yCD2

<400> 22

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cggcgccagt cctccgattg 600

actgagtcgc ccgggtaccc gtgtatccaa taaaccctct tgcagttgca tccgacttgt 660

ggtctcgctg ttccttggga gggtctcctc tgagtgattg actacccgtc agcgggggtc 720

tttcatttgg gggctcgtcc gggatcggga gacccctgcc cagggaccac cgacccacca 780

ccgggaggta agctggccag caacttatct gtgtctgtcc gattgtctag tgtctatgac 840

tgattttatg cgcctgcgtc ggtactagtt agctaactag ctctgtatct ggcggacccg 900

tggtggaact gacgagttcg gaacacccgg ccgcaaccct gggagacgtc ccagggactt 960

cgggggccgt ttttgtggcc cgacctgagt ccaaaaatcc cgatcgtttt ggactctttg 1020

gtgcaccccc cttagaggag ggatatgtgg ttctggtagg agacgagaac ctaaaacagt 1080

tcccgcctcc gtctgaattt ttgctttcgg tttgggaccg aagccgcgcc gcgcgtcttg 1140

tctgctgcag catcgttctg tgttgtctct gtctgactgt gtttctgtat ttgtctgaaa 1200

atatgggcca gactgttacc actcccttaa gtttgacctt aggtcactgg aaagatgtcg 1260

agcggatcgc tcacaaccag tcggtagatg tcaagaagag acgttgggtt accttctgct 1320

ctgcagaatg gccaaccttt aacgtcggat ggccgcgaga cggcaccttt aaccgagacc 1380

tcatcaccca ggttaagatc aaggtctttt cacctggccc gcatggacac ccagaccagg 1440

tcccctacat cgtgacctgg gaagccttgg cttttgaccc ccctccctgg gtcaagccct 1500

ttgtacaccc taagcctccg cctcctcttc ctccatccgc cccgtctctc ccccttgaac 1560

ctcctcgttc gaccccgcct cgatcctccc tttatccagc cctcactcct tctctaggcg 1620

ccaaacctaa acctcaagtt ctttctgaca gtgggggggcc gctcatcgac ctacttacag 1680

aagacccccc gccttatagg gacccaagac cacccccttc cgacagggac ggaaatggtg 1740

gagaagcgac ccctgcggga gaggcaccgg acccctcccc aatggcatct cgcctacgtg 1800

ggagacggga gccccctgtg gccgactcca ctacctcgca ggcattcccc ctccgcgcag 1860

gaggaaacgg acagcttcaa tactggccgt tctcctcttc tgacctttac aactggaaaa 1920

ataataaccc ttctttttct gaagatccag gtaaactgac agctctgatc gagtctgtcc 1980

tcatcaccca tcagcccacc tgggacgact gtcagcagct gttggggact ctgctgaccg 2040

gagaagaaaa acaacgggtg ctcttagagg ctagaaaggc ggtgcggggc gatgatgggc 2100

gccccactca actgcccaat gaagtcgatg ccgcttttcc cctcgagcgc ccagactggg 2160

attacaccac ccaggcaggt aggaaccacc tagtccacta tcgccagttg ctcctagcgg 2220

gtctccaaaa cgcgggcaga agccccacca atttggccaa ggtaaaagga ataacacaag 2280

ggcccaatga gtctccctcg gccttcctag agagacttaa ggaagcctat cgcaggtaca 2340

ctccttatga ccctgaggac ccagggcaag aaactaatgt gtctatgtct ttcatttggc 2400

agtctgcccc agacattggg agaaagttag agaggttaga agatttaaaa aacaagacgc 2460

ttggagattt ggttagagag gcagaaaaga tctttaataa acgagaaacc ccggaagaaa 2520

gagaggaacg tatcaggaga gaaacagagg aaaaagaaga acgccgtagg acagaggatg 2580

agcagaaaga gaaagaaaga gatcgtagga gacatagaga gatgagcaag ctattggcca 2640

ctgtcgttag tggacagaaa caggatagac agggaggaga acgaaggagg tcccaactcg 2700

atcgcgacca gtgtgcctac tgcaaagaaa aggggcactg ggctaaagat tgtcccaaga 2760

aaccacgagg acctcgggga ccaagacccc agacctccct cctgacccta gatgactagg 2820

gaggtcaggg tcaggagccc ccccctgaac ccaggataac cctcaaagtc ggggggcaac 2880

ccgtcacctt cctggtagat actggggccc aacactccgt gctgacccaa aatcctggac 2940

ccctaagtga taagtctgcc tgggtccaag gggctactgg aggaaagcgg tatcgctgga 3000

ccacggatcg caaagtacat ctagctaccg gtaaggtcac ccactctttc ctccatgtac 3060

cagactgtcc ctatcctctg ttaggaagag atttgctgac taaactaaaa gcccaaatcc 3120

actttgaggg atcaggagcc caggttatgg gaccaatggg gcagcccctg caagtgttga 3180

ccctaaatat agaagatgag tatcggctac atgagacctc aaaagagcca gatgtttctc 3240

tagggtccac atggctgtct gattttcctc aggcctgggc ggaaaccggg ggcatgggac 3300

tggcagttcg ccaagctcct ctgatcatac ctctgaaagc aacctctacc cccgtgtcca 3360

taaaacaata ccccatgtca caagaagcca gactggggat caagccccac atacagagac 3420

tgttggacca gggaatactg gtaccctgcc agtccccctg gaacacgccc ctgctacccg 3480

ttaagaaacc agggactaat gattataggc ctgtccagga tctgagagaa gtcaacaagc 3540

gggtggaaga catccacccc accgtgccca acccttacaa cctcttgagc gggctcccac 3600

cgtcccacca gtggtacact gtgcttgatt taaaggatgc ctttttctgc ctgagactcc 3660

accccaccag tcagcctctc ttcgcctttg agtggagaga tccagagatg ggaatctcag 3720

gacaattgac ctggaccaga ctcccacagg gtttcaaaaa cagtcccacc ctgtttgatg 3780

aggcactgca cagagaccta gcagacttcc ggatccagca cccagacttg atcctgctac 3840

agtacgtgga tgacttactg ctggccgcca cttctgagct agactgccaa caaggtactc 3900

gggccctgtt acaaacccta gggaacctcg ggtatcgggc ctcggccaag aaagcccaaa 3960

tttgccagaa acaggtcaag tatctggggt atcttctaaa agagggtcag agatggctga 4020

ctgaggccag aaaagagact gtgatggggc agcctactcc gaagacccct cgacaactaa 4080

gggagttcct agggacggca ggcttctgtc gcctctggat ccctgggttt gcagaaatgg 4140

cagccccctt gtaccctctc accaaaacgg ggactctgtt taattggggc ccagaccaac 4200

aaaaggccta tcaagaaatc aagcaagctc ttctaactgc cccagccctg gggttgccag 4260

atttgactaa gccctttgaa ctctttgtcg acgagaagca gggctacgcc aaaggtgtcc 4320

taacgcaaaa actgggacct tggcgtcggc cggtggccta cctgtccaaa aagctagacc 4380

cagtagcagc tgggtggccc ccttgcctac ggatggtagc agccattgcc gtactgacaa 4440

aggatgcagg caagctaacc atgggacagc cactagtcat tctggccccc catgcagtag 4500

aggcactagt caaacaaccc cccgaccgct ggctttccaa cgcccggatg actcactatc 4560

aggccttgct tttggacacg gaccgggtcc agttcggacc ggtggtagcc ctgaacccgg 4620

ctacgctgct cccactgcct gaggaagggc tgcaacacaa ctgccttgat atcctggccg 4680

aagcccacgg aacccgaccc gacctaacgg accagccgct cccagacgcc gaccacacct 4740

ggtacacgga tggaagcagt ctcttacaag agggacagcg taaggcggga gctgcggtga 4800

ccaccgagac cgaggtaatc tgggctaaag ccctgccagc cgggacatcc gctcagcggg 4860

ctgaactgat agcactcacc caggccctaa agatggcaga aggtaagaag ctaaatgttt 4920

atactgatag ccgttatgct tttgctactg cccatatcca tggagaaata tacagaaggc 4980

gtgggttgct cacatcagaa ggcaaagaga tcaaaaataa agacgagatc ttggccctac 5040

taaaagccct ctttctgccc aaaagactta gcataatcca ttgtccagga catcaaaagg 5100

gacacagcgc cgaggctaga ggcaaccgga tggctgacca agcggcccga aaggcagcca 5160

tcacagagac tccagacacc tctaccctcc tcatagaaaa ttcatcaccc tacacctcag 5220

aacattttca ttacacagtg actgatataa aggacctaac caagttgggg gccatttatg 5280

ataaaacaaa gaagtattgg gtctaccaag gaaaacctgt gatgcctgac cagtttactt 5340

ttgaattatt agactttctt catcagctga ctcacctcag cttctcaaaa atgaaggctc 5400

tcctagagag aagccacagt ccctactaca tgctgaaccg ggatcgaaca ctcaaaaata 5460

tcactgagac ctgcaaagct tgtgcacaag tcaacgccag caagtctgcc gttaaacagg 5520

gaactagggt ccgcgggcat cggcccggca ctcattggga gatcgatttc accgagataa 5580

agcccggatt gtatggctat aaatatcttc tagtttttat agataccttt tctggctgga 5640

tagaagcctt cccaaccaag aaagaaaccg ccaaggtcgt aaccaagaag ctactagagg 5700

agatcttccc caggttcggc atgcctcagg tattgggaac tgacaatggg cctgccttcg 5760
tctccaaggt gagtcagaca gtggccgatc tgttggggat tgattggaaa ttacattgtg 5820
catacagacc ccaaagctca ggccaggtag aaagaatgaa tagaaccatc aaggagactt 5880
taactaaatt aacgcttgca actggctcta gagactgggt gctcctactc cccttagccc 5940
tgtaccgagc ccgcaacacg ccgggccccc atggcctcac cccatatgag atcttatatg 6000
gggcaccccc gccccttgta aacttccctg accctgacat gacaagagtt actaacagcc 6060
cctctctcca agctcactta caggctctct acttagtcca gcacgaagtc tggagacctc 6120

tggcggcagc ctaccaagaa caactggacc gaccggtggt acctcaccct taccgagtcg 6180
gcgacacagt gtgggtccgc cgacaccaga ctaagaacct agaacctcgc tggaaaggac 6240
cttacacagt cctgctgacc acccccaccg ccctcaaagt agacggcatc gcagcttgga 6300
tacacgccgc ccacgtgaag gctgccgacc ccgggggtgg accatcctct agactgacat 6360
ggcgcgttca acgctctcaa aacccccctca agataagatt aacccgtgga agcccttaat 6420
agtcatggga gtcctgttag gagtagggat ggcagagagc ccccatcagg tctttaatgt 6480
aacctggaga gtcaccaacc tgatgactgg gcgtaccgcc aatgccacct ccctcctggg 6540

aactgtacaa gatgccttcc caaaattata ttttgatcta tgtgatctgg tcggagagga 6600
gtgggaccct tcagaccagg aaccgtatgt cgggtatggc tgcaagtacc ccgcagggag 6660
acagcggacc cggacttttg acttttacgt gtgccctggg cataccgtaa agtcggggtg 6720
tgggggacca ggagagggct actgtggtaa atgggggtgt gaaaccaccg gacaggctta 6780
ctggaagccc acatcatcgt gggacctaat ctcccttaag cgcggtaaca ccccctggga 6840
cacgggatgc tctaaagttg cctgtggccc ctgctacgac ctctccaaag tatccaattc 6900
cttccaaggg gctactcgag ggggcagatg caaccctcta gtcctagaat tcactgatgc 6960

aggaaaaaag gctaactggg acgggcccaa atcgtgggga ctgagactgt accggacagg 7020
aacagatcct attaccatgt tctccctgac ccggcaggtc cttaatgtgg gaccccgagt 7080
ccccataggg cccaacccag tattacccga ccaaagactc ccttcctcac caatagagat 7140
tgtaccggct ccacagccac ctagcccccct caataccagt taccccccctt ccactaccag 7200
tacaccctca acctccccta caagtccaag tgtcccacag ccacccccag gaactggaga 7260
tagactacta gctctagtca aaggagccta tcaggcgctt aacctcacca atcccgacaa 7320
gacccaagaa tgttggctgt gcttagtgtc gggacctcct tattacgaag gagtagcggt 7380

cgtgggcact tataccaatc attccaccgc tccggccaac tgtacggcca cttcccaaca 7440

taagcttacc ctatctgaag tgacaggaca gggcctatgc atgggggcag tacctaaaac 7500

tcaccaggcc ttatgtaaca ccacccaaag cgccggctca ggatcctact accttgcagc 7560

acccgccgga acaatgtggg cttgcagcac tggattgact ccctgcttgt ccaccacggt 7620

gctcaatcta accacagatt attgtgtatt agttgaactc tggcccagag taatttacca 7680

ctcccccgat tatatgtatg gtcagcttga acagcgtacc aaatataaaa gagagccagt 7740

atcattgacc ctggcccttc tactaggagg attaaccatg ggagggattg cagctggaat 7800

agggacgggg accactgcct taattaaaac ccagcagttt gagcagcttc atgccgctat 7860

ccagacagac ctcaacgaag tcgaaaagtc aattaccaac ctagaaaagt cactgacctc 7920

gttgtctgaa gtagtcctac agaaccgcag aggcctagat ttgctattcc taaaggaggg 7980

aggtctctgc gcagccctaa aagaagaatg ttgtttttat gcagaccaca cggggctagt 8040

gagagacagc atggccaaat taagagaaag gcttaatcag agacaaaaac tatttgagac 8100

aggccaagga tggttcgaag ggctgtttaa tagatccccc tggtttacca ccttaatctc 8160

caccatcatg ggacctctaa tagtactctt actgatctta ctctttggac cttgcattct 8220

caatcgattg gtccaatttg ttaaagacag gatctcagtg gtccaggctc tggttttgac 8280

tcagcaatat caccagctaa aacccataga gtacgagcca tgaacgcgtt actggccgaa 8340

gccgcttgga ataaggccgg tgtgcgtttg tctatatgtt attttccacc atattgccgt 8400

cttttggcaa tgtgagggcc cggaaacctg gccctgtctt cttgacgagc attcctaggg 8460

gtctttcccc tctcgccaaa ggaatgcaag gtctgttgaa tgtcgtgaag gaagcagttc 8520

ctctggaagc ttcttgaaga caaacaacgt ctgtagcgac cctttgcagg cagcggaacc 8580

ccccacctgg cgacaggtgc ctctgcggcc aaaagccacg tgtataagat acacctgcaa 8640

aggcggcaca accccagtgc cacgttgtga gttggatagt tgtggaaaga gtcaaatggc 8700

tctcctcaag cgtattcaac aaggggctga aggatgccca gaaggtaccc cattgtatgg 8760

gatctgatct ggggcctcgg tgcacatgct ttacatgtgt ttagtcgagg ttaaaaaaac 8820

gtctaggccc cccgaaccac ggggacgtgg ttttcctttg aaaaacacga ttataaatgg 8880

tgaccggcgg catggcctcc aagtgggatc aaaagggcat ggatatcgct tacgaggagg 8940

ccctgctggg ctacaaggag ggcggcgtgc ctatcggcgg ctgtctgatc aacaacaagg 9000

acggcagtgt gctgggcagg ggccacaaca tgaggttcca gaagggctcc gccaccctgc 9060

acggcgagat ctccaccctg gagaactgtg gcaggctgga gggcaaggtg tacaaggaca 9120

ccaccctgta caccaccctg tccccttgtg acatgtgtac cggcgctatc atcatgtacg 9180

gcatccctag gtgtgtgatc ggcgagaacg tgaacttcaa gtccaagggc gagaagtacc 9240

tgcaaaccag gggccacgag gtggtggttg ttgacgatga gaggtgtaag aagctgatga 9300

agcagttcat cgacgagagg cctcaggact ggttcgagga tatcggcgag taagcggccg 9360

cagataaaat aaaagatttt atttagtctc cagaaaaagg ggggaatgaa agaccccacc 9420

tgtaggtttg gcaagctagc ttaagtaacg ccattttgca aggcatggaa aaatacataa 9480

ctgagaatag agaagttcag atcaaggtca ggaacagatg gaacagctga atatgggcca 9540

aacaggatat ctgtggtaag cagttcctgc cccggctcag ggccaagaac agatggaaca 9600

gctgaatatg ggccaaacag gatatctgtg gtaagcagtt cctgccccgg ctcagggcca 9660

agaacagatg gtccccagat gcggtccagc cctcagcagt ttctagagaa ccatcagatg 9720

tttccagggt gccccaagga cctgaaatga ccctgtgcct tatttgaact aaccaatcag 9780

ttcgcttctc gcttctgttc gcgcgcttct gctccccgag ctcaataaaa gagcccacaa 9840

cccctcactc ggggcgccag tcctccgatt gactgagtcg cccgggtacc cgtgtatcca 9900

ataaaccctc ttgcagttgc atccgacttg tggtctcgct gttccttggg agggtctcct 9960

ctgagtgatt gactacccgt cagcgggggt ctttcattac atgtgagcaa aaggccagca 10020

aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc tccgcccccc 10080

tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga caggactata 10140

aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc 10200

gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt ctcatagctc 10260

acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga 10320

acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc 10380

ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag 10440

gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct acactagaag 10500

gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gagttggtag 10560

ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt gcaagcagca 10620

gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga 10680

cgctcagtgg aacgaaaact cacgttaagg gattttggtc atgagattat caaaaaggat 10740

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 10800

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 10860

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 10920

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 10980

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 11040

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 11100

agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt cacgctcgtc 11160

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 11220
catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 11280
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 11340
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 11400
tatgcggcga ccgagttgct cttgcccggc gtcaacacgg gataataccg cgccacatag 11460
cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 11520
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 11580

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 11640
aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 11700
ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa 11760
aaataaacaa atagggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga 11820
aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtct 11880
tcaagaattc cat 11893

<210> 23
<211> 5
<212> PRT
<213> Artificial Sequence

<220><223> Peptide linker
<400> 23
Gly Gly Gly Gly Ser
1 5

<210> 24
<211> 5
<212> PRT
<213> Artificial Sequence
<220><223> Peptide linker
<400> 24
Ser Gly Gly Gly Gly
1 5

<210> 25
<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Peptide linker

<400> 25

Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Ser

1 5 10

<210> 26

<211> 14

<212> PRT

<213> Artificial Sequence

<220

><223> Peptide linker

<400> 26

Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly

1 5 10

<210> 27

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Peptide linker

<400> 27

Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Ser Gly Ser Thr

1 5 10 15

Lys Gly

<210> 28

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Peptide linker

<400

> 28

Gly Ser Thr Ser Gly Ser Gly Lys Pro Gly Ser Gly Glu Gly Ser Thr

1 5 10 15

Lys Gly

<210> 29

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Peptide linker

<400> 29

Glu Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Glu Phe

1 5 10

<210> 30

<211> 546

<212> DNA

<213> Artificial Sequence

<220><223> Cytosine deaminase coding sequence with cloning sites

<400> 30

aacacgatta taaatggtga cgaggggaat ggcaagcaag tgggatcaga agggtatgga 60

cattgcctat gaggaggcgg ccttaggtta caaagagggt ggtgttccta ttggcggatg 120

tcttatcaat aacaaagacg gaagtgttct cggtcgtggt cacaacatga gatttcaaaa 180

gggatccgcc acactacatg gtgagatctc cactttggaa aactgtggga gattagaggg 240

caaagtgtac aaagatacca ctttgtatac gacgctgtct ccatgcgaca tgtgtacagg 300

tgccatcatc atgtatggta ttccacgctg tgttgtcggt gagaacgtta atttcaaaag 360

taagggcgag aaatatttac aaactagagg tcacgaggtt gttgttgttg acgatgagag 420

gtgtaaaaag atcatgaaac aatttatcga tgaaagacct caggattggt ttgaagatat 480

tggtgagtag gcggccgcgc catagataaa ataaaagatt ttatttagtc tccagaaaaa 540

gggggg 546

<210> 31

<211> 537

<212> DNA

<213> Artificial Sequence

<220><223> Cytosine deaminase codon optimized with cloning site

<400> 31

ttataaatgg tgaccggcgg catggcctcc aagtgggatc aaaagggcat ggatatcgct 60

tacgaggagg ccgccctggg ctacaaggag ggcggcgtgc ctatcggcgg ctgtctgatc 120

aacaacaagg acggcagtgt gctgggcagg ggccacaaca tgaggttcca gaagggctcc 180

gccaccctgc acggcgagat ctccaccctg gagaactgtg gcaggctgga gggcaaggtg 240

tacaaggaca ccaccctgta caccaccctg tccccttgtg acatgtgtac cggcgctatc 300

atcatgtacg gcatccctag gtgtgtggtg ggcgagaacg tgaacttcaa gtccaagggc 360

gagaagtacc tgcaaaccag gggccacgag gtggtggttg ttgacgatga gaggtgtaag 420

aagatcatga agcagttcat cgacgagagg cctcaggact ggttcgagga tatcggcgag 480

tgataagcgg ccgcagataa aataaaagat tttatttagt ctccagaaaa aggggggg 537

<210> 32

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide Primer

<400> 32

tcgaggatat cggcgagtga aacccgttat tctttttggc 40

<210> 33

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 33

gccaaaaaga ataacgggtt tcactcgccg atatcctcga 40

<210> 34

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide Primer

<400> 34

tcggcgagtg atccggcggc ggcgcctccg gcggcggcgc ctccggcggc ggcgcctccg 60

gcggcggcgc caacccgtta tt 82

<210> 35

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide Primer

<400> 35

aataacgggt tggcgccgcc gccggaggcg ccgccgccgg aggcgccgcc gccggaggcg 60

ccgccgccgg atcactcgcc ga 82

<210> 36

<211> 533

<212> DNA

<213> Artificial Sequence

<220><223> Cytosine deaminase heat stabilized

<400> 36

atggtgacag ggggaatggc aagcaagtgg gatcagaagg gtatggacat tgcctatgag 60

gaggcgttat taggttacaa agagggtggt gttcctattg gcggatgtct tatcaataac 120

aaagacggaa gtgttctcgg tcgtggtcac aacatgagat ttcaaaaggg atccgccaca 180

ctacatggtg agatctccac tttggaaaac tgtgggagat tagagggcaa agtgtacaaa 240

gataccactt tgtatacgac gctgtctcca tgcgacatgt gtacaggtgc catcatcatg 300

tatggtattc cacgctgtgt catcggtgag aacgttaatt tcaaaagtaa gggcgagaaa 360

tatttacaaa ctagaggtca cgaggttgtt gttgttgacg atgagaggtg taaaaagtta 420

atgaaacaat ttatcgatga aagacctcag gattggtttg aagatattgg tgagtaggcg 480

gccgcgccat agataaaata aaagatttta tttagtctcc agaaaaaggg ggg 533

<210> 37

<211> 480

<212> DNA

<213> Artificial Sequence

<220><223> CD human codon optimized heat stabilized

<400> 37

atggtgaccg gcggcatggc ctccaagtgg gatcaaaagg gcatggatat cgcttacgag 60

gaggccctgc tgggctacaa ggagggcggc gtgcctatcg gcggctgtct gatcaacaac 120

aaggacggca gtgtgctggg caggggccac aacatgaggt tccagaaggg ctccgccacc 180

ctgcacggcg agatctccac cctggagaac tgtggcaggc tggagggcaa ggtgtacaag 240

gacaccaccc tgtacaccac cctgtcccct tgtgacatgt gtaccggcgc tatcatcatg 300

tacggcatcc ctaggtgtgt gatcggcgag aacgtgaact tcaagtccaa gggcgagaag 360

tacctgcaaa ccaggggcca cgaggtggtg gttgttgacg atgagaggtg taagaagctg 420

atgaagcagt tcatcgacga gaggcctcag gactggttcg aggatatcgg cgagtgataa 480

<210> 38

<211> 546

<212> DNA

<213> Artificial Sequence

<220><223> CD human codon optimized heat stabilitized cloning sites

<400> 38

aacacgatta taaatggtga ccggcggcat ggcctccaag tgggatcaaa agggcatgga 60

tatcgcttac gaggaggccc tgctgggcta caaggagggc ggcgtgccta tcggcggctg 120

tctgatcaac aacaaggacg gcagtgtgct gggcaggggc cacaacatga ggttccagaa 180

gggctccgcc accctgcacg gcgagatctc caccctggag aactgtggca ggctggaggg 240

caaggtgtac aaggacacca ccctgtacac caccctgtcc ccttgtgaca tgtgtaccgg 300

cgctatcatc atgtacggca tccctaggtg tgtgatcggc gagaacgtga acttcaagtc 360

caagggcgag aagtacctgc aaaccagggg ccacgaggtg gtggttgttg acgatgagag 420

gtgtaagaag ctgatgaagc agttcatcga cgagaggcct caggactggt tcgaggatat 480

cggcgagtaa gcggccgcgc catagataaa ataaaagatt ttatttagtc tccagaaaaa 540

gggggg 546

<210> 39

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 39

tcgaggatat cggcgagtga aacccgttat tctttttggc 40

<210> 40

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 40

gccaaaaaga ataacgggtt tcactcgccg atatcctcga 40

<210> 41

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 41

ccaagctcct atagccgctc actatctact tgggcaataa gaaaaaccga ag 52

<210> 42

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide Primer

<400> 42

ggttcgagga tatcggcgag tgatagatga acccgttatt ctttttggct tc 52

<210> 43

<211> 1296

<212> DNA

<213> Artificial Sequence

<220><223> CD codon optimized-UPRT

<400> 43

aacacgatta taaatggtga ccggcggcat ggcctccaag tgggatcaaa agggcatgga 60

tatcgcttac gaggaggccc tgctgggcta caaggagggc ggcgtgccta tcggcggctg 120

tctgatcaac aacaaggacg gcagtgtgct gggcaggggc cacaacatga ggttccagaa 180

gggctccgcc accctgcacg gcgagatctc caccctggag aactgtggca ggctggaggg 240

caaggtgtac aaggacacca ccctgtacac caccctgtcc ccttgtgaca tgtgtaccgg 300

cgctatcatc atgtacggca tccctaggtg tgtgatcggc gagaacgtga acttcaagtc 360

caagggcgag aagtacctgc aaaccagggg ccacgaggtg gtggttgttg acgatgagag 420

gtgtaagaag ctgatgaagc agttcatcga cgagaggcct caggactggt tcgaggatat 480

cggcgagaac ccgttattct ttttggcttc tccattcttg taccttacat atcttatata 540

ttatccaaac aaagggtctt tcgttagcaa acctagaaat ctgcaaaaaa tgtcttcgga 600

accatttaag aacgtctact tgctacctca aacaaaccaa ttgctgggtt tgtacaccat 660

catcagaaat aagaatacaa ctagacctga tttcattttc tactccgata gaatcatcag 720

attgttggtt gaagaaggtt tgaaccatct acctgtgcaa aagcaaattg tggaaactga 780

caccaacgaa aacttcgaag gtgtctcatt catgggtaaa atctgtggtg tttccattgt 840

cagagctggt gaatcgatgg agcaaggatt aagagactgt tgtaggtctg tgcgtatcgg 900

taaaatttta attcaaaggg acgaggagac tgctttacca aagttattct acgaaaaatt 960

accagaggat atatctgaaa ggtatgtctt cctattagac ccaatgctgg ccaccggtgg 1020

tagtgctatc atggctacag aagtcttgat taagagaggt gttaagccag agagaattta 1080

cttcttaaac ctaatctgta gtaaggaagg gattgaaaaa taccatgccg ccttcccaga 1140

ggtcagaatt gttactggtg ccctcgacag aggtctagat gaaaacaagt atctagttcc 1200

agggttgggt gactttggtg acagatacta ctgtgtttaa gcggccgcgc catagataaa 1260

ataaaagatt ttatttagtc tccagaaaaa gggggg 1296

<210> 44

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 44

tcggcgagtg atccggcggc ggcgcctccg gcggcggcgc ctccggcggc ggcgcctccg 60

gcggcggcgc caacccgtta tt 82

<210> 45

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 45

aataacgggt tggcgccgcc gccggaggcg ccgccgccgg aggcgccgcc gccggaggcg 60

ccgccgccgg atcactcgcc ga 82

<210> 46

<211> 1356

<212> DNA

<213> Artificial Sequence

<220><223> CD human codon optimized-linker-UPRT

<400> 46

aacacgatta taaatggtga ccggcggcat ggcctccaag tgggatcaaa agggcatgga 60

tatcgcttac gaggaggccc tgctgggcta caaggagggc ggcgtgccta tcggcggctg 120

tctgatcaac aacaaggacg gcagtgtgct gggcaggggc cacaacatga ggttccagaa 180

gggctccgcc accctgcacg gcgagatctc caccctggag aactgtggca ggctggaggg 240

caaggtgtac aaggacacca ccctgtacac caccctgtcc ccttgtgaca tgtgtaccgg 300

cgctatcatc atgtacggca tccctaggtg tgtgatcggc gagaacgtga acttcaagtc 360

caagggcgag aagtacctgc aaaccagggg ccacgaggtg gtggttgttg acgatgagag 420

gtgtaagaag ctgatgaagc agttcatcga cgagaggcct caggactggt tcgaggatat 480

cggcgagtcc ggcggcggcg cctccggcgg cggcgcctcc ggcggcggcg cctccggcgg 540

cggcgccaac ccgttattct ttttggcttc tccattcttg taccttacat atcttatata 600

ttatccaaac aaagggtctt tcgttagcaa acctagaaat ctgcaaaaaa tgtcttcgga 660

accatttaag aacgtctact tgctacctca aacaaaccaa ttgctgggtt tgtacaccat 720

catcagaaat aagaatacaa ctagacctga tttcattttc tactccgata gaatcatcag 780

attgttggtt gaagaaggtt tgaaccatct acctgtgcaa aagcaaattg tggaaactga 840

caccaacgaa aacttcgaag gtgtctcatt catgggtaaa atctgtggtg tttccattgt 900

cagagctggt gaatcgatgg agcaaggatt aagagactgt tgtaggtctg tgcgtatcgg 960

taaaatttta attcaaaggg acgaggagac tgctttacca aagttattct acgaaaaatt 1020

accagaggat atatctgaaa ggtatgtctt cctattagac ccaatgctgg ccaccggtgg 1080

tagtgctatc atggctacag aagtcttgat taagagaggt gttaagccag agagaattta 1140

cttcttaaac ctaatctgta gtaaggaagg gattgaaaaa taccatgccg ccttcccaga 1200

ggtcagaatt gttactggtg ccctcgacag aggtctagat gaaaacaagt atctagttcc 1260

agggttgggt gactttggtg acagatacta ctgtgtttaa gcggccgcgc catagataaa 1320

ataaaagatt ttatttagtc tccagaaaaa gggggg 1356

<210> 47

<211> 1269

<212> DNA

<213> Artificial Sequence

<220><223> CD human codon optimized-OPRT

<400> 47

aacacgatta taaatggtga ccggcggcat ggcctccaag tgggatcaaa agggcatgga 60

tatcgcttac gaggaggccc tgctgggcta caaggagggc ggcgtgccta tcggcggctg 120

tctgatcaac aacaaggacg gcagtgtgct gggcaggggc cacaacatga ggttccagaa 180
gggctccgcc accctgcacg gcgagatctc caccctggag aactgtggca ggctggaggg 240
caaggtgtac aaggacacca ccctgtacac caccctgtcc ccttgtgaca tgtgtaccgg 300
cgctatcatc atgtacggca tccctaggtg tgtgatcggc gagaacgtga acttcaagtc 360
caagggcgag aagtacctgc aaaccagggg ccacgaggtg gtggttgttg acgatgagag 420
gtgtaagaag ctgatgaagc agttcatcga cgagaggcct caggactggt tcgaggatat 480
cggcgaggcg gtcgctcgtg cagctttggg gccattggtg acgggtctgt acgacgtgca 540

ggctttcaag tttggggact tcgtgctgaa gagcgggctt tcctccccca tctacatcga 600
tctgcggggc atcgtgtctc gaccgcgtct tctgagtcag gttgcagata ttttattcca 660
aactgcccaa aatgcaggca tcagttttga caccgtgtgt ggagtgcctt atacagcttt 720
gccattggct acagttatct gttcaaccaa tcaaattcca atgcttatta gaaggaaaga 780
aacaaaggat tatggaacta agcgtcttgt agaaggaact attaatccag gagaaacctg 840
tttaatcatt gaagatgttg tcaccagtgg atctagtgtt ttggaaactg ttgaggttct 900
tcagaaggag ggcttgaagg tcactgatgc catagtgctg ttggacagag agcagggagg 960

caaggacaag ttgcaggcgc acgggatccg cctccactca gtgtgtacat tgtccaaaat 1020
gctggagatt ctcgagcagc agaaaaaagt tgatgctgag acagttggga gagtgaagag 1080
gtttattcag gagaatgtct ttgtggcagc gaatcataat ggttctcccc tttctataaa 1140
ggaagcaccc aaagaactca gcttcggtgc acgtgcagag ctgcccagga tccacccagt 1200
tgcatcgaag taagcggccg cgccatagat aaaataaaag attttattta gtctccagaa 1260
aaaggggggg 1269

<210> 48
<211> 1329
<212> DNA
<213> Artificial Sequence

<220><223> CD codon optimized-linker-OPRT

<400> 48

aacacgatta taaatggtga ccggcggcat ggcctccaag tgggatcaaa agggcatgga 60
tatcgcttac gaggaggccc tgctgggcta caaggagggc ggcgtgccta tcggcggctg 120
tctgatcaac aacaaggacg gcagtgtgct gggcaggggc cacaacatga ggttccagaa 180
gggctccgcc accctgcacg gcgagatctc caccctggag aactgtggca ggctggaggg 240

caaggtgtac aaggacacca ccctgtacac caccctgtcc ccttgtgaca tgtgtaccgg 300

cgctatcatc atgtacggca tccctaggtg tgtgatcggc gagaacgtga acttcaagtc 360

caagggcgag aagtacctgc aaaccagggg ccacgaggtg gtggttgttg acgatgagag 420

gtgtaagaag ctgatgaagc agttcatcga cgagaggcct caggactggt tcgaggatat 480

cggcgagtcc ggcggcggcg cctccggcgg cggcgcctcc ggcggcggcg cctccggcgg 540

cggcgccgcg gtcgctcgtg cagctttggg gccattggtg acgggtctgt acgacgtgca 600

ggctttcaag tttggggact tcgtgctgaa gagcgggctt tcctccccca tctacatcga 660

tctgcggggc atcgtgtctc gaccgcgtct tctgagtcag gttgcagata ttttattcca 720

aactgcccaa aatgcaggca tcagttttga caccgtgtgt ggagtgcctt atacagcttt 780

gccattggct acagttatct gttcaaccaa tcaaattcca atgcttatta gaaggaaaga 840

aacaaaggat tatggaacta agcgtcttgt agaaggaact attaatccag gagaaacctg 900

tttaatcatt gaagatgttg tcaccagtgg atctagtgtt ttggaaactg ttgaggttct 960

tcagaaggag ggcttgaagg tcactgatgc catagtgctg ttggacagag agcagggagg 1020

caaggacaag ttgcaggcgc acgggatccg cctccactca gtgtgtacat tgtccaaaat 1080

gctggagatt ctcgagcagc agaaaaaagt tgatgctgag acagttggga gagtgaagag 1140

gtttattcag gagaatgtct ttgtggcagc gaatcataat ggttctcccc tttctataaa 1200

ggaagcaccc aaagaactca gcttcggtgc acgtgcagag ctgcccagga tccacccagt 1260

tgcatcgaag taagcggccg cgccatagat aaaataaaag attttattta gtctccagaa 1320

aaaggggggg 1329

【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 1

【변경전】

e). 시토신 디아미나제

【변경후】

e). 시토신 디아미나아제