



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년07월17일  
(11) 등록번호 10-2556383  
(24) 등록일자 2023년07월12일

- (51) 국제특허분류(Int. Cl.)  
C12N 15/85 (2006.01) C07K 16/00 (2006.01)  
C12P 21/00 (2006.01)
- (52) CPC특허분류  
C12N 15/85 (2013.01)  
C07K 16/00 (2013.01)
- (21) 출원번호 10-2022-7018877(분할)
- (22) 출원일자(국제) 2015년08월19일  
심사청구일자 2022년07월01일
- (85) 번역문제출일자 2022년06월03일
- (65) 공개번호 10-2022-0080215
- (43) 공개일자 2022년06월14일
- (62) 원출원 특허 10-2017-7006620  
원출원일자(국제) 2015년08월19일  
심사청구일자 2020년08월12일
- (86) 국제출원번호 PCT/US2015/045793
- (87) 국제공개번호 WO 2016/028838  
국제공개일자 2016년02월25일
- (30) 우선권주장  
62/039,416 2014년08월19일 미국(US)
- (56) 선행기술조사문헌  
KR1020090053230 A\*  
JP2005027599 A  
JBC. ,265(33):20621-20626(1990.11.)\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
리제너론 파마슈티칼스 인코포레이티드  
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리버 로드 777
- (72) 발명자  
데슈판데 디팔리  
미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내  
부라코프 다리아  
미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내  
(뒷면에 계속)
- (74) 대리인  
특허법인와이에스장

전체 청구항 수 : 총 23 항

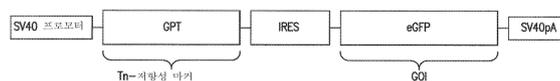
심사관 : 최성호

(54) 발명의 명칭 **재조합 단백질의 효율적인 선택성**

(57) 요약

본 발명은 당단백질의 바람직한 번역후 변형을 촉진하는 포유류 선택성 마커를 포함하는 새로운 발현 체계를 제공한다. 특히, 본 발명은 포유류 기원의 GPT 유전자에 기초한 선별 마커 시스템을 사용함으로써 포유류 세포에서 최적의 재조합 단백질 발현을 위한 방법 및 조성물을 포함한다. 본 발명은 포유류 세포에서 재조합 단백질의 선택성 및 향상된 발현 사본 뿐만 아니라 단백질 수율을 용이하게 하는 방법, 및 GPT 발현 체계를 사용하는 방법을 포함한다.

대표도



(52) CPC특허분류

*C12P 21/005* (2013.01)

*C07K 2319/30* (2013.01)

*C12N 2510/02* (2013.01)

*C12N 2840/203* (2013.01)

(72) 발명자

**첸 강**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

---

**판들 제임스**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**명세서**

**청구범위**

**청구항 1**

포유류 세포 배양시 선택 마커로서 튜니카마이신(Tn)을 사용하는 방법으로서,

(a) 포유류 숙주 세포 집단을 제공하는 단계,

(b) 형질감염에 의해 핵산을 단계 (a)의 세포 집단으로 도입하는 단계로서, 핵산은 (i) 서열번호 3의 아미노산 서열과 적어도 93% 동일성을 갖는 단백질을 암호화하는 포유류 튜니카마이신(Tn)-저항성 유전자, 및 (ii) 제1 관심 단백질(POI)을 암호화하는 제1 관심 유전자(GOI)를 포함하는, 단계; 및

(c) Tn의 존재 시에 단계 (b)의 세포 집단을 배양하여, 상기 핵산을 포함하는 세포 형질감염체를 포함하는 세포 집단을 얻는 단계

를 포함하는, 방법.

**청구항 2**

제1 항에 있어서, 단계 (c)에서 얻은 세포 집단으로부터 세포 형질감염체를 선택하는 단계를 더 포함하는 것을 특징으로 하는 방법.

**청구항 3**

제1 항에 있어서, 단계 (c)로부터 얻은 세포 집단 또는 선택된 세포 형질감염체를 배양하는 단계, 상기 제1 GOI로부터 상기 제1 POI를 발현시키는 단계, 및 세포 배양물로부터 상기 제1 POI를 단리시키는 단계를 더 포함하는 것을 특징으로 하는 방법.

**청구항 4**

제1 항에 있어서, 핵산이 숙주 세포 집단의 게놈 내 표적 유전자자리에 통합되는 것을 특징으로 하는 방법.

**청구항 5**

제4 항에 있어서, 상기 표적 유전자자리에서의 통합이 표적 유전자자리 내 특정 부위를 인식하는 재조합효소에 의해 매개되거나; 또는 상기 통합이 징크 핑거 뉴클레아제(ZFN), 전사 촉진제-유사 이펙터 뉴클레아제(TALEN), 또는 RNA 유전자가위(RNA-guided endonuclease)와 같이 표적 유전자 내 DNA 서열을 인식하여 절단하는 뉴클레아제에 의해 선택적으로 촉진되는 상동 재조합에 의해 매개되는 것을 특징으로 하는 방법.

**청구항 6**

재조합 관심 단백질(POI)을 생산하는 방법으로서,

(a) 외생적 핵산을 포함하는 포유류 숙주 세포를 제공하는 단계로서, 여기서 외생적 핵산은 (i) 서열번호 3의 아미노산 서열과 적어도 93% 동일성을 갖는 단백질을 암호화하는 포유류 튜니카마이신(Tn)-저항성 유전자, 및 (ii) 제1 POI를 암호화하는 제1 관심 유전자(GOI)를 포함하는, 단계;

(b) 상기 제1 POI를 발현시키기 위해 Tn의 존재 시에 세포를 배양하는 단계

를 포함하는, 방법.

**청구항 7**

제1 항 내지 제6 항 중 어느 한 항에 있어서, Tn은 1 µg/mL, 2.5 µg/mL, 또는 5 µg/mL의 농도로 존재하는 것을 특징으로 하는 방법.

**청구항 8**

제1 항 내지 제5 항 중 어느 한 항에 있어서, 상기 단계 (c)에서 세포 집단을 배양하는 단계는 순차적으로 증가하는 농도의 Tn의 존재 시에 배양하는 단계 또는 제6 항의 방법을 포함하고, 단계 (b)는 순차적으로 증가하는 농도의 Tn의 존재 시에 배양하는 단계를 포함하고, 선택적으로 단계 (b) 또는 (c)에서 순차적으로 증가하는 농도의 Tn의 존재 시에 배양하는 단계는 1 µg/mL의 제1 Tn 농도에서 배양하는 단계에 이어서 2.5 µg/mL 또는 5 µg/mL의 제2 Tn 농도에서 배양하는 단계를 포함하는 것을 특징으로 하는 방법.

**청구항 9**

제1 항 내지 제6 항 중 어느 한 항에 있어서,

(i) Tn-저항성 유전자는 서열번호 2, 서열번호 11, 서열번호 12, 서열번호 13, 서열번호 14, 서열번호 15, 서열번호 16 및 서열번호 17로 이루어진 군으로부터 선택된 핵산 서열을 포함하고;

(ii) 제1 GOI는 프로모터에 연결되고;

(iii) 제1 GOI는 당단백질을 암호화하고, 선택적으로 당단백질은 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, Fc-융합 단백질 또는 그의 단편, 리간드, 및 수용체 또는 그의 리간드-결합 단편으로부터 선택되고;

(iv) 포유류 숙주 세포는 CHO, COS-7, HEK293, 종양 세포, 림프구, 망막 세포, 및 줄기 세포로 이루어진 군으로부터 선택되고, 선택적으로 포유류 숙주 세포는 CHO 세포이고; 및/또는

(v) 핵산은 제2 관심 유전자(GOI)를 더 포함하고, 선택적으로 제2 GOI는 당단백질을 암호화하고 선택적으로 당단백질은 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, Fc-융합 단백질 또는 그의 단편, 리간드, 및 수용체 또는 그의 리간드-결합 단편으로부터 선택되는

것을 특징으로 하는 방법.

**청구항 10**

제1 항 내지 제6 항 중 어느 한 항에 있어서, GOI로부터 발현된 POI를 단리하는 단계를 더 포함하는 것을 특징으로 하는 방법.

**청구항 11**

(a) 서열번호 3의 아미노산 서열과 적어도 93% 동일성을 갖는 단백질을 암호화하는 포유류 튜니카마이신(Tn)-저항성 유전자; 및

(b) 제1 관심 유전자(GOI)

를 포함하는 외생적 핵산을 포함하는, 단리된 세포.

**청구항 12**

제11 항에 있어서, 제1 GOI는 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, 또는 Fc-융합 단백질 또는 그의 단편을 암호화하는 것을 특징으로 하는 단리된 세포.

**청구항 13**

제11 항에 있어서,

(i) Tn-저항성 유전자는 서열번호 2, 서열번호 11, 서열번호 12, 서열번호 13, 서열번호 14, 서열번호 15, 서열번호 16, 및 서열번호 17로 이루어진 군으로부터 선택된 핵산 서열을 포함하고;

(ii) 제1 GOI는 프로모터에 연결되고; 및/또는

(iii) 세포는 제2 외생적으로 부가된 관심 유전자(GOI)를 더 포함하는

것을 특징으로 하는 단리된 세포.

**청구항 14**

제13 항에 있어서, 제2 GOI는 당단백질을 암호화하는 것을 특징으로 하는 단리된 세포.

**청구항 15**

제14 항에 있어서, 당단백질은 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, Fc-융합 단백질 또는 그의 단편, 리간드, 및 수용체 또는 그의 리간드-결합 단편으로부터 선택되는 것을 특징으로 하는 단리된 세포.

**청구항 16**

(a) 서열번호 3의 아미노산 서열과 적어도 93% 동일성을 갖는 단백질을 암호화하는 포유류 튜니카마이신(Tn)-저항성 유전자; 및  
 (b) 제1 관심 유전자(GOI)  
 를 포함하는 핵산을 포함하는, 벡터.

**청구항 17**

제16 항에 있어서, 제1 GOI는 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, 또는 Fc-융합 단백질 또는 그의 단편을 암호화하는 것을 특징으로 하는 벡터.

**청구항 18**

제16 항에 있어서,  
 (i) Tn-저항성 유전자는 서열번호 2, 서열번호 11, 서열번호 12, 서열번호 13, 서열번호 14, 서열번호 15, 서열번호 16, 및 서열번호 17로 이루어진 군으로부터 선택된 핵산 서열을 포함하고;  
 (ii) 제1 GOI는 프로모터에 연결되고; 및/또는  
 (iii) 벡터는 제2 외생적으로 부가된 관심 유전자(GOI)를 더 포함하는 것을 특징으로 하는 벡터.

**청구항 19**

제18 항에 있어서, 제2 GOI는 당단백질을 암호화하는 것을 특징으로 하는 벡터.

**청구항 20**

제19 항에 있어서, 당단백질은 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, Fc-융합 단백질 또는 그의 단편, 리간드, 및 수용체 또는 그의 리간드-결합 단편으로부터 선택되는 것을 특징으로 하는 벡터.

**청구항 21**

제11 항에 있어서,  
 (i) 외생적 핵산은 세포의 게놈 내 표적 유전자자리에 통합되고; 및/또는  
 (ii) 세포는 CHO, COS-7, HEK293, 종양 세포, 림프구, 망막 세포, 및 줄기 세포로 이루어진 군으로부터 선택되고, 선택적으로 세포는 CHO-K1 세포인  
 것을 특징으로 하는 단리된 세포.

**청구항 22**

제12 항에 있어서,  
 (i) 외생적 핵산은 세포의 게놈 내 표적 유전자자리에 통합되고; 및/또는  
 (ii) 세포는 CHO, COS-7, HEK293, 종양 세포, 림프구, 망막 세포, 및 줄기 세포로 이루어진 군으로부터 선택되고, 선택적으로 세포는 CHO-K1 세포인  
 것을 특징으로 하는 단리된 세포.

**청구항 23**

제13 항에 있어서,

(i) 외생적 핵산은 세포의 게놈 내 표적 유전자자리에 통합되고; 및/또는

(ii) 세포는 CHO, COS-7, HEK293, 종양 세포, 림프구, 망막 세포, 및 줄기 세포로 이루어진 군으로부터 선택되고, 선택적으로 세포는 CHO-K1 세포인

것을 특징으로 하는 단리된 세포.

**청구항 24**

삭제

**청구항 25**

삭제

**청구항 26**

삭제

**청구항 27**

삭제

**청구항 28**

삭제

**청구항 29**

삭제

**청구항 30**

삭제

**청구항 31**

삭제

**청구항 32**

삭제

**청구항 33**

삭제

**청구항 34**

삭제

**청구항 35**

삭제

**청구항 36**

삭제

**청구항 37**

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

**발명의 설명**

**기술 분야**

[0001] **서열 목록**

[0002] 본 출원은 2015년 8월 3일에 작성된 8700WO\_ST25.txt 파일(75,769 바이트)로서 컴퓨터 판독가능 형식(Computer Readable Form)으로 제출된 서열 목록을 참조문서로 통합한다.

[0003] 본 발명은 일관되고 효율적인 방식으로 포유류 세포에서 재조합 단백질을 발현하는 것을 제공한다. 특히, 본 발명은 포유류 선택 마커를 사용함으로써 포유류 세포에서 단백질의 개선된 발현을 위한 방법 및 조성물을 포함한다. 본 발명은 포유류 세포에서 재조합 단백질의 선택성 및 향상된 발현 사본 뿐만 아니라 단백질 수율을 촉진하는 방법, 및 이러한 발현 체계를 사용하는 방법을 포함한다.

**배경 기술**

[0004] 세포 발현 체계의 개발은 연구 및 치료용으로 주어진 단백질의 신뢰성 있고 효율적인 공급원을 제공하기 위한 중요한 목표이다. 포유류 세포에서 재조합 단백질 발현은, 예를 들어, 재조합 단백질을 적절하게 번역후 변형하는 포유류 발현 체계의 능력 때문에 종종 치료용 단백질의 제조에 선호된다.

[0005] 다양한 벡터가 포유류 숙주에서의 발현에 이용될 수 있으며, 각 벡터는 세포 배양 중 재조합 단백질-발현 세포의 분리를 용이하게 할 수 있는 선택 마커를 포함한다. 선택성 마커 유전자(SMG)는 관심 단백질을 발현하는 세포에 선택적 우위를 부여하기 때문에 이러한 시스템에서 사용되나, SMG은 여러 이유중 이들의 표현형 중립성, 효율성 및 융통성에 있어 최적화되어야 한다.

[0006] SMG를 호스팅하는 수많은 벡터 및 발현 체계의 유효성에도 불구하고, 포유류 시스템에서 얻은 재조합 단백질의 발현은 종종 양적, 질적으로 또는 이들 모두를 만족시키지 못한다 분자의 생물학적 “지문(fingerprint)”, 예를 들어, 당질화와 같은 번역후 변형은 재조합 단백질 치료제의 개발에 있어 분자의 유용성과 효능을 정의하는데 있어 특히 중요하다(Cumming, D.A., 1990, *Glycobiology*, 1(2):115-130). 발현된 관심 단백질의 생물학적 특성에 부정적인 영향을 끼치지 않는 SMG가 특히 유익하다.

[0007] 대부분의 SMG는 세균 유래로 환경 세균에 대한 세균성 항생제 저항성 유전자의 수평 이동 위험에 대한 우려 증가 때문에 포유류 시스템에서 사용되는데 불리하다(Breyer, D. 등, 2014, *Critical Reviews in Plant Sciences* 33:286-330). 세균성 항생제 저항성 유전자의 사용 배제는 사용자 수용 및 이러한 인지된 위험을 완화하는데 궁

정적인 효과를 가질 수 있다.

[0008] 유전자-조작된 자가 세포는 빠르게 임상적 성공을 거두고 있다(예를 들어, Kershaw, M.H. 등, 2013, *Nature Reviews: Cancer* 13:525-541 참조). 인간 자가 세포 산물에서 유전적 변형을 위한 백터의 선정 및 제작이 중요한데, 특히, 인간 자가 세포로의 비-인간성 성분의 원치 않는 도입은 환자의 안전에 있어 심각한 결과를 초래할 수 있기 때문이다(Eaker, 등, 2013, *Stem cells Trans. Med.* 2:871-883; SCTMEXPRESS October 7, 2013에 최초로 온라인 출판). 세균 대신, 포유류 유래의 성분만을 가진 백터 시스템은 임상 면역요법을 위한 환자-특이적 T 세포에 사용하기에 유리할 것이다.

[0009] 따라서, 포유류 선택성 유전자, 특히, 상기 형질전환된 세포에 목적 포유류 단백질의 생산을 위한 발현 체계에 있어 표현형 또는 대사적 이점을 부여하는 유전자를 도입하는 것이 바람직하다. 게다가, 충분히 높은 양의 치료성 단백질을 신뢰성 있게 발현하고 단백질 번역 후 상기 치료성 단백질을 적절하고 지속적으로 변형하는 세포주가 매우 바람직하다. 따라서, 당해 기술 분야에서는 향상된 포유류 발현 체계가 요구된다.

**발명의 내용**

[0010] 포유류 발현 체계에서 선택성 마커로서 포유류 튜니카마이신(Tn) 저항성 유전자의 사용은 형질감염체의 효율 및 사본 수를 증가시킬 수 있다. 관심 유전자에 작동가능하게 연결된 Tn 저항성 유전자의 사용은 포유류 세포의 군집에 대한 선택압을 생성해 상기 형질감염체(즉, 관심 유전자)의 무작위 통합이 증가되는 것이 관찰되었다. 선택 마커 시스템은 목적하는 형질감염체의 선택을 촉진할 수 있으나, 본 발명의 방법은 상기 목적하는 단백질의 신뢰성 있는 생물학적 성질뿐 아니라 상기 관심 유전자의 효율 및 무작위 통합 모두를 예상치 못하게 증가시키는 것으로 이해된다. 따라서 본 발명의 조성물 및 방법은 발현 단백질에 대한 질적으로 바람직한 번역후 변형이라는 이로온 선택을 허용한다.

[0011] 일 측면에서, 본 발명은 관심 유전자(GOI) 및 적어도 하나의 조절 요소에 작동가능하게 연결된, 서열번호 3의 아미노산 서열과 적어도 93% 동일성을 가진 단백질을 암호화하는 포유류 튜니카마이신(Tn)-저항성 유전자를 포함하는 단리된 세포를 제공한다.

[0012] 다른 측면에서, 본 발명은 (i) 포유류 튜니카마이신(Tn)-저항성 유전자 및 (ii) 상기 POI를 암호화하는 유전자를 포함하는 핵산 분자를 암호화하는 포유류 숙주 세포를 제공하는 단계; 제1 농도의 Tn 존재 하에 상기 세포를 배양하는 단계; 상기 Tn-저항성 유전자의 적어도 하나의 사본을 발현하는 세포 군집을 단리시키는 단계; 및 증가하는 농도의 Tn의 존재 하에 세포 군집을 배양하는 단계로, 상기 Tn 농도의 증가는 상기 POI의 생산을 증가시키는, 단계; 및 상기 세포 배양으로부터 POI를 단리시키는 단계를 포함하는 제조법 관심 단백질(POI)을 생산하는 방법을 제공한다.

[0013] 또 다른 측면에서, 본 발명은 당질화가 요구되는 상기 단백질 기질을 암호화하는 유전자에 작동가능하게 연결된 포유류 튜니카마이신(Tn)-저항성 유전자를 포함하는 핵산 분자를 암호화하는 포유류 숙주 세포를 제공하는 단계; 제1 농도의 Tn 존재 하에 상기 세포를 배양하는 단계; 상기 Tn-저항성 유전자의 적어도 하나의 사본을 발현하는 세포 군집을 단리시키는 단계; 증가하는 농도의 Tn의 존재 하에 상기 세포 군집을 배양하는 단계로, 상기 Tn 농도의 증가는 상기 POI의 생산을 증가시키는 단계; 및 상기 세포 배양으로부터 상기 단백질 기질을 단리시키는 단계를 포함하는, N-글리칸 단백질 기질을 당질화하는 방법을 제공한다.

[0014] 상기 방법의 일부 실시예에서, 상기 Tn-저항성 유전자는 상기 POI를 암호화하는 유전자에 작동가능하게 연결되고, 상기 POI를 암호화하는 유전자는 적어도 하나의 조절 요소에 작동가능하게 연결된다.

[0015] 일부 실시예에서, Tn-저항성 유전자는 세포에 외생적으로 부가된 것이다. 다른 실시예들에서, Tn-저항성 유전자는 서열번호 3의 아미노산 서열과 적어도 93%의 동일성을 갖는 단백질을 암호화한다. 다른 실시예들에서, Tn-저항성 유전자는 서열번호 3의 아미노산 서열과 적어도 94%의 동일성을 갖는 단백질을 암호화한다. 일부 실시예에서, Tn-저항성 유전자는 서열번호 4의 아미노산 서열과 적어도 93%의 동일성을 갖는 단백질을 암호화한다. 또 다른 실시예들에서, Tn-저항성 유전자는 서열번호 4의 아미노산 서열과 적어도 94%의 동일성을 갖는 단백질을 암호화한다.

[0016] 일부 실시예에서, 포유류 Tn-저항성 유전자는 중국 햄스터 (*Cricetulus griseus*) Tn-저항성 유전자를 포함한다. 다른 실시예들에서, 포유류의 Tn-저항성 유전자는 인간 Tn-저항성 유전자를 포함한다.

[0017] Tn-저항성 유전자는 또한 서열번호 2, 서열번호 11, 서열번호 12, 서열번호 13, 서열번호 14, 서열번호 15, 서열번호 16 및 서열번호 17로 이루어진 군으로부터 선택된 핵산 서열을 포함할 수도 있다.

- [0018] 상기 발명의 소정의 실시예들에서 포유류 Tn-저항성 유전자는 서열번호 2의 핵산 서열과 적어도 92%의 동일성을 갖는 핵산 서열을 포함한다. 일부 실시예에서, 포유류 Tn-저항성 유전자는 서열번호 12의 핵산 서열과 적어도 92%의 동일성을 갖는 핵산 서열을 포함한다.
- [0019] Tn-저항성 유전자에 작동 가능하게 연결된 적어도 하나의 조절 요소가 본 발명의 단리된 세포에 제공되며, 여기서 조절 요소는 프로모터, 리보솜-결합 부위 및 인핸서를 포함하지만 이에 한정되지는 않는다. 또 다른 실시예에서, GOI는 프로모터에 작동 가능하게 연결된다. 또 다른 실시예에서, 예를 들어, GOI는 IRES와 같은, 리보솜-결합 부위에 작동 가능하게 연결된다.
- [0020] 일부 실시예에서, 본 발명의 단리된 세포 및 방법은 제2 관심 유전자(GOI)를 더 포함하는 반면, GOI는 관심 단백질(POI)을 암호화한다. 일 실시예에서, 관심 유전자(GOI)는 외생적으로 부가된 GOI이다. 또 다른 실시예에서, 외생적으로 부가된 GOI는 인간 유전자이다. 또 다른 실시예에서, 조절 요소는 외생적으로 부가된 조절 요소이다.
- [0021] 다른 실시예들에서, 제1 및/또는 제2 GOI는 항체 중쇄, 항체 경쇄, 항원-결합 단편 및/또는 Fc-융합 단백질을 포함하지만 이에 한정되지 않는 POI를 암호화한다.
- [0022] 또 다른 실시예에서, 제1 GOI 및 제2 GOI는 항체 경쇄 또는 그의 항원-특이적 단편, 항체 중쇄 또는 그의 항원-특이적 단편, Fc-융합 단백질 또는 그의 단편, 및 수용체 또는 그의 리간드-특이적 단편을 암호화하는 유전자로 이루어진 군으로부터 독립적으로 선택된다. 일 실시예에서, 재조합효소 인식 부위는 제1 GOI와 제2 GOI 사이에 존재한다. 다른 실시예들에서, 본 발명은 제1 GOI에 대한 재조합효소 인식 부위 5' 및 제2 GOI에 대한 재조합효소 인식 부위 3'를 더 제공한다.
- [0023] 또 다른 실시예에서, GOI는 항체 경쇄 또는 그의 항원-결합 단편, 항체 중쇄 또는 그의 항원-결합 단편, Fc-융합 단백질 또는 그의 단편, 리간드, 및 수용체 또는 그의 리간드-결합 단편으로부터 선택된 당단백질을 암호화한다.
- [0024] 본 발명의 단리된 비-자연발생 세포는 진핵 세포로부터 유래될 수도 있다. 일 실시예에서, 세포는 포유류 세포이다. 일부 실시예에서, 단리된 세포는 생체 외 인간 세포이다. 다른 실시예들에서, 세포는 CHO(예를 들어, CHO K1, DXB-11 CHO, Veggie-CHO), COS(예를 들어, COS-7), 림프구, 줄기 세포, 망막 세포, Vero, CV1, 신장(예를 들어, HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L 세포, C127 세포, SP2/0, NS-0, MMT 세포, 종양 세포, 및 전술된 세포로부터 유래한 세포주로 이루어진 군으로부터 선택된다. 소정의 실시예들에서, 본 발명의 단리된 세포는 CHO-K1 세포, 림프구, 망막 세포, 또는 줄기 세포이다.
- [0025] 일 실시예에서, Tn의 제1 농도는 1 µg/mL이다. 또 다른 실시예에서, 증가하는 Tn의 농도는 Tn의 제2 농도 및 제3 농도를 포함한다.
- [0026] 일부 실시예에서, 제2 농도는 Tn의 제1 농도보다 크고, 제3 농도는 Tn의 제2 농도보다 크다. 소정의 실시예들에서, Tn의 제2 농도는 2.5 µg/ml이고, 제3 농도는 5 µg/ml이다.
- [0027] 또 다른 실시예들에서, 증가하는 Tn의 농도는 Tn의 제2 농도를 포함하며, 여기서 Tn의 제2 농도는 2.5 µg/ml 또는 5 µg/ml이다.
- [0028] 본 발명의 임의의 측면 및 실시예는 달리 명시되거나 문맥으로부터 명백하지 않는 한, 본 발명의 임의의 다른 측면 또는 실시예와 관련하여 사용될 수 있다.
- [0029] 다른 목적 및 이점은 후속하는 상세한 설명의 검토로부터 명백해질 것이다.

**도면의 간단한 설명**

[0030] **도 1**은 관심 유전자, 예를 들어, eGFP를 암호화하는 상기 핵산 서열의 세포 계놈으로의 도입을 위해 사용되는 클로닝 벡터 구조물 내 유효 발현 카세트의 개략도를 나타낸다. SV40 프로모터: 시미언 바이러스(Simian virus) 40 프로모터; GPT: GlcNAc-1-P transferase(예를 들어, CHO-GPT, 서열번호 2; 또는 hGPT, 서열번호 12); IRES: 내부 리보솜 도입 부위(internal ribosomal entry site); eGFP: 향상된 녹색 형광 단백질(enhanced Green Fluorescent Protein); SV40polyA: 시미언 바이러스 40 폴리A.

**도 2a 내지 2c**는 중국 햄스터(GPT\_CRIGR; UniProtKB Accn. No. P24140; 서열번호 3) GPT 아미노산 서열과 비교

해 포유류 GPT 아미노산 서열, 다시 말해, 인간(GPT\_HUMAN; UniProtKB Accn. No. Q9H3H5; 서열번호 4), 히말라야 원숭이(Rhesus macaque; GPT\_MACMU; UniProtKB Accn. No. F6TXM3; 서열번호 5), 침팬지(GPT\_PANTR; UniProtKB Accn. No. H2R346; 서열번호 6), 개(GPT\_CANFA; UniProtKB Accn. No. E2RQ47; 서열번호 7), 기니아 피그(GPT\_CAVPO; UniProtKB Accn. No. E2RQ47; 서열번호 8), 랫트(GPT\_RAT; UniProtKB Accn. No. Q6P4Z8; 서열번호 9), 및 마우스(GPT\_MOUSE; UniProtKB Accn. No. P42867; 서열번호 10)의 배열을 나타낸다.

**도 3a 및 3b**는 본 발명의 방법 및 조성물을 사용하여 단백질을 최적화하는 방법을 예시한다. **도 3a**는 1 µg/mL 튜니카마이신(Tn)과 함께 배양된 첫번째 세포 풀(pool)로부터 양성 세포 형질감염체를 선택하는 방법을 나타낸다. 이어서, 단백질 발현을 향상시키기 위해 증가된 농도, 예를 들어, 2.5 µg/mL 또는 5 µg/mL의 튜니카마이신과 함께 두번째 세포를 배양한다. **도 3b**는 단백질 발현을 최적화하기 위해 1 µg/mL 튜니카마이신(Tn)과 함께 배양된 첫번째 세포 풀로부터 양성 세포 형질감염체를 선택한 다음, 이후의 세포 배양시 Tn 농도를 연속적으로 증가시키는 방법을 나타낸다.

**도 4a 및 4b**는 히그로마이신 선택성의 다양한 파라미터를 나타내는 FACS 산포도를 보여준다. 변형된 CHO 세포는 lox 자리 옆에 위치한 YFP 유전자를 포함한다. Lox 자리 옆에 위치한 선택 마커(항생제 저항성 유전자 및 eGFP)는 상기 YFP 자리에 포함되고 표적화 통합(targeted integration)을 통해 YFP를 Cre 재조합효소로 치환한다. 무작위 통합 성분은 YFP 및 eGFP 모두를 표현한다. **도 4a**는 세포가 Cre 재조합효소 벡터 및 eGFP로 구성된 *hpt* 발현 벡터로 형질감염되나, 배양시 히그로마이신 없이 배양된다는 것을 나타낸다. **도 4b**는 세포가 400 µg/mL 히그로마이신의 존재 하에 Cre 재조합효소 벡터 및 eGFP로 구성된 *hpt* 발현 벡터로 형질감염된다는 것을 나타낸다.

**도 5a 내지 5f**는 튜니카마이신(Tn) 선택성의 다양한 파라미터를 나타내는 FACS 산포도를 보여준다. 변형된 CHO 세포는 lox 자리 옆에 위치한 YFP 유전자를 포함한다. Lox 자리 옆에 위치한 선택 마커(항생제 저항성 유전자 및 eGFP)는 상기 YFP 자리에 포함되고 표적화 통합을 통해 YFP를 Cre 재조합효소로 치환한다. 무작위 통합 성분은 YFP 및 eGFP 모두를 표현한다. **도 5a**는 세포가 Cre 재조합효소 벡터 및 eGFP로 구성된 CHO-GPT 발현 벡터로 형질감염되나; 배양시 튜니카마이신 없이 배양된다는 것을 나타낸다. **도 5b**는 세포가 1 µg/mL Tn 존재 하에 Cre 재조합효소 벡터 및 eGFP로 구성된 CHO-GPT 발현 벡터로 형질감염된다는 것을 나타낸다. **도 5c**는 세포가 2.5 µg/mL Tn의 존재 하에 Cre 재조합효소 벡터 및 eGFP로 구성된 CHO-GPT 발현 벡터로 형질감염된다는 것을 나타낸다. **도 5d**는 세포가 Cre 재조합효소 벡터 및 eGFP로 구성된 인간 GPT 발현 벡터로 형질감염되나; 배양시 튜니카마이신 없이 배양된다는 것을 나타낸다. **도 5e**는 세포가 1 µg/mL Tn 존재 하에 Cre 재조합효소 벡터 및 eGFP로 구성된 인간 GPT 발현 벡터로 형질감염된다는 것을 나타낸다. **도 5f**는 세포가 2.5 µg/mL Tn의 존재 하에 Cre 재조합효소 벡터 및 eGFP로 구성된 인간 GPT 발현 벡터로 형질감염된다는 것을 나타낸다.

**도 6a 및 6b**는 eGFP와 같은 작동가능하게 연결된 GOI의 발현을 향상시키는 상대적 능력에 대해 세포 풀을 발현하는 비-GPT와 비교한 세포 풀을 발현하는 GPT를 나타낸다. **도 6a**는 세포 풀에 대한 PCR에 의해 측정된 CHO-GPT의 상대적 유전자 사본 수를 나타내며, 다음과 같다: Tn 선택없는 풀-49 세포(외인성 GPT 미첨가); 5 µg Tn 선택하는 풀-49 세포(외인성 GPT 없음); 풀-1 세포는 자연적으로 더 많은 양의 GPT 발현하고(데이터 미도시), Tn 선택없이 검사한다; Tn 선택없는 풀-78 세포(외인성 GPT 없음); 외생적으로 부가된 *hpt*를 발현하는 CHO 세포 및 400 µg/mL 히그로마이신 선택; 1 µg/mL Tn 선택 조건 하에 외인성 GPT를 발현하는 CHO 세포; 1 µg/mL Tn 선택 풀에서 선택된 외인성 GPT를 발현하는 CHO 세포는 1 µg/mL Tn에서 더 배양된다; 1 µg/mL Tn 선택 풀에서 선택된 외인성 GPT를 발현하는 CHO 세포는 2.5 µg/mL Tn에서 더 배양된다; 1 µg/mL Tn 선택 풀에서 선택된 외인성 GPT를 발현하는 CHO 세포는 5 µg/mL Tn에서 더 배양된다. **도 6b**는 동일 세포 풀(도 6a에 도시)에 대한 qPCR에 의해 측정된 관심 유전자 eGFP의 상대적 유전자 사본 수를 나타낸다.

**도 7a 내지 7d**는 세포 배양으로부터 생산된 Fc-융합 단백질 1(FcFP1)의 글리코폼 특성을 나타내며 다음과 같다: **도 7a**는 도 7b의 CHO-GPT를 발현하고 Tn 선택이 없는 CHO 세포(Lot 110728)와 비교한 결과, 표준 프로토콜을 사용한 GPT 발현없는 CHO 세포(Lot B10002M410)을 나타낸다. **도 7c**는 도 7d의 CHO-GPT를 발현하고 5 µg/mL Tn으로 선택된 CHO 세포(Lot 110728-02)와 비교한 결과, CHO-GPT를 발현하고 1 µg/mL Tn으로 선택된 CHO 세포(Lot 110728-01)를 나타낸다. 각각의 크로마토그램은 시알화(sialylated) 잔기를 함유한 부분을 나타내며 다음과 같다: 0SA = 시알산 잔기 없음; 1SA = 하나의 시알산 잔기; 2SA = 두 개의 시알산 잔기; 3SA = 세 개의 시알산 잔기; 4SA = 네 개의 시알산 잔기.

**도 8**은(A) Lot B10002M410, (B) Lot 110728, (C) Lot 110728-01 및 (D) Lot 110728-02으로부터 샘플링된 Fc-융합 단백질 1(FcFP1)의 중첩된 당질화 프로파일을 나타낸다. 상기 GPT 룯트로부터 생산된 각 단백질의 글리코프로파

일(glycoprofile)은 참조 표준 단백질과 호환될 수 있으며 다수의 글리코폼 중은 지속적으로 생산된다. 상기 참조 표준 단백질과 비교해서 상기 GPT 루트에서 새롭고 독특한 글리코폼 종이 생산되지 않았음이 명백하다.

**발명을 실시하기 위한 구체적인 내용**

- [0031] 본 발명의 방법을 설명하기 전에, 본 발명은 특정 방법 및 설명된 실험 조건이 다양하여 이러한 방법 및 조건에 제한되지 않는 것으로 이해해야 한다. 또한 본 발명의 범위는 단지 첨부된 청구항에 의해 제한되므로 본 발명에서 사용된 용어는 특정 실시예를 설명하고자할 뿐 제한하는 것은 아닌 것으로 이해해야 한다.
- [0032] 본 명세서 및 첨부된 청구항에 사용된 바와 같이, 문맥에 달리 명시되어 있지 않는 한 단수 형태인 “a”, “an”, 및 “the” 는 다수의 참조를 포함한다. 따라서, 예를 들어, “하나의 방법”에 대한 참조는 하나 이상의 방법 및/또는 본 발명에서 설명 및/또는 본 명세서를 읽으면 당업자에게 명백해질 종류의 단계를 포함한다.
- [0033] 달리 정의되거나 규정되지 않는 한, 본 명세서에 사용된 모든 기술적 및 과학적 용어는 본 발명이 속한 당해 기술분야의 당업자에 의해 통상 이해되는 것과 동일한 의미를 가진다.
- [0034] 본 발명에 기재된 방법 및 재료와 유사 또는 동등한 어떠한 방법 및 재료도 본 발명의 실시 또는 검사에 사용될 수 있으나, 특정 방법 및 재료를 지금부터 설명한다. 본 발명에 언급된 모든 문헌은 그 전체가 본 발명에 참조로 인용되어 있다.
- [0035] 당해 기술 분야에 잘 알려진 다양한 유전자는 배양시 포유류 세포에 선택성 표현형을 부여할 수 있다. 통상, 선택성 마커 유전자는 단백질, 보통 세포 배양시 다양한 항생제에 저항성을 부여할 수 있는 효소를 발현한다. 몇몇 선택적 조건에서, 형광 단백질 마커를 발현하는 세포는 가시화되어 선택가능하다. 당해 기술분야에서 실시예는 베타-락타마아제(bla; 베타-락탐 항생제 저항성 유전자 또는 ampR; 암피실린 저항성 유전자), bls(블라스티시딘 저항성 아세틸 트랜스퍼라제 유전자), 히그로마이신 포스포트랜스퍼라제(hpt; 히그로마이신 저항성 유전자), 및 기타를 포함한다.
- [0036] 본 명세서에 기재된 방법은 튜니카마이신 및 튜니카마이신에 저항성 있는 세포를 세포 배양시 성장시킬 수 있는 효소(마커)의 사용에 의존한다. 튜니카마이신(Tn)은 N-아세틸글루코사민 지질 중간체의 형성 및 새로 합성된 당 단백질의 당질화를 저해하는 세균 및 진핵생물의 N-아세틸글루코사민 트랜스퍼라제의 저해제로 작용하는 항생제 혼합물이다. (King, I.A., 및 Tabiwo, A., 1981, 생체의 표피 당단백질 및 글리코사미노글리칸 합성에 대한 튜니카마이신의 효과. *Biochem. J.*, 198(2):331-338). Tn은 특히 UDP-N-아세틸글루코사민을 저해하기 때문에 세포독성이다: 돌리콜 포스페이트 N-아세틸글루코사민-1-P 트랜스퍼라제(dolichol phosphate N-acetylglucosamine-1-P transferase; GPT), 돌리콜-연결된 올리고당의 생합성의 초기 단계를 촉매화하는 효소. 튜니카마이신의 존재 하에, 소포체(ER) 내에서 만들어진 아스파라긴-연결된 당단백질은 N-연결된 글리칸으로 당질화되지 않으므로 상기 ER 내에서 정확하게 접히지 않을 수 있어 분해될 수 있다(Koizumi, 등, 1999, *Plant Physiol.* 121(2):353-362). 따라서, Tn은 세균 및 진핵 세포에서 세포사멸을 일으키는 접히지 않은 단백질 반응(UPR)의 주목할 만한 인ductor이다.
- [0037] 우리딘 디포스페이트 GPT(GlcNAc-1-P 트랜스퍼라제로도 알려진)에 대한 유전자는 Tn에 저항성을 부여하기 위해 임의의 세포 조건 하에 과발현됨에 따라 확인되었다(Crisuolo and Krag, 1982, *J Biol Chem*, 263(36):19796-19803; Koizumi, 등, 1999, *Plant Physiology*, Vol. 121, pp. 353-361). GenBank Accn. No. M36899(서열번호 2)에도 기재된GPT를 암호화하는 유전자는 Tn-저항성 중국 햄스터 난소 세포주로부터 분리되었고 408 아미노산 단백질(서열번호 3)을 암호화한다(Scocca and Krag, 1990, *J Biol Chem* 265(33):20621-20626; Lehrman, M. 등, 1988, *J Biol Chem* 263(36):19796-803). 햄스터 GPT는 효모 세포(*S. pombe*) 내에서 과발현되었고 이들 세포 내에서 Tn 저항성을 부여하였으며, 또한 상기 GPT 효소의 정제를 위한 편리한 소스를 제공한다(Scocca JR, 등. 1995, *Glycobiology*, 5(1):129-36). GPT의 전사 수준을 하이브리도마 세포에서 분석하는 반면(IgG을 발현하는 B 세포 vs. 휴지기 B 세포) IgG-생산 세포는 상승된 GPT 전사 또는 활성 레벨을 보이지 않았으나, GPT의 적은 증가는 휴지 상태에서 활성화된 B 세포로의 이행에서 관찰되었다. GPT 레벨은 B 세포 내 LPS(항원) 자극에 대한 증식 반응의 초기 발달과 일치할 수 있는 것으로 결론내렸다(Crick, D.C. 등, 1994, *J Biol Chem* 269(14):10559-65).
- [0038] 또한, 세포 발현 체계에서 Tn의 존재 또는 부존재에 따라 GPT의 발현이 변화하는지 여부가 단백질 산물의 당질화에 영향을 미치므로 산물 품질에 영향을 미칠 것이라는 것에 대해서는 이전에 알려진 바 없었다. 최적화되고 일관된 당질화는 치료적 당단백질의 생산에 있어 중요한 단백질 속성으로 이해된다.

- [0039] 본 발명은 포유류 Tn-저항성 유전자, GPT를 규제가능한 선택 마커로 사용하는 포유류 세포 시스템에서 재조합 단백질을 생산하는 개선된 방법을 제공하는 반면, GPT에 작동가능하게 연결된 관심 유전자의 사본수의 증가는 GPT 발현 카세트의 상기 세포 내로의 무작위 통합의 증가와 상관 관계가 있다.
- [0040] 당해 기술은 치료적 단백질, 특히, 당단백질의 제조는 이러한 단백질의 자연 당질화를 모방하는 포유류타입 발현 체계에 의존하는 것을 인정해왔다. (리뷰를 위해, Bork, K. 등, 2009, J Pharm Sci. 98(10):3499-3508. 참조) 예를 들어, N연결된 복합체 글리칸과 같은 어떤 당단백질의 최종 단당류는 일반적으로 시알산에 의해 점유된다. 시알릴화는 당단백질의 흡수, 혈청반감기 및 제거율 같은 약동학적 성질, 또는 당단백질의 기타 물리화학적 또는 면역원성에 영향을 미칠 수도 있다. 과발현된 재조합 당단백질은 종종 불완전하거나 불연속적인 당질화를 가진다. 신뢰성 있는 방법은 포유류 세포주에서 생산된 치료적 당단백질의 공정 일관성 및 품질에 있어 매우 중요하다.
- [0041] 본 발명은 목적하는 단백질의 일관된 품질 산출량을 제공하기 위해 재조합 단백질의 개선된 당질화 방법, 즉, 포유류 세포 시스템에서 당단백질을 만드는 방법 또한 제공한다.
- [0042] 정의
- [0043] DNA 영역은 이들이 각각 기능적으로 연관되어 있을 때 작동가능하게 연결된다. 예를 들어, 프로모터는 상기 프로모터가 상기 서열의 전사에 참여할 수 있으면 코딩 서열에 작동가능하게 연결되고; 리보솜-결합 부위는 번역이 가능하도록 배치되는 경우 코딩 서열에 작동가능하게 연결된다. 일반적으로, 작동가능하게 연결된다는 것은 접근을 포함하나, 요구되지는 않는다. 분비 리더(leaders)와 같은 서열의 경우, 접근 및 리딩 프레임(reading frame)에서 적절한 배치는 일반적인 특징이다. 프로모터와 같은 생산 향상 서열은 GOI에 기능적으로 관련된, 예를 들어, 그것의 존재가 상기 GOI의 발현을 증가시키는 관심 유전자(GOI)에 작동가능하게 연결된다.
- [0044] 이와 같이, DNA 발현 벡터 구조체의 문맥에 기술된 바와 같이 문구 “작동가능하게 연결된”은 제어 서열, 예를 들어, 프로모터 또는 작동자 또는 마커는 코딩 서열에 상대적인 위치에 적절하게 배치되어 상기 제어 서열은 상기 코딩 서열에 의해 암호화된 목적하는 폴리펩티드/단백질의 생산을 지시 또는 허용한다. 예를 들어, 선택 마커는 어떤 배양 조건에서 세포가 생존할 것을 요구하나, 작동가능한 선택 마커 단백질의 존재 없이는 발현이 일어나지 않을 것이므로 상기 관심 유전자는 상기 선택 마커 유전자에 작동가능하게 연결된다.
- [0045] 본 발명에 사용된 “프로모터”는 작동가능하게 연결된, 즉, 적절한 신호가 있을 때 상기 관심 유전자 및/또는 선택 마커 유전자의 전사를 허용하는 방식으로 연결된 DNA 서열의 직접 전사에 충분한 DNA 서열을 나타낸다. 유전자의 발현은 당해 기술 분야에 알려진 어떤 프로모터 또는 인핸서 요소의 조절 하에 배치될 수 있다.
- [0046] 본 발명의 문맥에서 “발현 벡터”는 염색체, 비-염색체, 및 합성 핵산 벡터(적합한 발현 제어 요소 세트를 포함하는 핵산 서열)을 포함하는 어떤 적합한 벡터일 수 있다. 이러한 벡터의 예는 SV40의 유도체, 세균 플라스미드, 파아지 DNA, 바콜로바이러스, 효모 플라스미드, 플라스미드 및 파지 DNA의 조합에서 유도된 벡터, 및 바이러스 핵산(RNA 또는 DNA) 벡터를 포함한다. 일 실시예에서, Fc-융합 단백질 또는 폴리펩티드-암호화된 핵산 분자는, 예를 들어, 선형 발현 요소(가령, Sykes and Johnston, 1997, *Nat Biotech* 12, 355-59에 기재된 것과 같이), 작은 핵산 벡터(가령, US6,077,835 및/또는 W000/70087에 기재된 것과 같이), 또는 pBR322, pUC 19/18, 또는 pUC 118/119와 같은 플라스미드 벡터를 포함하는, 노출된 DNA 또는 RNA 벡터에서 구성된다. 이러한 핵산 벡터 및 이들의 용도는 당해 기술분야에 잘 알려져 있다(가령, US5,589,466 및 US5,973,972에 기재되어 있음).
- [0047] 본 발명에서 사용된 바와 같이 “작동자”는 상기 작동자에 억제 단백질의 결합하여 상기 유전자를 조절할 수 있는 방식으로 유전자 내 또는 가까이에 도입된 DNA 서열을 나타내어, 그 결과, 상기 GOI, 즉, 관심있는 폴리펩티드 또는 단백질을 암호화하는 뉴클레오티드의 전사를 저해 또는 허용한다.
- [0048] 리보솜 결합 부위는 “내부 리보솜 유입점(IRES)”을 포함하거나 또는 5' 캡을 포함할 수 있다. 많은 IRES 서열은 당해 기술 분야에 잘 알려져 있다. IRES는 번역 제어 서열을 나타내며, 상기 IRES 부위는 일반적으로 관심 유전자의 5'에 위치하고 캡-독립적 방식으로 상기 RNA를 번역한다. 전사된 IRES는 직접적으로 리보솜 서브유닛을 결합하여 mRNA의 개시 코돈의 위치가 번역을 위해 적절하게 리보솜 내에서 배향될 수 있다. IRES 서열은 일반적으로 상기 mRNA의 5' UTR에 위치한다(상기 개시 코돈의 바로 상류에). IRES는 진핵생물의 번역기와 상호작용하는 다양한 단백질 요소의 필요성을 기능적으로 대체한다.
- [0049] 단백질 발현을 설명하기 위해 사용하는 경우 “향상된(enhanced)” 또는 “개선된(improved)”이라는 용어는 본 발명의 발현 체계 또는 방법에 의해 생산된 상기 단백질(즉, 유전자 산물)의 양 및/또는 품질의 일관성의 증가

를 포함한다. 이와 같이, 이는, 예를 들어, 다른 선택 마커 구조체를 사용하는 통합 성분의 풀과 비교할 때, 게놈으로의 무작위 통합에 의해 일반적으로 관찰되는 것보다 발현이 적어도 약 1.5-배에서 적어도 약 3-배 향상되는 것을 포함한다. 이와 같이, 목적하는 단백질에 대해 관찰된 접합-발현 향상을 GPT 유전자를 포함하는 본 발명의 발현 카세트 또는 세포의 부재 또는 다른 선택 마커를 포함하는 발현 카세트 또는 세포의 존재 시 실질적으로 동일한 조건 하에서 측정된 동일 유전자의 발현양과 비교한다. 또한 그 결과 무작위 통합 사건의 수에 의해 발현 향상을 측정할 수 있다. 향상된 재조합 효율은 재조합을 위해 유전자자리의 능력 향상을 포함한다(예를 들어, 재조합효소-인식 부위 이용). 향상은 무작위 재조합보다 측정가능한 효율을 의미하며, 일반적으로 0.1%이다. 특정 조건에서, 향상된 재조합 효율은 무작위보다 약 10배 또는 약 1%이다. 명시되지 않은 한, 청구된 발명은 특정한 재조합 효율에 제한되지 않는다. 정량적 중합효소 연쇄 반응(qPCR) 또는 다른 잘 알려진 기술에 의해 측정됨에 따라 결과적인 유전자 사본 수에 의해 발현 향상 역시 측정될 수 있다.

[0050] 향상된 또는 개선된 산물은 또한 보다 일관된 품질, 예를 들어, 본 발명의 GPT 발현 체계로 관찰되는 번역후 변형을 의미한다. 일관된 품질은, 예를 들어, 복제 생산 라인 다음 목적하는 당질화 프로파일을 가지는 것을 포함한다. 품질에 대한 일관성은 불균일도 및 표준화를 의미하는 반면, 복제 생산 배치(batches)는 본질적으로 변형이 없다. 일관성을 측정하기 위한 Z-수의 계측을 본 발명에 고시한다. 일관성을 측정하기 위해 다른 통계적 방법이 당해 기술 분야에 알려져 있다.

[0051] 상기 문구 “선택압”은 주어진 환경 안에서 상기 생물 또는 시스템의 행위 및 생존(생존력과 같은)을 변화되는 생물(예를 들어, 세포) 또는 시스템(예를 들어, 발현 체계)에 적용된 힘 또는 자극을 의미한다.

[0052] 상기 문구 “유전자 증폭”은 유전자 서열의 동등한 사본의 수의 증가를 의미한다. 어떤 세포성 공정은 특정 유전자의 복수개의 사본 또는, 예를 들어, 항생제 저항성과 같이 상기 유전자가 세포에 대해 부여하는 표현형을 증폭시키는 유전자의 생성에 의해 특징 지어진다.

[0053] 상기 문구 “외생적으로 부가된 유전자” 또는 “외생적으로 부가된 GOI”은 발현 카세트를 참조하여 사용되나, 상기 문구는 자연에서 발견되는 세포 게놈 내에 존재하지 않는 어떤 유전자, 또는 상기 게놈(내 다른 자리)에 통합된 추가적인 유전자 사본을 의미한다. 예를 들어, CHO 게놈(예를 들어, 선택 마커 유전자) 내에 “외생적으로 부가된 유전자”는 자연에서 특정 CHO 자리 내에서 발견되지 않는 햄스터 유전자(즉, 햄스터 게놈 내 다른 자리에서 유래된 햄스터 유전자), 다른 종에서 유래된 유전자(예를 들어, 인간 유전자) 또는 키메라 유전자(예를 들어, 인간/마우스) 일 수 있고, 또는 자연에서 CHO 게놈 내에서 확인되지 않는 햄스터 유전자(즉, 햄스터 게놈의 다른 자리 유래의 유전자와 99.9% 미만의 동일성을 가진 햄스터 유전자), 또는 CHO 자연 게놈 내에 존재하는 것으로 자연에서는 발견되지 않는 다른 유전자일 수 있다.

[0054] 무작위 통합 사건은 표적 통합 사건과 다른 반면, 세포의 게놈 내 유전자 삽입은 무작위 통합 사건에서 부위-특이적인 것은 아니다. 표적 통합의 예는 동종 재조합이다. 무작위(비-동종) 통합은 상기 결과로 얻은 통합 성분의 위치(자리)가 알려져 있지 않거나 특정되는 것을 의미한다. 무작위 통합은 비동종 말단 연결(NHEJ)에 의해 일어난 것으로 생각되나, 상기 방법에 제한되는 것은 아니다.

[0055] 선택 효율은 선택 마커를 발현하는 생존 세포 및 적용 가능하다면, 상기 선택 마커의 제어 하에 상기 목적 단백질의 퍼센트 균집을 의미한다.

[0056] Tn-저항성 단백질을 기술할 때, 동일성 %은 연속적인 상동성의 영역을 따라 인용된 동일성을 표시하는 상동 서열을 포함하는 것을 의미하지만, 비교된 서열에는 상동성이 없는 갭, 결실 또는 삽입의 존재는 동일성 %을 계산할 때 고려되지 않는다. 이 문맥에서 “동일성 %”의 사용을 설명함에 있어서, 하기의 아미노산 서열 비교가 참조될 것이다:

[0057] 1 MWAFPELPLPLLVLNIGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSSQQIPESQ 60 GPT\_MOUSE

[0058] 1 MWAFPELPL--PLLVLNFGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSSRQQIPESQ 58 GPT\_CRIG

[0059]

[0060] 본원에서 사용된 마우스 상동체 (“GPT\_MOUSE”)와 상기 “GPT\_CRIG” 서열(중국 햄스터 GPT) 간의 “동일성 %” 결정은 햄스터 아미노산 10 및 11의 비교를 포함하지 않는데, 햄스터 상동체는 정렬(즉, 마우스 GPT가 그 지점에서 삽입을 가지거나, 경우에 따라, 햄스터 상동체가 갭 또는 결실을 가짐)에서 비교할 상동 서열이 없기 때문이다. 따라서, 상기의 비교에서, 동일성 % 비교는 5' 말단의 “MWA”에서 3' 말단의 “ESQ”까지 확장된다. 그 경우, 마우

스 상동체는 햄스터 GPT 위치 51에서 "R"이 있다는 것만 상이하다. 비교는 60개 염기쌍의 스트레치에서 58개 넘는 연속적인 염기이기 때문에, 오직 한 개의 아미노산 차이가 있는 경우 (갭, 결실 또는 삽입이 아님), 햄스터 GPT 위치 1에서 햄스터 GPT 위치 58까지 두 서열 (햄스터 및 마우스) 간에 98% 넘는 동일성이 있다 ("동일성 %"는 갭, 결실 및 삽입에 대한 페널티를 포함하지 않기 때문). 상기의 예는 아미노산 서열에 기초하고 있지만, 핵산 서열 동일성 %은 동일한 방식으로 계산된다는 것이 이해된다.

[0061] 용어 "세포"는 재조합 핵산 서열을 발현하는데 적합한 임의의 세포를 포함한다. 세포는 원핵 세포 및 진핵 세포 (단세포 또는 다세포), 박테리아 세포(예를 들어, *E. coli*, *Bacillus* spp., *Streptomyces* spp. 등), 마이코박테리아세포, 진균 세포, 효모 세포(예를 들어, *S. cerevisiae*, *S. pombe*, *P. partoris*, *P. methanolica* 등), 식물 세포, 곤충 세포(예를 들어, SF-9, SF-21, 배콜로바이러스에 감염된 곤충 세포, *Trichoplusia ni* 등), 비인간 동물 세포, 포유류 세포, 인간 세포, 또는 세포 용합물, 예를 들어, 하이브리도마 또는 쿼드로마 (quadroma)를 포함한다. 소정의 실시예들에서, 세포는 인간, 원숭이, 유인원, 햄스터, 랫트, 또는 마우스 세포이다. 다른 실시예들에서, 세포는 진핵 세포이고, 다음의 세포로부터 선택된다: CHO(예를 들어, CHO K1, DXB-11 CHO, Veggie-CHO), COS(예를 들어, COS-7), 망막 세포, Vero, CV1, 신장(예를 들어, HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L 세포, C127 세포, SP2/0, NS-0, MMT 세포, 종양 세포, 및 전술된 세포로부터 유래한 세포주. 일부 실시예에서, 세포는 하나 이상의 바이러스 유전자, 예를 들어, 바이러스 유전자를 발현하는 망막 세포(예를 들어, PER.C6® cell)를 포함한다.

[0062] 상기 문구 "통합된 세포 밀도" 또는 "ICD"는 mL당 세포-일수로 표현되는 기간에 대한 적분으로 얻어진 배양 배지 내 세포의 밀도를 의미한다. 일부 실시예에서, 상기 ICD는 세포 배양 12일째에 측정된다.

[0063] 상기 단어 "당질화" 또는 문구 "단백질의 당질화"는 당단백질의 형성을 포함하는 반면, 올리고당이 단백질의 아스파라긴(Asn) 잔기(즉, N-연결) 또는 세린(Ser)/트레오닌(Thr) 잔기(즉, O-연결)의 측쇄중 어느 하나에 부착된다. 글리칸은 선형 또는 분지형의 당당류 잔기의 동종 또는 이종중합체일 수 있다. N-연결 당질화는 주로 소포체에서 시작되는 것으로 알려진 반면, O-연결 당질화는 소포체(ER) 또는 골지체중 어느 하나에서 시작되는 것으로 보여진다.

[0064] 용어 "N-글리칸 단백질" 또는 "N-글리칸 단백질 기질"은 N-연결 올리고당을 함유하거나 수용할 수 있는 단백질들을 포함한다. N-글리칸은 N-아세틸 갈락토사민(GalNAc), 만노오스(Man), 푸코오스(Fuc), 갈락토오스(Gal), 뉴라민산(NANA), 및 다른 당당류로 구성될 수 있으나, N-글리칸은 대개 세 개의 만노오스 및 두 개의 N-아세틸 글루코사민(GlcNAc) 당류를 포함하는 공통 중심 오탄당 구조를 가진다. X가 프로린을 제외한 어떤 아미노산인 연속된 아미노산 서열(즉, 시퀀) Asn-X-Ser 또는 Asn-X-Thr을 가진 단백질은 N-글리칸에 대한 부착 부위를 제공할 수 있다.

[0065] **일반적인 설명**

[0066] 본 발명은 적어도 일부로 어떤 조건 하에서 재조합 단백질이 세포 내에서 생산될 수 있다는 발견을 기반으로 하며 상기 단백질을 암호화하는 상기 유전자는 Tn-저항성 유전자, GPT에 작동가능하게 연결되고 단백질-생산 세포의 선택은 세포 계능에서 무작위 통합 사건을 증가시도록 구성되어 상기 관심 유전자의 사본 수를 증가시키고 결국 단백질이 생산된다.

[0067] 본 발명은 또한 적어도 일부로 단백질-생산 세포가 일관성 및 신뢰성 있는 번역후 변형이 있는 단백질을 발현하는데 최적화될 수 있다는 발견을 기반으로 한다. GPT 발현 카세트는 또한 당해 기술 분야에 알려진 다양한 유전자 편집 기술을 사용해 발현 벡터 경유와 같은 발현 구조체 내 세포 계능에 통합될 수 있다. GPT를 포함하는 발현 벡터는 특정 재조합 부위를 인식하는 재조합 효소에 의해 매개된 동종 재조합 또는 재조합(예를 들어, Cre-lox-매개 재조합)과 같은 무작위 또는 표적 재조합에 의해 계능에 통합될 수 있다.

[0068] 진핵 세포에서 동종 재조합은 상기 통합 부위에서 염색체 DNA 내 절단을 도입하여 촉진될 수 있다. 모델 시스템은 염색체 목적 서열 내 이중 가닥 절단이 도입되면 유전자 표적화시 동종 재조합 빈도가 증가하는 것을 설명했다. 이는 통합 특정 부위에 어떤 뉴클레아제를 표적화하여 수행될 수 있다. 상기 목표 자리에서 DNA 서열을 인식하는 DNA-결합 단백질은 당해 기술 분야에 알려져 있다. 유전자 표적화 벡터는 또한 동종 재조합을 용이하게 하는데 사용된다. 상동성 지시된 수선을 위한 유전자 표적화 벡터의 부재시, 상기 세포는 상기 절단 부위에서 다수의 뉴클레오티드의 삭제 또는 삽입시킬 수 있는 비-동종 말단-연결(NHEJ)에 의해 상기 이중 가닥 절단을 종종 폐쇄한다. 유전자 표적화 벡터 제작 및 뉴클레아제 선택은 본 발명이 속하는 기술자의 기술에 의한다.

- [0069] 몇몇 예에서, 모듈러 구조를 가지며 개별 징크 핑거 도메인을 함유하는 징크 핑거 뉴클레아제(ZFNs)는 상기 목적 서열(예를 들어, 표적화 통합 부위) 내 특정 3-뉴클레오티드 서열을 인식한다. 몇몇 실시예는 다수의 목적 서열을 표적화하는 개별 징크 핑거 도메인의 조합과 함께 ZFN을 사용할 수 있다.
- [0070] 전사 촉진제-유사(TAL) 이펙터 뉴클레아제(TALENs)는 또한 부위-특이적 게놈 편집을 위해 사용될 수 있다. TAL 이펙터 단백질 DNA-결합 도메인은 일반적으로 FokI과 같은 제한 뉴클레아제의 비-특이적 절단 도메인과 조합하여 사용된다. 일부 실시예에서, TAL 이펙터 단백질 DNA-결합 도메인 및 제한 뉴클레아제 절단 도메인을 포함하는 융합 단백질은 본 발명의 유전자자리 내 목적 서열에서 DNA를 인식하고 절단하는데 사용된다(Boch J 등, 2009 *Science* 326:1509-1512).
- [0071] RNA 유전자가위(RNA-guided endonucleases; RGENs)는 세균의 적응 면역 기계로부터 개발된 기구를 조작하는 프로그램 가능한 게놈이다. 이 시스템-간헐적으로 반복되는 회문 구조 염기 서열 집합체(the clustered regularly interspaced short palindromic repeats; CRISPR)/CRISPR-관련(Cas) 면역 반응-에서 두 개의 RNA로 복합체를 형성할 때 상기 단백질 Cas9은 서열-특이적 엔도뉴클레아제를 형성하며 이들 RNA중 하나는 표적 선택을 유도한다. RGEN는 구성성분(Cas9 및 tracrRNA) 및 표적-특이적 CRISPR RNA(crRNA)로 구성된다. DNA 표적 절단 효율 및 절단 부위의 위치 모두는 프로토스페이스 인접 모티프(protospacer adjacent motif; PAM)의 위치, 표적 인식에 대한 추가 요건에 따라 다양하다(Chen, H. 등, *J. Biol. Chem.* 원고 M113.539726로 2014년 3월 14일자 온라인 출판).
- [0072] 정확한 DNA-결합 특이성을 가진 BuD-유래 뉴클레아제(BuDNs)와 같은 동종 재조합의 또 다른 방법은 숙련된 기술자가 이용할 수 있다(Stella, S. 등, *Acta Cryst.* 2014, D70, 2042-2052). 정확한 게놈 변형 방법을 상기 게놈 내에서 독특한 목적 서열과 호환가능한 기구에 따라 선택하여 세포 표현형의 파괴를 피할 수 있다.
- [0073] 핵산 서열(관심 유전자)을 포유류 세포로 안정되게 통합되기 위한 세포 및 방법이 제공되며, 상기 핵산 서열은 GPT 서열과 통합되어 발현을 향상시킨다. 또한 GPT를 예를 들어, 발현 벡터와 같은 발현 구조체와 결합하여 사용하고 외인성 GPT를 목적 포유류 세포에 첨가하기 위한 조성물 및 방법이 제공된다. 세포 및 방법은 당단백질, 특히, 치료적 당단백질을 만드는 지속적이나 강력한 방법에 사용되기 위해 제공된다.
- [0074] **GPT 선택 마커 카세트의 제작**
- [0075] 본 발명은 작동하는 GPT 발현 카세트로 구성된 발현 벡터를 제공한다. 상기 발현 카세트는 포유류 GPT 및 목적하는 유전자 산물의 전사 및 번역을 허용하고 촉진하는데 필요한 조절 요소를 포함한다.
- [0076] 상기 유전자 및 본 발명에 기재된 조절 서열의 다양한 조합 역시 개발될 수 있다. 역시 개발될 수 있는 본 발명에 기재된 적합한 서열의 다른 조합 예는 최적의 조절 요소 조합을 얻기 위해 본 발명에 개시된 상기 GPT 유전자의 다수의 사본, 또는 상기 개시된 GPT과 다른 뉴클레오티드 서열의 혼합에 의해 파생된 서열을 포함한다. 이러한 조합은 상기 관심 유전자 및 조절 요소에 배향된 GPT의 최적 간격을 제공하기 위해 연속적으로 연결 또는 배열될 수 있다.
- [0077] GPT를 암호화하는 유전자의 동종 서열은 다른 포유류 조직 타입으로부터 유래된 세포주 내뿐 아니라 다른 포유류 종(예를 들어, 인간; 도 2 참조)에서 유래된 세포 내에 존재하는 것으로 알려져 있으며, 당해 기술 분야에 잘 알려진 기술에 의해 분리될 수 있다. 포유류 GPT 아미노산 서열의 예시적인 리스트가 도 2에서 제공된다. 코돈 최적화와 같은 뉴클레오티드 서열의 변화는 서열번호 3 내지 10에 기재된 해당 GPT 단백질의 최적의 발현을 허용하기 위해 서열번호 2 및 11 내지 17에 기재된 뉴클레오티드 서열 내에서 이루어질 수 있다. 게다가, 변화는 GPT를 암호화하는 뉴클레오티드 서열에 변화를 주어 서열번호 3 내지 10에 기재된 아미노산 서열 내에서 일어날 수 있다. 부위-지정 또는 무작위 돌연변이 유발을 포함하나 이에 제한되지 않는 이러한 기술은 당해 기술 분야에 잘 알려져 있다.
- [0078] 그래서 그 결과로 얻은 GPT 변이체는 본 발명에서 설명된 바와 같이 GPT 활성화에 대해 실험, 예를 들어, 튜니카 마이신에 대한 저항성에 대해 실험 할 수 있다. 아미노산 서열에 있어 GPT 활성을 가진 서열번호 3에 적어도 약 93%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 또는 적어도 약 98% 동일성이 있는 GPT 단백질은 일상적인 실험으로 분리가능하고, Tn에 대한 동일한 저항성, 선택 효율 및 서열번호 3에 대해 번역후 이점을 나타내는 것으로 기대된다. 따라서, GPT의 포유류 동족체 및 GPT의 변이체는 또한 본 발명의 실시예에 의해 포함된다. 도 2a 내지 2c는 다양한 포유류 GPT 아미노산 서열 (즉, 서열번호 3 내지 10)의 정렬을 보여준다. 포유류 GPT 서열 (핵산 및 아미노산)은 햄스터, 인간, 마우스 및 랫 게놈 간에 보존된다. 표 1은 예시적인 포유류 GPT 단백질 및 이의 상동성 정도를 확인한다.

표 1

【표 1a】

GPT 상동체의 아미노산 일치도

동물	서열번호	인간 일치도%	마우스 일치도%	랫 일치도%	햄스터 일치도%
햄스터	3	93.87	96.08	96.08	-
마우스	10	94.12	-	97.07	96.08
인간	4	-	94.12	93.63	93.87
랫	9	93.63	97.07	-	96.08

【표 1b】

대표적인 GPT 상동체의 핵산 일치도

동물	서열번호	햄스터 일치도%
햄스터	2	-
마우스	11	92
인간	12	92
랫	13	94
마카크(Macaque)	14	92
침팬지	15	92

[0079]

[0080]

강화된 수준의 관심 단백질을 발현하는 세포 집단은 본원에 제공된 GPT/튜니카마이신 방법을 사용하여 개발할 수 있다. 절대 발현 수준은 단백질이 세포에 의해 얼마나 효율적으로 처리되는지에 따라, 특이적 단백질에 따라 달라진다.

[0081]

따라서, 본 발명은 또한 서열번호 2 및 11-17로 이루어진 군으로부터 선택된 GPT-발현 뉴클레오티드 서열을 포함한다. 본 발명은 또한 서열번호 2 및 11-17로 이루어진 군으로부터 선택된 뉴클레오티드 서열과 적어도 92% 동일하거나, 적어도 93% 동일하거나, 적어도 94% 동일하거나, 적어도 95% 동일하거나, 적어도 96% 동일하거나, 적어도 98% 동일하거나 또는 적어도 99% 동일한 GPT-발현 뉴클레오티드 서열을 포함한다.

[0082]

본 발명은 서열번호 1, 서열번호 2 또는 서열번호 12를 포함하는 벡터를 포함한다. 포유류 GPT 유전자, 및 선택적 조절 요소를 포함하는 벡터는 일시적 또는 안정적 형질감염을 위한 벡터를 포함한다.

[0083]

일 실시예에서, 도 1에 예시된 바와 같이, GPT 유전자는 GOI의 발현을 향상시키는데 사용된다. 도 1은 IRES 서열 및 GPT 선택성 마커와 작동 가능하게 연결된 GOI를 도시한다. GPT 카세트는 예를 들어, SV40 프로모터 및 폴리아데닐화(폴리(A)) 서열, 예를 들어 SV40 폴리(A)을 더 포함한다.

[0084]

발현 강화 카세트 (GPT와 상류 프로모터 포함)는 세포 계능에 최적으로 편입된다. 본 발명의 방법을 사용하여, GOI는 증가하는 농도의 Tn에 기초한 배양 조건 하에서 GPT 발현 카세트 내에서 발현된다 (도 3a 또는 도 3b). 도 5b, 5c, 5e 및 5f에 도시된 것과 같은, FACS 판독은 안정적으로 형질감염된 세포 집단에서의 발현의 분포, 특히 포유류 Tn-저항성 선택 마커, CHO-GPT 및 hGPT를 이용하여 극적인 선택 효율의 증가를 예시한다. 포유류 GPT 발현은 관심 유전자 산물, 예를 들어 형광 단백질, eGFP의 생산의 발현을 더욱 향상시킨다. Tn의 농도를 증가시키는 연속 배양의 결과 도 6b에서 예시된 바와 같이, Tn의 한 농도에 기초한 배양 조건 하에서 GPT를 사용하는 발현 체계에서 발현된 GOI와 비교하여 약 2배의 향상된 발현을 초래한다.

[0085]

본 발명은 GPT 유전자가 외인성이고 본 발명의 방법에 의해 세포 계능에 통합되는 이러한 GPT 유전자를 포함하는 포유류 세포를 포함한다. GPT 유전자의 상류 또는 하류에 있는 적어도 하나의 외생적으로 부가된 관심 유전자(GOI)를 갖는 이러한 GPT 유전자를 포함하는 세포.

[0086]

다양한 실시예에서, GOI의 발현은 포유류 선택성 마커 GPT의 제어 하에 GOI를 위치시킴으로써 향상될 수 있다. 다른 실시예들에서, 무작위 GOI의 통합 행위는 포유류 선택성 마커 GPT의 제어 하에 GOI를 위치시키고 0.5 μg/mL 초과 Tn 농도를 포함하는 세포 배양 조건을 제공함으로써 향상될 수 있다. 일부 실시예에서, 세포 배양 조건은 1 μg/mL 초과 Tn 농도를 포함한다. 조절 요소는 GOI에 작동 가능하게 연결될 수 있는데, 이때 GOI의 발현-GOI 및 GPT (5' 또는 3' 방향)로부터 선택된 거리에서-은 예를 들어, 일반적으로 무작위 통합 행위로 인해 관찰된 발현에 비해 GOI의 발현을 향상시키는 능력을 보유한다. 다양한 실시예에서, 향상은 약 1.5배 내지 약 2배 또는 그 이상이다. 무작위 통합, 또는 무작위 발현과 비교하여 발현 향상은 약 1.5배 또는 약 2배 또는 그 이상

이다.

- [0087] 또 다른 실시예에서, 균일하게 당질화된 단백질이 본 발명의 방법 및 조성물을 사용하여 달성될 수 있다. 표 4에서 볼 수 있듯이, Tn으로 처리한 GPT/GOI 재조합 단백질 배치(batch)는 동등한 당질화 프로파일을 갖는 복제물 배치를 허용한다. 이와 같이, 일관된 당질화 프로파일과 같은 향상된 단백질 발현은 본원에서 교시된 Z-수를 계산함으로써 직접적으로 비교될 수 있다. Z-수 방정식은 시알 산 (sialic acid; SA) 잔기를 나타내는 크로마토그램의 상대적 피크 수, 뿐만 아니라 각 피크의 상대적 모양 및 강도를 고려한다. Z-수는 각 피크가 점유한 면적에 기반하며 복합체 당단백질에 대한 일관성의 척도로 사용될 수도 있다 (예 도 7a 내지 도 7d, 도 8 및 실시예 3 참조).
- [0088] 단백질 발현 최적화는 또한 예를 들어, 발현 카세트 배향 또는 코돈 최적화를 포함하여, 각 GOI에 대해 달성될 수 있다. 단백질 최적화는 또한 세포 배양 방법에서 증분식 Tn 농도를 변화시킴으로써 달성될 수도 있다.
- [0089] 재조합 발현 벡터는 포유류, 바이러스 또는 곤충 유전자로부터 유래된 적절한 전사 및/또는 번역 조절 요소에 작동 가능하게 연결된, 단백질을 암호화하는 합성 또는 cDNA 유래 DNA 단편을 포함할 수 있다. 이러한 조절 요소는 본원에서 상세히 기술된 바와 같이, 전사 프로모터, 인핸서, 적절한 mRNA 리보솜 결합 부위를 암호화하는 서열, 및 전사 및 번역의 종결을 조절하는 서열을 포함한다. 또한 포유류 발현 벡터는 복제 기원, 다른 5' 또는 3' 측면에 위치하는 비-전사 서열, 5' 또는 3' 비-번역 서열, 예컨대 스플라이스 공여자 및 수여자 부위 같은 비-전사 요소를 포함할 수 있다. 형질감염체의 인식을 용이하게 하기 위한 부가적인 선택성 마커 유전자 (예, 형광 마커) 또한 통합될 수도 있다.
- [0090] 또 다른 실시예에서, 벡터는 기술된 핵산 분자(유전자)를 포함하는 발현 벡터를 포함하여, 관심 단백질을 암호화하는 핵산 분자(또는 관심 유전자)를 포함하며, 여기서 상기 핵산 분자(유전자)는 발현 조절 서열에 작동 가능하게 연결된다.
- [0091] 관심 유전자(GOI)를 포함하는 벡터가 제공되며, 여기서 상기 GOI는 포유류 숙주 세포에서의 발현에 적절한 발현 조절 서열에 작동 가능하게 연결된다.
- [0092] 본 발명에서 사용될 수도 있는 유용한 프로모터는 SV40 초기 프로모터 영역, Rous 육종 바이러스의 3' 긴 말단 반복체에 함유된 프로모터, 메탈로티오닌 유전자의 조절 서열, 마우스 또는 인간 세포거대바이러스 (cytomegalovirus) IE 프로모터 (Gossen 등, (1995) Proc. Nat. Acad. Sci. USA 89:5547-5551), 콜리플라워 모자이크 바이러스 35S RNA 프로모터, 광합성 효소 리볼로오스 이인산 카르복실라아제의 프로모터, 효모 또는 다른 균류 유래 프로모터, 예컨대 Gal 4 프로모터, ADC (알코올 탈수소효소) 프로모터, PGK (phosphoglycerol kinase) 프로모터, 알칼리 포스포타아제 프로모터, 및 조직 특이성을 나타내며 트랜스제닉 동물에서 사용된 하기 동물 전사 조절 영역: 엘라스타제 I; 인슐린; 면역글로블린; 마우스 유방 종양 바이러스; 알부민;  $\alpha$ -태아단백질;  $\alpha$ 1-항트립신;  $\beta$ -글로빈; 및 미오신 경쇄-2를 포함하지만, 이들에만 한정되지는 않는다.
- [0093] 본 발명의 핵산 분자는 또한 효과적인 폴리 (A) 종결 서열, 예를 들어 SV40 폴리(A), 대장균에서 플라스미드 산물을 위한 복제 기점 및/또는 간편한 클로닝 부위(예, 폴리링커)에 작동 가능하게 연결될 수도 있다. 또한 핵산은 CMV IE와 같은 구성적(constitutive) 프로모터와는 대조적으로 조절가능한 유도성(inducible) 프로모터 (유도성, 억제성, 발달적으로 조절됨)를 포함할 수도 있다 (당업자라면 그러한 용어들이 소정의 조건 하에서 실제로 유전자 발현 정도의 설명어임을 인식할 것이다).
- [0094] 본 발명은 관심 단백질을 생산하는 방법을 제공하며, 관심 유전자(GOI)를 포함하는 발현 벡터가 제공된다. 그러한 발현 벡터는 임의의 관심 단백질의 재조합 생산에 사용될 수도 있다. 척추동물 세포를 형질감염시키는 데 유용한 발현 벡터에서의 전사 및 번역 조절 서열은 바이러스 출처에 의해 제공될 수도 있다. 예를 들어, 일반적으로 사용되는 프로모터 및 인핸서는 폴리오마, 아데노바이러스 2, 시미안 바이러스 40 (SV40) 및 인간 세포거대 바이러스 (CMV)와 같은 바이러스로부터 유래한다. 그러한 조절 서열이 선택된 숙주 세포와 양립할 수 있다면, 바이러스 게놈 프로모터, 제어 및/또는 신호 서열이 발현을 유도하는데 사용될 수도 있다. 또한 재조합 단백질이 발현되어야 하는 세포 유형에 따라, 비-바이러스 세포 프로모터가 사용될 수도 있다 (예,  $\beta$ -글로빈 및 EF-1 $\alpha$  프로모터).
- [0095] SV40 바이러스 게놈, 예를 들어 SV40 기원, 초기 및 후기 프로모터, 인핸서, 스플라이스 및 폴리아데닐화 부위로부터 유래된 DNA 서열을 사용해서 이중 DNA 서열의 발현에 유용한 다른 유전 요소를 제공할 수도 있다. 초기 및 후기 프로모터는 둘 다 SV40 바이러스 복제 기점을 또한 포함하는 단편으로서 SV40 바이러스에서 쉽게 얻어지기 때문에 특히 유용하다 (Fiers 등, Nature 273:113, 1978). 더 작거나 더 큰 SV40 단편 또한 사용할 수

도 있다. 통상적으로, Hind III 부위로부터 SV40 복제 기점에 위치하는 BglI 부위 쪽으로 연장되는 약 250bp 서열이 포함된다.

- [0096] 다중 전사체의 발현에 사용되는 바이시스트론(Bicistronic) 발현 벡터는 이전에 기술되어 있으며 (Kim S. K. 및 Wold B. J., Cell 42:129, 1985; Kaufman 등 1991, 상기), GPT 발현 체계와 함께 사용될 수 있다. 다른 유형의 발현 벡터도 또한 유용할 것이며, 예를 들어 미국 특허 제4,634,665호 (Axel 등) 및 미국 특허 제4,656,134호(Ringold 등)에 기재된 것이다.
- [0097] 통합 부위, 예를 들어 재조합 효소 인식 부위는 POI를 암호화하는 유전자 서열에 5' 또는 3'에 위치할 수 있다. 적절한 통합 부위의 한 예는 lox p 부위이다. 적절한 통합 부위의 다른 예는 2개의 재조합 효소 인식 부위, 예를 들어 lox p 부위, lox 및 lox 5511 부위로 이루어진 군으로부터 선택된다.
- [0098] **유전자 증폭 카세트 및 그의 발현 벡터**
- [0099] 이전에 기술되었거나 당업계에 공지된 유용한 조절 요소는 또한 포유류 세포를 형질감염시키는 데 사용되는 핵산 구조체에 포함될 수 있다. 도 1은 프로모터 서열, IRES 서열, 관심 유전자 및 폴리(A) 서열을 추가로 포함하는 GPT 벡터의 작동 카세트를 예시한다.
- [0100] 본 발명의 문맥에서 발현 벡터는 염색체, 비-염색체 및 합성 핵산 벡터(적절한 발현 조절 요소 세트를 포함하는 핵산 서열)를 포함하는, 임의의 적합한 벡터일 수도 있다. 이러한 벡터의 예는 SV40의 유도체, 세균 플라스미드, 파아지 DNA, 바칼로바이러스, 효모 플라스미드, 플라스미드 및 파지 DNA의 조합에서 유도된 벡터, 및 바이러스 핵산(RNA 또는 DNA) 벡터를 포함한다. 일 실시예에서, 항체-암호화 핵산 분자는 예를 들어, (Sykes and Johnston, *Nat Biotech* 12, 355-59 (1997)에 기술된 바와 같이) 선형 발현 요소를 포함하여, 네이키드(naked) DNA 또는 RNA 벡터, (예를 들어, 미국 특허 제6,077,835호 및/또는 WO 00/70087에 기재된 바와 같은) 압축된 핵산 벡터, 또는 플라스미드 벡터, 예컨대 pBR322, pUC 19/18 또는 pUC 118/119에 포함되어 있다. 그러한 핵산 및 그의 용도는 당업계에 잘 알려져 있다 (예를 들어, 미국 특허 제5,589,466호 및 미국 특허 제5,973,972호 참조).
- [0101] 발현 벡터는 대안적으로 효모 체계에서의 발현에 적합한 벡터일 수도 있다. 효모 체계에서의 발현에 적합한 어떠한 벡터도 사용될 수도 있다. 적합한 벡터는 예를 들어, 구성적 또는 유도성 프로모터를 포함하는 벡터, 예컨대 효모 알파 인자, 알코올 산화효소 및 PGH를 포함한다 (F. Ausubel 등, ed. *Current Protocols in Molecular Biology*, Greene Publishing and Wiley InterScience New York (1987), 및 Grant 등, *Methods in Enzymol* 153, 516-544 (1987)에서 검토되었음).
- [0102] 소정의 실시예들에서, 벡터는 기술된 핵산 분자(유전자)를 포함하는 발현 벡터를 포함하여, 관심 단백질을 암호화하는 핵산 분자(또는 관심 유전자)를 포함하며, 여기서 상기 핵산 분자(유전자)는 숙주 세포에서의 발현에 적합한 발현 조절 서열에 작동 가능하게 연결된다.
- [0103] 발현 조절 서열은 관심 유전자의 전사와, 다양한 세포 체계에서의 단백질의 후속 발현을 제어하고 유도하도록 조작된다. 플라스미드는 발현가능한 관심 유전자를 예를 들어, 프로모터, 인핸서, 선택성 마커, 작동자 등과 같은 바람직한 조절 요소를 포함하는 발현 조절 서열(즉, 발현 카세트)과 결합한다. 본 발명의 발현 벡터에서, GPT 및 관심 단백질, 예컨대 항체-암호화 핵산 분자는 임의의 적합한 프로모터, 인핸서, 작동자, 억제자 단백질, 폴리(A) 종결 서열 및 기타 발현 촉진 요소를 포함할 수도 있고 이들과 결합될 수도 있다.
- [0104] 항체-암호화 뉴클레오티드 서열과 같은 관심 유전자의 발현은 당업계에 공지된 임의의 프로모터 또는 인핸서 요소의 제어 하에 놓일 수도 있다. 이러한 요소의 예로는 강한 발현 프로모터 (예, 인간 CMV IE 프로모터/인핸서 또는 CMV 메이저 IE (CMV-MIE) 프로모터, 뿐만 아니라 RSV, SV40 후기 프로모터, SL3-3, MMTV, 유비퀴틴(Ubi), 유비퀴틴 C (UbC), 및 HIV LTR 프로모터)를 포함한다.
- [0105] 일부 실시예에서, 벡터는 SV40, CMV, CMV-IE, CMV-MIE, RSV, SL3-3, MMTV, Ubi, UbC 및 HIV LTR로 구성된 군으로부터 선택된 프로모터를 포함한다.
- [0106] 본 발명의 핵산 분자는 또한 효과적인 폴리(A) 종결 서열, 대장균에서 플라스미드 산물을 위한 복제 기점, 선택성 마커로서 항생제 저항성 유전자 및/또는 간편한 클로닝 부위(예, 폴리링커)에 작동 가능하게 연결될 수도 있다. 또한 핵산은 CMV IE와 같은 구성적(constitutive) 프로모터와는 대조적으로 조절가능한 유도성(inducible) 프로모터 (유도성, 억제성, 발달적으로 조절됨)를 포함할 수도 있다 (당업자라면 그러한 용어들이 소정의 조건 하에서 실제로 유전자 발현 정도의 설명어임을 인식할 것이다).

- [0107] 선택성 마커는 당업계에 잘 알려진 요소이다. 어떤 상황에서는, GPT 이외에 추가적인 선택성 마커가 사용될 수도 있는데, 여기서 마커는 세포를 보이게 한다. 양성 또는 음성 선택을 사용할 수도 있다.
- [0108] 일부 실시예에서, 박터는 녹색 형광 단백질 (GFP), 향상된 녹색 형광 단백질 (eGFP), 시아노 형광 단백질 (CFP), 향상된 시아노 형광 단백질 (eCFP), 황색 형광 단백질 (YFP) 또는 향상된 황색 형광 단백질 (eYFP)을 암호화하는 하나 이상의 선택성 마커 유전자를 포함한다.
- [0109] 본 발명의 목적상, 진핵 세포에서의 유전자 발현은 조절 융합 단백질(regulatory fusion protein; RFP)에 의해 차례로 조절되는 작동자에 의해 제어되는 강한 프로모터를 사용하여 엄격히 조절될 수도 있다. RFP는 본질적으로 전사 차단 도메인과 그 활성을 조절하는 리간드-결합 도메인으로 구성된다. 이러한 발현 체계의 예는 US20090162901A1에 기재되어 있으며, 그 전문이 본원에 참고로 인용된다.
- [0110] 원핵 세포와 박테리오파지에서 여러 가지 작동자가 잘 특성화되어 있다 (Neidhardt, ed. *Escherichia coli and Salmonella; Cellular and Molecular Biology* 2d. Vol 2 ASM Press, Washington D.C. 1996). 이들은 LexA 펩티드를 결합하는 *대장균*의 LexA 유전자의 작동자 영역, *대장균*의 LacI 및 trpR 유전자에 의해 암호화된 억제자 단백질을 결합하는 락토오스 및 트립토판 작동자를 포함하지만 이에 한정되지는 않는다. 이들은 또한 람다 cI와 P22 arc에 의해 암호화된 억제자 단백질을 결합하는 람다 P<sub>R</sub>과 파지 P22 ant/mnt 유전자의 박테리오파지 작동자를 포함한다. 일부 실시예에서, 억제자 단백질의 전사 차단 도메인이 NotI와 같은 제한 효소일 때, 작동자는 그 효소에 대한 인식 서열이다. 당업자라면 작동자가 프로모터에 의해 전사를 제어할 수 있도록 상기 프로모터에 인접하거나 또는 3'에 위치해야 한다는 것을 인식할 것이다. 예를 들어, 본원에 참고로 인용된 미국 특허 제 5,972,650호는 tetO 서열이 TATA 박스로부터 특정 거리 내에 있어야 한다고 규정하고 있다. 특정 실시예에서, 작동자는 바람직하게는 프로모터 바로 하류에 위치한다. 다른 실시예들에서, 작동자는 프로모터의 10개 염기 쌍 내에 위치한다.
- [0111] 소정의 실시예들에서, 작동자는 tet 작동자 (tetO), NotI 인식 서열, LexA 작동자, 락토오스 작동자, 트립토판 작동자 및 Arc 작동자 (AO)로 이루어진 군으로부터 선택된다. 일부 실시예에서, 억제자 단백질은 TetR, LexA, LacI, TrpR, Arc, LambdaC1 및 GAL4로 이루어진 군으로부터 선택된다. 다른 실시예들에서, 전사 차단 도메인은 진핵 세포 억제자 단백질, 예를 들어, GAL4에서 유래한 억제자 도메인에서 유래한다.
- [0112] 예시적인 세포 발현 체계에서, 세포는 테트라사이클린 억제자 단백질(TetR)을 발현하도록 조작되고 관심 단백질은 그 활성이 TetR에 의해 조절되는 프로모터의 전사 제어 하에 놓인다. 2개의 탠덤 TetR 작동자(tetO)는 박터에서 CMV-MIE 프로모터/인핸서의 바로 하류에 위치한다. 그러한 박터에서 CMV-MIE 프로모터에 의해 유도된 관심 단백질을 암호화하는 유전자의 전사는 테트라사이클린 또는 몇몇 다른 적합한 유도자(예, 독시사이클린)의 부재 하에 TetR에 의해 차단될 수도 있다. 유도자의 존재 하에서, TetR 단백질은 tetO에 결합할 수 없으므로, 따라서 관심 단백질의 전사와 이어서 번역(발현)이 일어난다. (예, 미국 특허 제7,435,553호 참조, 이는 그 전문이 본원에 참고로 인용됨).
- [0113] 또 다른 예시적인 세포 발현 체계는 TetR-ER<sub>LD</sub>T2 융합 단백질과 같은 조절 융합 단백질을 포함하는데, 상기 융합 단백질의 전사 차단 도메인이 TetR이고 리간드-결합 도메인이 에스트로젠 수용체 리간드-결합 도메인(ER<sub>LD</sub>)이며, T2 돌연변이를 보유한다 (ER<sub>LD</sub>T2; Feil 등 (1997) *Biochem. Biophys. Res. Commun.* 237:752-757). tetO 서열을 강한 CMV-MIE 프로모터 하류와 근위 거리에 놓았을 때, CMV-MIE/tetO 프로모터 유래 관심 뉴클레오티드 서열의 전사는 타목시펜의 존재 시에 차단되었으며 타목시펜의 제거에 의해 차단해제되었다. 또 다른 예에서, 15개 아미노산 링커와 ER<sub>LD</sub>T2 (상기)로 연결된 2개의 Arc 단백질로 구성된 단일 사슬 이합체로 구성된 융합 단백질인, 융합 단백질 Arc2-ER<sub>LD</sub>T2는, Arc 작동자(AO)와 연관되며, 보다 특이적으로는 CMV-MIE 프로모터/인핸서의 바로 하류에 있는 두 개의 탠덤 arc 작동자와 연관된다. 세포주는 Arc2-ER<sub>LD</sub>T2에 의해 조절될 수도 있는데, 관심 단백질을 발현하는 세포는 CMV-MIE/ArcO2 프로모터에 의해 유도되고 타목시펜의 제거로 유도 가능하다. (예, US 20090162901A1 참조, 이는 본원에 참고로 인용됨).
- [0114] 일부 실시예에서, 본 발명의 박터는 CMV-MIE/TetO 또는 CMV-MIE/AO2 혼성 프로모터를 포함한다.
- [0115] 본 발명의 박터는 관심 유전자의 복제를 용이하게 하기 위해 재조합 기술을 위한 Cre-lox 도구를 사용할 수도 있다. Cre-lox 전략은 적어도 두 가지 구성요소를 필요로 한다: 1) 두 개의 loxP 부위 간에 재조합을 촉매하는 효소인, Cre 재조합 효소; 및 2) loxP 부위 (예, 재조합이 일어나는, 8-bp 코어 서열과, 두 개의 측면에 위치한

는 13-bp의 역방향 반복으로 구성된 특이적 34-염기쌍 bp 서열) 또는 돌연변이 *lox* 부위. (예. Araki 등 *PNAS* 92:160-4 (1995); Nagy, A. 등 *Genesis* 26:99-109 (2000); Araki 등 *Nuc Acids Res* 30(19):e103 (2002); 및 US20100291626A1 참조, 이들 전부는 본원에 참고로 인용됨). 다른 재조합 전략에서, 효모-유래 FLP 재조합 효소가 컨센서스 서열 FRT과 함께 활용될 수도 있다 (역시, 예, Dymecki, S. *PNAS* 93(12): 6191-6196 (1996) 참조).

- [0116] 또 다른 측면에서, 유전자(즉, 본 발명의 재조합 폴리펩티드를 암호화하는 뉴클레오티드 서열)가 발현 카세트의 GPT 유전자의 상류 또는 하류에 삽입되고, 선택적으로 프로모터에 작동 가능하게 연결되며, 여기서 프로모터-연결 유전자는 제1 재조합 효소 인식 부위 5' 옆에 위치하고 제2 재조합 효소 인식 부위 3' 옆에 위치한다. 이러한 재조합 효소 인식 부위는 발현 체계의 숙주 세포에서 Cre-매개 재조합을 허용한다. 일부 경우에는, 제2 프로모터-연결 유전자가 제1 유전자의 하류(3')에 있고 제2 재조합 효소 인식 부위 3' 옆에 위치한다. 또 다른 경우에, 제2 프로모터-연결 유전자는 제2 재조합 단백질 부위 5' 옆에 위치하며, 제3 재조합 단백질 인식 부위 3' 옆에 위치한다. 일부 실시예에서, 재조합 효소 인식 부위는 *loxP* 부위, *lox511* 부위, *lox2272* 부위 및 FRT 부위로부터 선택된다. 다른 실시예들에서, 재조합 효소 인식 부위는 다르다. 추가의 실시예들에서, 숙주 세포는 Cre 재조합 효소를 발현할 수 있는 유전자를 포함한다.
- [0117] 일 실시예에서, 벡터는 본 발명의 항체의 경쇄 또는 항체의 중쇄를 암호화하는 제1 유전자 및 본 발명의 항체의 경쇄 또는 항체의 중쇄를 암호화하는 제2 유전자를 포함한다.
- [0118] 일부 실시예에서, 벡터는 소포체(ER)에서의 단백질 접힘에 관여하는 유전자들의 발현을 조절하여 단백질 생산/단백질 분비를 더욱 향상시킬 수 있는 X-box-결합-단백질 1 (mXBP1) 유전자를 더 포함한다. (예, Ron D 및 Walter P. *Nat Rev Mol Cell Biol*.8:519-529 (2007) 참조).
- [0119] 임의의 세포는 본 발명의 재조합 핵산 서열을 발현하는 데 적합하다. 본 발명에서 사용된 세포는 인간 이외의 동물 세포, 인간 세포와 같은 포유류 세포, 또는 예를 들어 하이브리도마 또는 쿼드로마와 같은 세포 융합체를 포함한다. 소정의 실시예들에서, 세포는 인간, 원숭이, 햄스터, 랫트, 또는 마우스 세포이다. 다른 실시예들에서, 세포는 진핵 세포이고, 다음의 세포로부터 선택된다: CHO(예를 들어, CHO K1, DXB-11 CHO, Veggie-CHO), COS(예를 들어, COS-7), 망막 세포, Vero, CV1, 신장(예를 들어, HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L 세포, C127 세포, SP2/0, NS-0, MMT 세포, 종양 세포, 및 전술된 세포로부터 유래한 세포주. 일부 실시예에서, 세포는 하나 이상의 바이러스 유전자, 예를 들어, 바이러스 유전자를 발현하는 망막 세포(예를 들어, PER.C6® cell)를 포함한다.
- [0120] 또 다른 측면에서, 본 발명은 항체 또는 이중특이적 분자와 같은 면역 글로불린을 생산하는 트랜스펙토마(transfectoma)와 같은 재조합 포유류 숙주 세포에 관한 것이다. 이러한 숙주 세포의 예로 CHO 또는 HEK 세포와 같은 조작된 포유류 세포를 포함한다. 예를 들어, 일 실시예에서, 본 발명은 본 발명의 재조합 폴리펩티드를 포함하는 항체의 발현을 암호화하는 서열을 포함하는 세포의 계놈에 안정적으로 통합되는 핵산을 포함하는 세포를 제공한다. 또 다른 실시예에서, 본 발명은 본 발명의 재조합 폴리펩티드를 포함하는 항체의 발현을 암호화하는 서열을 포함하는 플라스미드, 코스미드, 파아지미드, 또는 선형 발현 요소와 같은 비-통합(즉, 에피솜) 핵산을 포함하는 세포를 제공한다. 다른 실시예들에서, 본 발명은 본 발명의 발현 벡터를 포함하는 플라스미드로 숙주 세포를 안정적으로 형질감염하여 생산되는 세포주를 제공한다.
- [0121] 따라서, 하나의 측면에서, 본 발명은 (a) 외생적으로 부가된 포유류 GPT 유전자를 암호화하는 재조합 폴리뉴클레오티드 및 (b) 다중-서브유닛 단백질을 암호화하는 폴리뉴클레오티드를 함유하는 세포를 제공한다. 일부 실시예에서, 상기 외생적으로 부가된 GPT 유전자는 서열번호 2의 핵산 서열에 90% 동일성이 있고, 이의 비-제한적 예는 서열번호 11 내지 17에 제공되어 있고, 상기 다중-서브유닛 단백질은 항체이다. 다른 실시예들에서, 상기 세포는 또한 외생적으로 부가된 GPT 유전자, 및 조절 요소를 함유한다. 일 실시예에서, 상기 세포는 생물약제 제조에 사용되는 CHO 세포와 같은 포유류 세포이다.
- [0122] 다른 측면에서, 본 발명은 이전 측면에서 설명된 상기 세포에서 유래한 세포주를 제공한다. “유래된”에 의해, 이는 개별 세포에서 클론에 의해 유래되고 주어진 역가에서 활성 단백질을 생산하는 능력 또는 특정 밀도로 확산되는 능력과 같은 몇몇 선택된 특성을 가진 세포 군집을 의미한다. 일부 실시예에서, 포유류 GPT 유전자를 암호화하는 재조합 폴리뉴클레오티드 및 다중-서브유닛 단백질을 암호화하는 폴리뉴클레오티드를 가진 세포에서 유래한 상기 세포주는 배지 리터당 적어도 3 그램(g/L), 적어도 5 g/L 또는 적어도 8 g/L의 역가로 상기 다중-서브유닛 단백질을 생산할 수 있다. 일부 실시예에서, 상기 세포주는 상기 GPT를 암호화하는 재조합 폴리뉴클

레오티드는 없으나 본질적으로 같은 세포에서 유래된 세포주에 의해 달성할 수 있는 상기 통합된 세포 밀도보다 적어도 30%, 적어도 50%, 적어도 60% 또는 적어도 90% 큰 통합된 세포 밀도(ICD)를 달성할 수 있다.

[0123] 상기 GOI를 증폭하는 방법을 제공한다. 상기 예시된 방법은 튜니카마이신 농도 증가를 진행 GPT 발현 체계에 적용하여, 외생적으로 부가된 포유류 GPT 유전자에 작동가능하게 연결된 GOI의 상기 유전자 사본을 증폭시킨다.

[0124] **관심 단백질**

[0125] 관심 단백질을 암호화하는 핵산 서열은 Tn 저항성 마커 유전자 및 IRES를 포함하는 세포로 편리하게 통합될 수 있고, 재조합효소 인식부위 옆에 선택적으로 존재할 수 있다. 포유류 세포에서 발현에 적합한 어떤 관심 단백질이 사용될 수 있으나, 당단백질은 특히 본 발명의 방법으로부터 이익을 얻을 것이다. 예를 들어, 상기 관심 단백질은 항체 또는 이들의 항원-결합 단편, 이중 특이적 항체 또는 이들의 단편, 키메라 항체 또는 이들의 단편, ScFv 또는 이들의 단편, Fc-태그된 단백질(예를 들어, Trap 단백질) 또는 이들의 단편, 성장 인자 또는 이들의 단편, 사이토카인 또는 이들의 단편, 또는 세포 표면 수용체의 세포외 도메인 또는 이들의 단편일 수 있다.

[0126] 아스파라긴-연결(N-연결) 글리칸을 가진 당단백질은 진행 세포에서 아주 흔하다. 이들 글리칸의 생합성 및 폴리펩티드로의 이들의 전달은 소포체(ER)에서 일어난다. N-글리칸 구조는 소포체 및 골지체에서 다수의 글리코시다아제 및 글리코실-트랜스퍼라제에 의해 더욱 변형된다. 본 발명을 사용한 단백질 생산은 면역원성 결정인자(“글리코토프(glycotopes)”)를 제거하기 위해 상기 네이티브(native) N-글리칸 구조의 일관성을 지향한다.

[0127] 본 발명의 방법을 사용하여, 재조합 단백질 루트는 유리한 특성을 나타낸다. 복제 단백질 생산 배치의 HPLC(형광 검출과 함께)는 도 7 및 8에 예시된 바와 같이 상기 당단백질이 균일한 발현 및 당질화 패턴을 가진다는 것을 보여주었다. N-글리칸 단백질 기질을 당질화하는 방법이 제공되어, 당질화를 필요로 하는 단백질 기질을 암호화하는 유전자에 작동가능하게 연결된 포유류 튜니카마이신(Tn)-저항성 유전자를 포함하는 핵산 분자를 암호화하는 포유류 숙주 세포가 제공되며; 상기 세포를 제1 농도의 Tn 존재 하에 배양하고; Tn-저항성 유전자의 적어도 하나의 사본을 발현하는 세포 군집을 분리하고; 상기 세포 군집을 Tn 농도의 증가시 배양하고; 및 상기 N-글리칸 단백질 기질을 세포 배양액으로부터 분리한다. 상기 단백질 기질의 N-글리칸 성분을 당해 기술 분야에 알려진 어떤 방법에 의해 당당류 및 올리고당의 존재에 대해 평가할 수 있다.

[0128] 글리칸-연결 단백질의 상세한 구조 분석은 상기 단백질의 기능적 특징과 상호 관련될 수 있다. 단백질 당질화를 특징으로하는 이러한 분석은 일반적으로 여러 단계를 포함한다: i) 상기 접착된 글리칸의 효소적 또는 화학적 방출; ii) 방향족 또는 지방족 아민을 이용한 환원성 아미노화에 의한 방출된 글리칸의 유도체화 또는 페메틸화; iii) 상기 글리칸의 분석. 당업자에게 알려져있는 당질화 패턴 분석의 다양한 변화. 당단백질은 특정한 양으로 다양한 부위를 차지하는 여러 종류의 글리코폼을 담지하므로 이들의 복잡성은 어떤 생산 방법에서 이의 재생산을 어렵게 만들 수 있다. 글리코폼의 종류 및 양의 일관성은 측정 가능하고 치료적 단백질 생산을 위한 바람직한 결과를 나타낸다.

[0129] **숙주 세포 및 형질감염**

[0130] 본 발명의 방법에서 사용된 포유류 숙주 세포는 통상 포유류 세포, 예를 들어, CHO 세포 및 마우스 세포를 포함하는 진행 숙주 세포이다. 일 실시예에서, 본 발명은 *Cricetulus griseus* (중국 햄스터; 서열번호 3에 기재)에서 유래된 Tn 저항성 마커 단백질을 암호화하는 핵산 서열을 포함하는 세포, 이들의 동족체 또는 변이체를 제공한다. 일부 실시예에서, 상기 세포는 상기 Tn 저항성 마커 유전자의 다수의 유전자 사본을 포함한다. 다른 실시예들에서, 본 발명은 인간(서열번호 4), 히말라야 원숭이(서열번호 5), 침팬지(서열번호 6), 개(서열번호 7), 기니아 피그(서열번호 8), 랫트(서열번호 9) 또는 마우스(서열번호 10)에서 유래된 Tn 저항성 마커 단백질을 암호화하는 핵산 서열을 제공한다.

[0131] 본 발명은 본 발명의 발현 벡터로 형질감염시킨 포유류 숙주 세포를 포함한다. 형질감염된 숙주 세포는 관심 단백질 또는 폴리펩티드를 암호화하는 서열을 포함하는 발현 벡터로 형질감염시킨 세포를 포함한다. 발현된 단백질은 일반적으로 상기 선택된 핵산 서열에 따라 배양 배지로 분비될 것이나, 상기 세포 내에 유지 또는 세포막 내에 침전될 수 있다. 다양한 포유류 세포 배양 시스템은 재조합 단백질을 발현하는데 사용될 수 있다. 적합한 포유류 숙주 세포주의 예로 Gluzman(1981) Cell 23:175에 기재된 원숭이 신장 세포주 COS-7, 및 예를 들어, CV-1/EBNA(ATCC CRL 10478), L 세포, C127, 3T3, CHO, HeLa 및 BHK 세포주를 포함하는, 적합한 벡터를 발현할 수 있는 다른 세포주를 포함한다. 특정한 선택 또는 증폭 계획을 위해 개발된 다른 세포주 역시 본 발명에서 제공된 방법 및 조성물과 함께 사용될 수 있다. 본 발명의 일 실시예에서, 상기 세포는 CHO 세포주-지정 K1(즉, CHO K1 세포)이다. 재조합 단백질의 고부피 생산이라는 목표를 달성하기 위해, 상기 숙주 세포주는 적합한 경우

에 생물 반응기 배지에 사전 적응된다.

- [0132] 여러 형질감염 프로토콜이 당해 기술 분야에 알려져 있고, Kaufman(1988) Meth. Enzymology 185:537에서 검토된다. 선택된 형질감염 프로토콜은 숙주 세포 종류 및 상기 GOI의 성질에 의존하며 일상적인 실험을 기반으로 선택될 수 있다. 이러한 프로토콜의 기본 요건은 상기 관심 단백질을 암호화하는 DNA를 적합한 숙주 세포로 도입한 다음, 상대적으로 안정하고, 발현가능한 방식으로 상기 이중 DNA를 편입하는 숙주 세포를 동정 및 단리시키는 것이다.
- [0133] 이중 DNA를 포유류 세포로 도입하는데 유용한 어떤 시약은 리포펙틴(Lipofectin™) 시약 및 리포펙타민(Lipofectamine™) 시약(Gibco BRL, Gaithersburg, Md.)을 포함한다. 이들 시약은 모두 지질-핵산 복합체(또는 리포솜)를 형성하는데 사용되는 상업적으로 이용가능한 시약으로, 배양된 세포에 적용시, 핵산을 상기 세포 내로 쉽게 흡수시킨다.
- [0134] 선택된 형질감염 프로토콜 및 본 발명에서 사용되도록 선택된 구성요소는 사용된 숙주 세포의 종류에 의존할 것이다. 당업자는 수많은 다른 프로토콜 및 숙주 세포를 알고 있으며, 사용된 세포 배양 시스템의 요건을 기준으로 목적하는 단백질의 발현을 위한 적합한 시스템을 선택할 수 있다. 또 다른 측면에 따라, 본 발명은 항체, 이중 특이적 항체, 키메라 항체, ScFv, 항원-결합 단백질, 또는 Fc 융합 단백질을 포함하는 폴리펩티드를 암호화하는 발현 벡터에 관한 것으로, 이에 제한되는 것은 아니다. 이러한 발현 벡터는 본 발명의 방법 및 조성물을 사용하여 폴리펩티드의 재조합 생산을 위해 사용될 수 있다.
- [0135] 본 발명의 다른 특징은 본 발명의 설명을 위해 제공된 바람직한 실시예에 대한 이하의 설명 과정에서 명백해질 것이나 본 발명을 제한하기 위한 것은 아니다.
- [0136] **실시예**
- [0137] 하기 실시예는 당해 기술 분야의 당업자에게 본원에 기재된 방법 및 조성물을 제조하고 사용하는 방법을 제공하기 위해 제시되나 발명자가 발명으로 간주하는 것의 범위를 제한하려는 것은 아니다. 사용된 숫자(예를 들어, 양, 온도 등)와 관련하여 정확성을 보장하기 위한 노력을 기울였으나 약간의 실험 오차와 편차에 대해서는 설명되어야 한다. 달리 언급이 없는 한, 부(parts)는 중량부, 분자량은 평균 분자량, 온도는 섭씨 온도, 압력은 대기압 또는 그 주변을 의미한다.
- [0138] **실시예 1. GPT를 발현하는 형질감염체 세포의 선택 효율.**
- [0139] 변형된 CHO K1 세포는 CHO-GPT를 함유한 플라스미드 벡터(서열번호 2), 인간 GPT(서열번호 12) 또는 히그로마이신 포스포트랜스퍼라제를 포함한 플라스미드 벡터(Hpt, 히그로마이신 저항성 유전자)로 형질감염되었다; 예를 들어, 상기 선택 마커 유전자(CHO-GPT or hpt)는 이들 각각의 벡터에서 IRES 서열을 통해 eGFP 유전자 하부에 전사적으로 연결된다. 예를 들어, 각 플라스미드는 5'에서 3' 방향으로 다음의 유전자 서열을 함유하도록 제조되었다: Lox 자리, SV40 후기 프로모터, CHO-GPT(또는 Hpt), IRES, 향상된 녹색 형광 단백질(eGFP), 및 두번째 Lox 자리. 정제된 재조합 플라스미드는 Cre 재조합효소를 발현하는 플라스미드와 함께 5'에서 3' 방향으로 전사적으로 활성 자리에 lox 자리, YFP 및 두번째 lox 자리를 함유하는 변형된 CHO 숙주 세포주로 형질감염되었다. 따라서, 상기 숙주 CHO 세포는 녹색-양성 또는 노란색-음성 세포로서 유세포 분석으로 분리될 수 있다. 상기 eGFP를 발현하는 재조합 플라스미드(상기 GPT 또는 hpt 유전자에 의해 전사적으로 조절)가 Cre 재조합효소를 발현하는 플라스미드와 함께 형질감염되면, 상기 Cre 재조합효소에 의해 매개된 재조합으로 상기 lox 자리를 가진 염색체 자리에서 GPT/eGFP 카세트의 부위-특이적 통합이 이루어지며 YFP 유전자의 치환이 일어난다(즉, 녹색-양성 세포). 상기 eGFP가 무작위적으로 통합되면, 녹색-양성 및 노란색-양성 세포 모두를 얻을 것이다.
- [0140] 세포 균집을, 표 2에 요약된 바와 같이, 0, 1 µg/ml, 2.5 µg/ml 또는 5 µg/ml 튜니카마이신(Tn) 또는 400 µg 히그로마이신(Hyg)과 함께 배양하였다. 관찰된 재조합 균집(ORPs)을 형광표지 세포 분류 분석법(FACS)에 의해 측정하였다. 세포의 각 균집을 정량하기 위해 세포를 분류하였으며, YFP가 아닌 GFP만 발현하는 세포에 대해 선택 효율을 계산하였다(도 4 또는 5).
- [0141] 선택 효율(생존한 GFP 발현 세포의 퍼센트 균집)을 Tn나 Hyg에 저항성 있는 세포 풀 내에서 비교하였다(표 2).

**표 2**

[0142]

선택 효율

Hpt 또는 GPT	Cre	선택 제제 ( $\mu\text{g/ml}$ Hyg 또는 Tn)	선택 효율 %(총 GFP+)
Hpt	+	-	1.35
Hpt	+	+ (400 Hyg)	98.8
choGPT	+	-	0.89
choGPT	+	+ (1Tn)	86.9
choGPT	+	+ (2.5 Tn)	96.1
hGPT	+	-	2.6
hGPT	+	+ (1Tn)	97
hGPT	+	+ (2.5 Tn)	96.7

[0143]

튜니카마이신 선택이 히드로마이신 선택만큼 효율적이라는 것을 관찰하였다. CHO-GPT와 인간 GPT 모두는 1  $\mu\text{g/ml}$  또는 2.5  $\mu\text{g/ml}$  튜니카마이신의 존재 하에 통합 성분의 선택에 효율적이었다. 실시예 2. 유전자 산물의 증폭.

[0144]

튜니카마이신의 농도 증가를 GPT 발현 체계에 적용하여 증분 선택이 이루어졌다. CHO K1 세포를 상술한 CHO-GPT 유전자(서열번호 2) 를 함유한 플라스미드 벡터로 형질감염하였다. 상기 플라스미드는 5' 에서 3' 방향으로 첫 번째 Lox 자리, SV40 후기 프로모터, 상기 CHO-GPT 유전자, IRES, eGFP, 및 두 번째 Lox 자리를 함유한다. 상기 관심 유전자의 게놈으로의 CRE-lox 자리-지정 통합으로 세포당 적어도 하나의 GPT 삽입을 가지는 안정된 세포의 형질감염체 풀이 된다. (앞서 확인된 바와 같이, 무작위 통합으로 인해 더 많은 통합 성분이 발생할 수 있다). CHO 세포를 초기에 1  $\mu\text{g/ml}$  튜니카마이신(Tn)의 존재 하에 배양하였다. 이후 형질감염체는 안정된 풀(세포 풀 2으로 명명)에서 선택된 다음 1  $\mu\text{g/ml}$ , 2.5  $\mu\text{g/ml}$  또는 5  $\mu\text{g/ml}$  Tn의 존재 하에 확장되었다. eGFP의 발현(다수의 사본)을 향상시킬 수 있는 세포 균집을 확인하기 위해 여러 차례의 선택과정이 수행되었다. 2.5  $\mu\text{g/ml}$  또는 5  $\mu\text{g/ml}$  Tn의 존재 하에, 상기 무작위 통합 사건이 크게 증가되었다.

[0145]

유전자 산물, 예를 들어, CHO GPT, eGFP 또는 mGapdh(표준화된 대조군) 각각의 사본 수를 표준 qPCR법을 사용하여 측정하였다. 2.5  $\mu\text{g/ml}$  Tn와 함께 더욱 배양된 1  $\mu\text{g/ml}$  Tn-저항성 풀에서 유래된 세포 내 eGFP의 사본 수는 1  $\mu\text{g/ml}$  Tn와 함께 더욱 배양된 1  $\mu\text{g/ml}$  Tn-저항성 풀에서 유래된 세포 내 eGFP의 사본 수의 적어도 2배이다. 1  $\mu\text{g/ml}$  Tn-처리된 풀을 5  $\mu\text{g/ml}$  Tn와 함께 더욱 배양하였을 때 상기 유전자 사본 수가 더욱 증가하였다. eGFP에 대한 유전자 사본 수의 증가는 CHO-GPT의 유전자 사본의 증가와 상호 관련이 있다. (도 6a 및 6b 참조)

[0146]

유전자 사본 수의 증가가 단백질 발현 증가으로 해석되는지 확인하기 위해, 평균 형광 강도(MFI)를 여러 차례의 Tn 선택 과정, 즉, 1, 2.5 또는 5  $\mu\text{g}$  Tn(예를 들어, 도 6b의 샘플 7, 8, 및 9 참조)로 처리된 GPT 및 eGFP를 발현하는 세포 풀에 대해 FACS법으로 측정하였다. 이들 세포 풀에 대한 eGFP 발현 비교를 표 3에 나타내었다.

[0147]

상기 GPT를 발현하는 세포풀은 5  $\mu\text{g}$  Tn으로 두 번째 선택에 들어가 그 결과 eGFP 생산에 있어 생산량이 1  $\mu\text{g}$  Tn 처리한 경우 비교해 2.5 배 이상, 2.5  $\mu\text{g}$  Tn 처한 경우와 비교해 1.5 배가 되었다(표 3).

표 3

eGFP 단백질 생산

[0148]

GPT 1 $\mu\text{g}$ 풀 + 두 번째 Tn( $\mu\text{g}$ ) 처리	MFI
1 $\mu\text{g}$	1098
2.5 $\mu\text{g}$	1867
5 $\mu\text{g}$	2854

[0149]

어떤 이론에 구속되지 않고, Tn 농도의 증분 증가는 통제된 방식으로 세포에 대한 선택압을 증폭시켜, 생산량이 증가된다. Tn-저항성 발현 벡터는 당질화 패턴에 대한 Tn 선택의 효과를 검사하기 위해 아래 설명되는 추가 실험에도 사용되었다.

[0150]

실시예 3. 예시적인 이량체 단백질의 발현 및 당질화 프로파일.

[0151] “트랩” 단백질(Fc 융합 단백질-1, 이하 FcFP1라 지칭)을 발현하는 CHO 세포는 GPT-함유 발현 벡터로 형질감염되었다. 상기 플라스미드는 5’에서 3’ 방향으로, Lox 자리, SV40 후기 프로모터, Tn-저항성 유전자(CHO-GPT), IRES eGFP, SV40 폴리A 및 두번째 Lox 자리를 가진다. 1 µg/mL Tn 또는 5 µg/mL Tn는 상기 GPT 선택 마커의 선택에 사용되었다. 상기 선택된 세포 풀은 무혈청 생산 배지에서 현탁 배양하여 증대시켰다. GPT 형질 감염을 FACS 분석에 의해 eGFP 발현으로 확인하였다. 선택된 풀에서 수집된 펠렛을 GPT 발현을 위한 사본 수 분석을 의뢰하였으며, 다른 농도의 튜니카마이신으로 선택된 풀에서 FcFP1의 발현량을 결정하기 위해 12일간 생산성 분석을 실시하였다.

[0152] FcFP1은 풍부한 당질화 부위를 가진 그것의 복합 당질화 패턴으로 인해 선택되었다. 당질화 프로파일을 결정하기 위해, FcF1 단백질을 발현하는 세포는 표준 프로토콜(Tn 없음) 또는 표 4에 나타난 것과 같이 Tn 처리 조건 하에 세포 배양으로 증대시킨 후 단백질을 분리 정제하였다.

**표 4**

**FcFP1 단백질 생산**

단백질	롯트 번호	처리
FcFP1 트랩	110728	없음
FcFP1 트랩	110728-1	1 µg/ml Tn
FcFP1 트랩	110728-2	5 µg/ml Tn

[0154] 상세한 글리칸 분석은 HPLC 및 형광 안트라닐산(anthranilic acid; AA) 태그(Anumula, 및 Dhume, *Glycobiology*, 8(7):685-694, 1998)를 위해 잘 알려진 방법을 기반으로 크로마토그래피를 사용하여 당 단백질의 각 롯트가 Tn이 당질화 프로파일에 부정적인 영향을 주는지 여부를 확인하기 위해 수행하였다. 생산 롯트 또한 치료적으로 허용가능한 단백질 배치(batch)를 나타내는 참조 표준과 비교하였다. 대표적인 글리칸 분석을 도 7a 내지 7d에 도시하였다. 참조 롯트와 비교해, 각 롯트는 동일한 피크 수, 상대적 형태 및 강도를 지속적으로 생산한다. 각 크로마토그램의 오버랩(도 8)은 독특하거나 특이한 피크가 발견되지 않았음을 나타낸다.

[0155] FcFP1 단백질에 대한 참조 표준 롯트와 비교해 잘 알려진 HPLC법에 의해 올리고당 프로파일링을 수행하였다. FcFP1 트랩 단백질 롯트에 대해 시알화 레벨을 측정하고 각 롯트에 대해 Z-수를 계산하였다(3개의 복제). Z-수는 롯트간의 변화의 측정을 나타낸다. 상기 Z-수는 각 피크의 상대적 형태 및 강도뿐 아니라 상대 피크 수를 고려한다. 예를 들어, 도 7a 내지 7d에서 각 OSA, 1SA, 2SA, 3SA 및 4SA 피크의 영역을 표 5와 같이 정량하였다.

**표 5**

**정량적 올리고당 분석**

단백질 롯트	복제	OSA	1SA	2SA	3SA	4SA	OS 프로파일 (Z-수)
참조 B100002M410	1	6506.43	13388.34	11268.60	5176.21	1728.15	1.53
	2	5869.80	11932.32	10159.21	4196.10	1550.09	1.51
	3	6870.18	14536.84	12090.21	5200.58	1707.74	1.51
	평균	6415.47	13285.83	11172.65	4857.63	1661.99	1.52±0.01
FcFP1 트랩 110728	1	6159.09	9394.92	7368.03	3074.66	675.48	1.34
	2	7530.49	12117.03	9589.08	2951.63	810.09	1.36
	3	5508.95	8580.56	6902.59	3794.81	630.79	1.34
	평균	6399.51	10030.84	7953.23	3074.66	705.45	1.35±0.01
FcFP1 트랩 110728-1	1	5330.22	8149.81	6539.33	2490.06	641.37	1.35
	2	5034.39	9009.42	7059.61	2698.05	812.21	1.40
	3	6222.44	10235.08	8428.04	3276.75	848.83	1.39
	평균	5529.02	9131.44	7342.33	2821.62	767.47	1.38±0.03
FcFP1 트랩 110728-2	1	6300.77	10001.93	8109.12	3000.96	790.99	1.36
	2	5999.09	9952.47	7968.58	2885.50	717.70	1.36
	3	4322.29	6176.33	5187.48	1742.26	458.52	1.32
	평균	5540.72	8710.24	7088.39	2542.91	655.74	1.35±0.02

[0157] OS = 올리고당; 0SA = 시알산 잔기 없음; 1SA = 하나의 시알산 잔기; 2SA = 두 개의 시알산 잔기; 3SA = 세 개의 시알산 잔기; 4SA = 네 개의 시알산 잔기

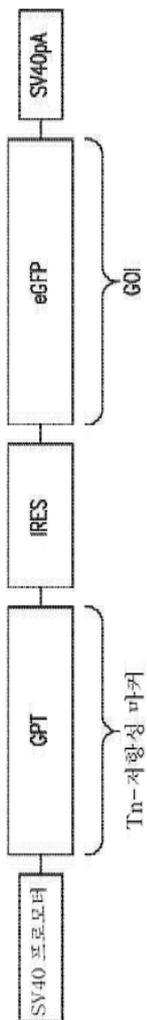
$$Z \text{ number} = \frac{(\text{Area}1\text{SA}+2*\text{Area}2\text{SA}+3*\text{Area}3\text{SA}+4*\text{Area}4\text{SA})}{(\text{Area}0\text{SA}+\text{Area}1\text{SA}+\text{Area}2\text{SA}+\text{Area}3\text{SA}+\text{Area}4\text{SA})}$$

[0158] 각 롯트에 대해 계산된 Z-수는 참조 롯트와 비교하여 허용가능한 범위 내에 있으므로, 각 단백질 롯트는 치료용 분자와 동일한 물질을 얻는 것으로 이해된다. Tn의 존재가 N-연결 당단백질의 당질화에 부정적인 영향을 미치는 것으로 알려져 있기 때문에, Tn에 의한 선택압 증가라는 조건 하에서는, 단백질 생산이 신뢰성 있고 일관성 있을 뿐만 아니라 생산적일 것이라는 예상되지 않았다.

[0159] 본 발명은 그것의 사상 또는 본질을 벗어나지 않고 다른 특정 실시예들에서 구현될 수도 있다.

**도면**

**도면1**



도면2a

```

*****
1 MWAFSELP--MPLLNLIVSLLIGFVATVTLIPAFRGHFIAARLCGDDLNKTSRQ0IPES0
1 MWAFSELP--MPLLVNLIVSLLIGFVATVTLIPAFRGHFIAARLCGDDLNKTSRQ0IPES0
1 MWAFSELP--MPLLNLIVSLLIGFVATVTLIPAFRGHFIAARLCGDDLNKTSRQ0IPES0
1 MWAFSELP--MPLLVNLVGSLLIGFVATVTLIPAFRGHFIAARLCGDDLNKTSRQ0IPES0
1 MWAFSELP--IPLLVNLVGSLLIGFVATVTLIPAFRGHFIAARLCGDDLNKTSRQ0IPES0
1 MWAFPELPLPPLLNLVGSLLIGFVATVTLIPAFRSHFIAARLCGDDLNKLSRQ0IPES0
1 MWAFPELPLPPLLNLVGSLLIGFVATVTLIPAFRSHFIAARLCGDDLNKLSRQ0IPES0
1 MWAFPELPL--PLLNLVGSLLIGFVATVTLIPAFRSHFIAARLCGDDLNKLSRQ0IPES0
*****
59 GVISGAVELIILFCFIPPEFLMCFVKEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
59 GVISGAVELIILFCFIPPEFLMCFVKEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
59 GVISGAVELIILFCFIPPEFLMCFVKEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
59 GVISGAVELIILFCFIPPEFLMCFMEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
61 GVISGAVELIILFCFIPPEFLMCFVEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
61 GVISGAVELIILFCFIPPEFLMCFVEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
59 GVICGAVELIILFCFIPPEFLMCFVEOCKAFPHHEFVALIGALLAICCMIFLGFAD0YL
*****
119 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRPILGLHLDLGLYYYMGLLAV
119 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRPILGLHLDLGLYYYMGLLAV
119 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRPILGLHLDLGLYYYMGLLAV
119 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRPILGLHLDLGLYYYMGLLAV
121 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRWILGLHLDLGLYYYMGLLAV
121 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRWILGLHLDLGLYYYMGLLAV
119 NLRWRHKLLLPTAASEPLLMVFTNEGNTTIVPKPFRWILGLHLDLGLYYYMGLLAV
*****
118 GPT_HUMAN
118 GPT_MACMU
118 GPT_PANTR
118 GPT_CANFA
118 GPT_CAVPO
120 GPT_BAT
120 GPT_MOUSE
118 GPT_CRIGR
*****
178 GPT_HUMAN
178 GPT_MACMU
178 GPT_PANTR
178 GPT_CANFA
178 GPT_CAVPO
180 GPT_BAT
180 GPT_MOUSE
178 GPT_CRIGR

```

도면2b

```

*****
179 FCTNAINI LAGINGLEAGOSLVISASLIVFNLVLELEGGDRODHVFSLYFMIPIFFFTLGL
181 FCTNAINI LAGINGLEAGOSLVISASLIVFNLVLELEGGDRODHVFSLYFMIPIFFFTLGL
181 FCTNAINI LAGINGLEAGOSLVISASLIVFNLVLELEGGDRODHVFSLYFMIPIFFFTLGL
179 FCTNAINI LAGINGLEAGOSLVISASLIVFNLVLELEGGDRODHVFSLYFMIPIFFFTLGL
*****
239 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
239 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
239 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
239 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
241 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
241 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
239 LYHHWYPSRVFVGDIFCYFAGMTFAVVGILGHFSKTMLLFFMPQVFNFLYSLPQLLHIIP
*****
299 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
299 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
299 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
299 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
301 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
301 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
299 CPHRIPRLNITKTKLEMSYKFKTKLSLFLGTFILKVAESLQLVTVHOSSETEDEFEFTEC
*****
358 GPT_HUMAN
358 GPT_MACMU
358 GPT_PANTR
358 GPT_CANFA
358 GPT_CAVPO
360 GPT_RAT
360 GPT_MOUSE
238 GPT_CRIGR

298 GPT_HUMAN
298 GPT_MACMU
298 GPT_PANTR
298 GPT_CANFA
298 GPT_CAVPO
300 GPT_RAT
300 GPT_MOUSE
298 GPT_CRIGR

358 GPT_HUMAN
358 GPT_MACMU
358 GPT_PANTR
358 GPT_CANFA
358 GPT_CAVPO
360 GPT_RAT
360 GPT_MOUSE
358 GPT_CRIGR

```

도면2c

```

*****:*** *****:*****:..** *****:*****
359 NNMTLNNLLKVLGPIHERNL TLLLLLLOILGSAITFSIRYQLVRLFYDV
359 NNMTLNLLKIFGPIHERNL TLLLLLLOILGSAITFSIRYQLVRLFYDV
359 NNMTLNLLKILGPIHERNL TLLLLLLOILGSAITFSIRYQLVRLFYDV
359 NNMTLNLLKVLGPMHERNL TLLLLLLOILGSAITFSIRYQLVRLFYDV
359 NNMTLNLLKIFGPIHERNL TLLLLLLOIVGSAITFSIRYQLVRLFYDV
361 NNMTLNLLKVFQPIHERNL TFLLLLOVLSSAVTFSIRYQLVRLFYDV
361 NNMTLNLLKVFQPIHERNL TLLLLLOVLSSAATFSIRYQLVRLFYDV
359 NNMTLNLLKIFGPIHERNL TLLLLLLOILSSAVTFSIRYQLVRLFYDV

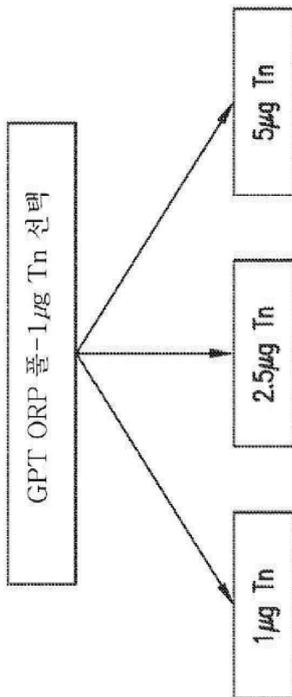
```

```

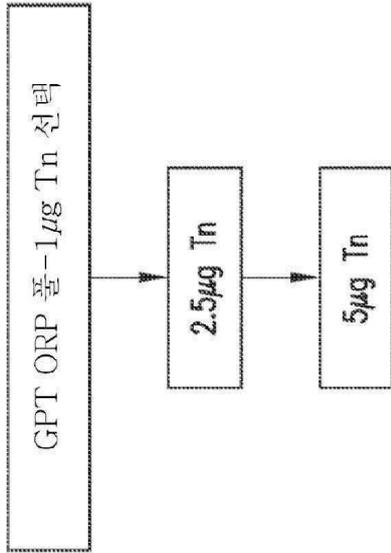
408 GPI_HUMAN
408 GPI_MACMU
408 GPI_PANTR
408 GPI_CANFA
408 GPI_CAVPO
410 GPI_RAT
410 GPI_MOUSE
408 GPI_CRIGR

```

도면3a

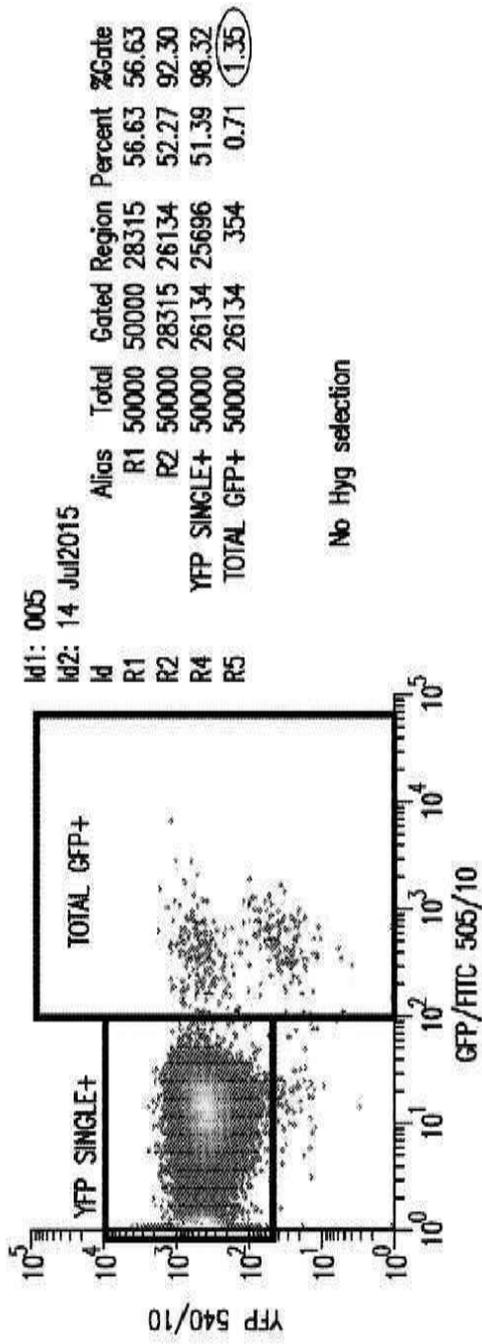


도면3b



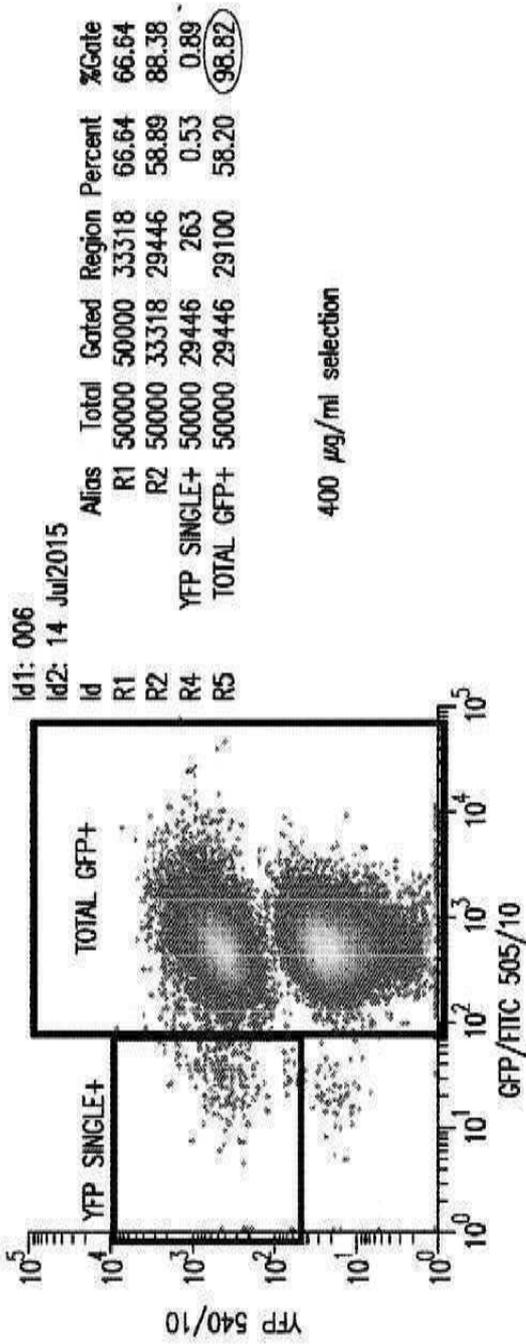
도면4a

히그로마이신 선택 효율



도면4b

히그로마이신 선택 효율



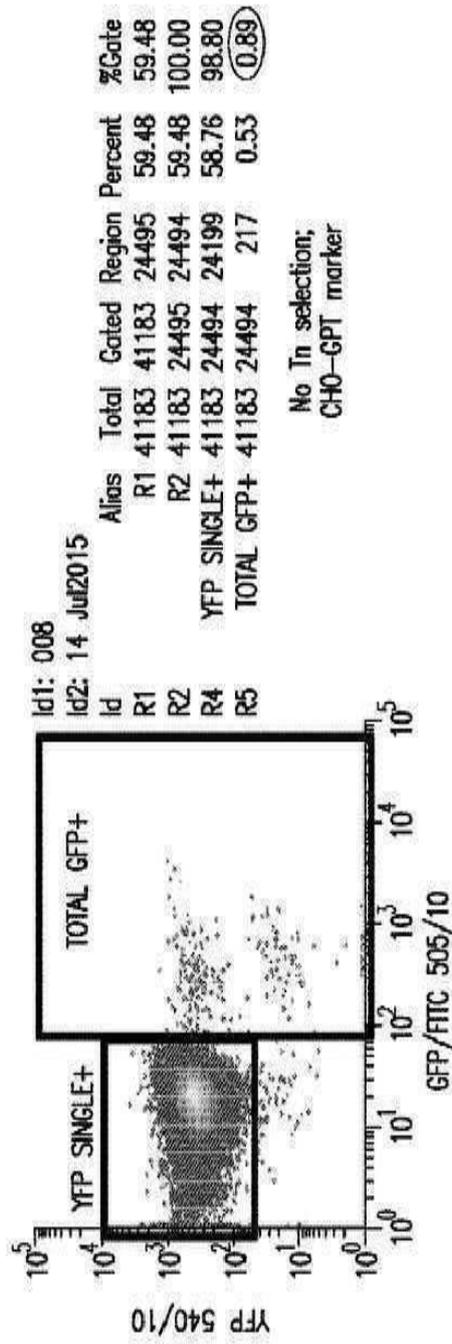
Id1: 006  
Id2: 14 Jul2015

Id	Alias	Total	Gated	Region	Percent	%Gate
R1	R1	50000	50000	33318	66.64	66.64
R2	R2	50000	33318	29446	58.89	88.38
R4	YFP SINGLE+	50000	29446	263	0.53	0.89
R5	TOTAL GFP+	50000	29446	29100	58.20	98.82

400 µg/ml selection

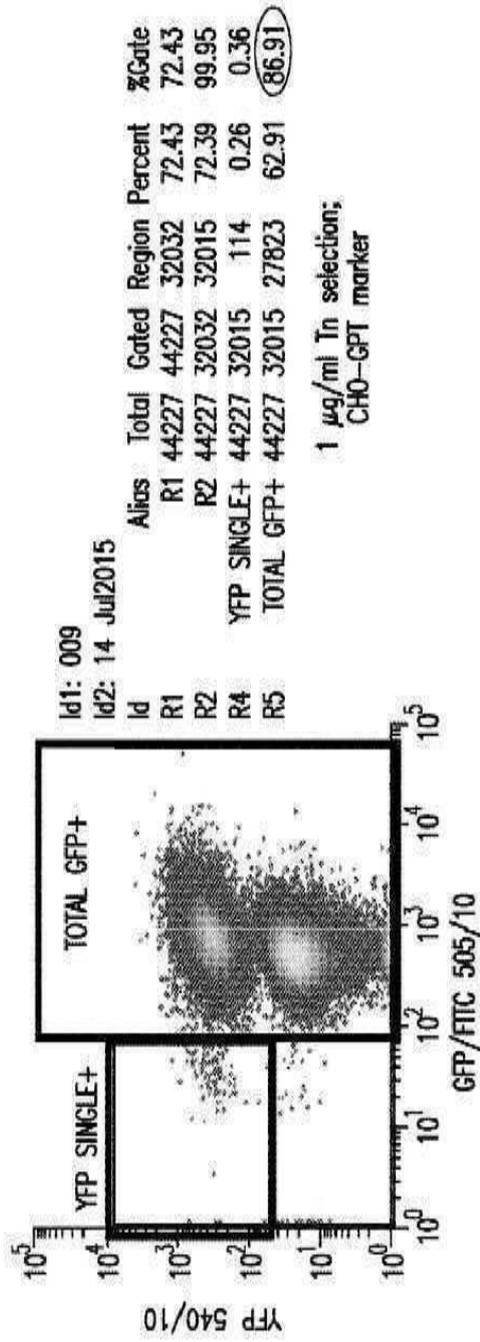
도면5a

선택성 마커로서 Tn 및 GPT  
선택 효율



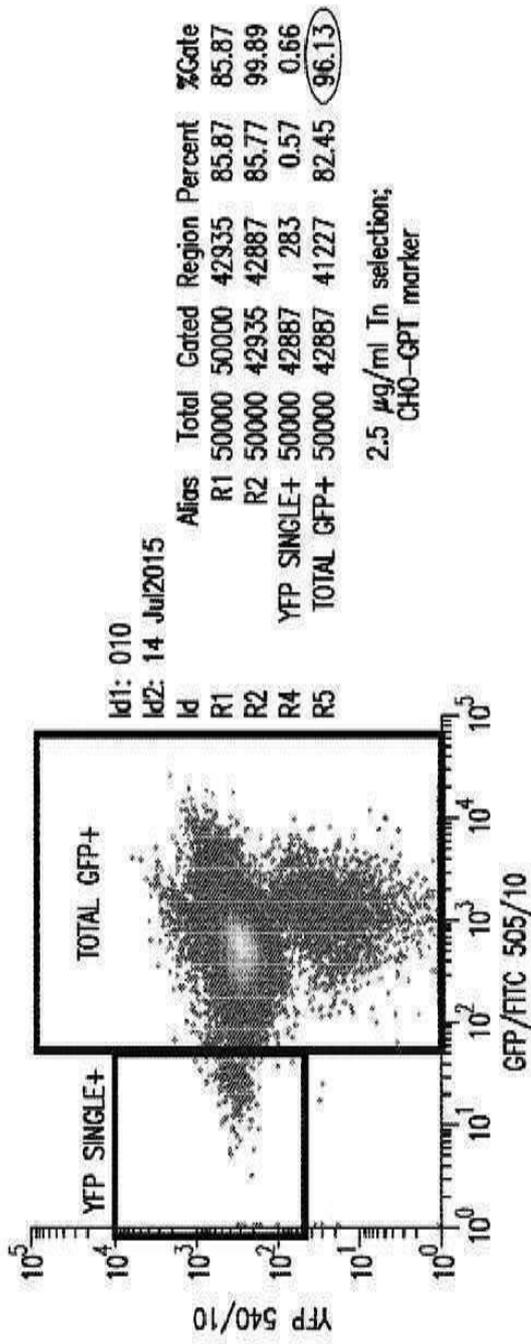
도면5b

선택성 마커로서 Tn 및 GPT  
선택 효율



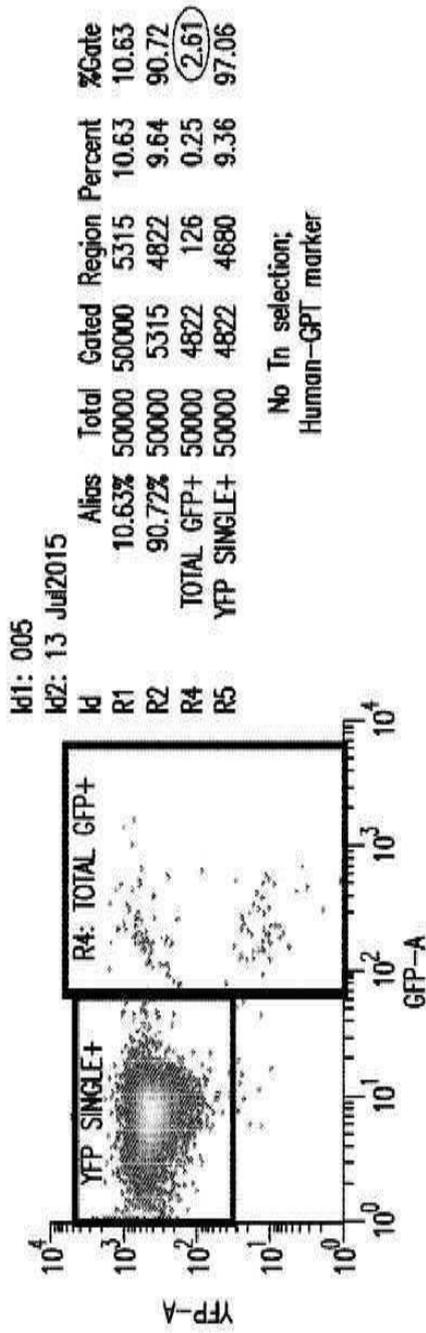
도면5c

선택성 마커로서 Tn 및 GPT  
선택 효율



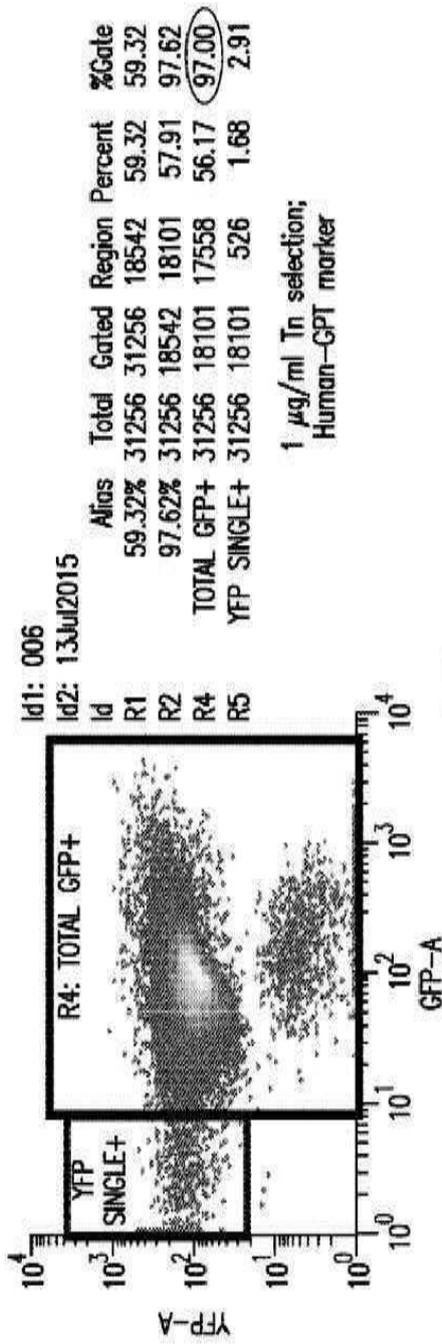
도면5d

선택성 마커로서 Tn 및 GPT  
선택 효율



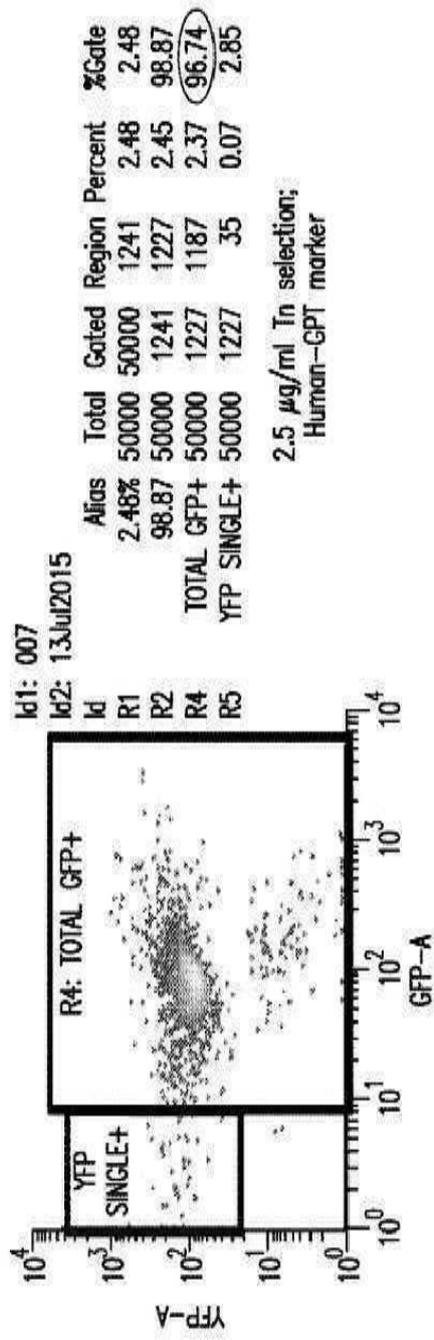
도면5e

선택성 마키로서 Tn 및 GPT  
선택 효율

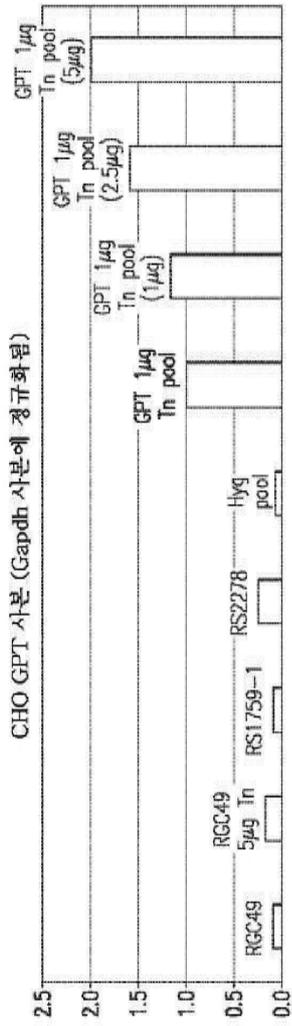


도면5f

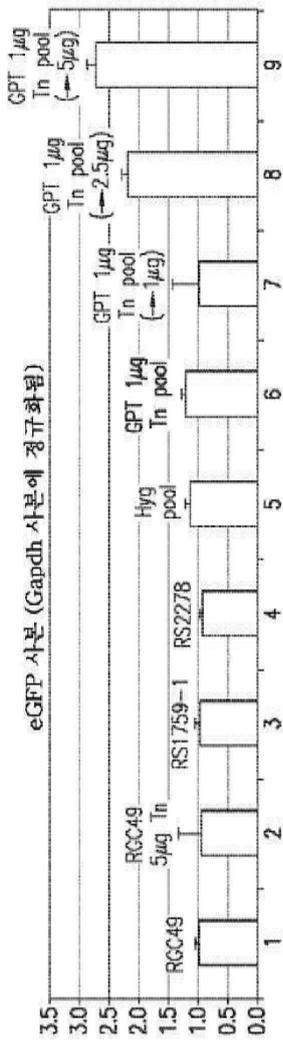
선택성 마커로서 Tn 및 GPT  
선택 효율



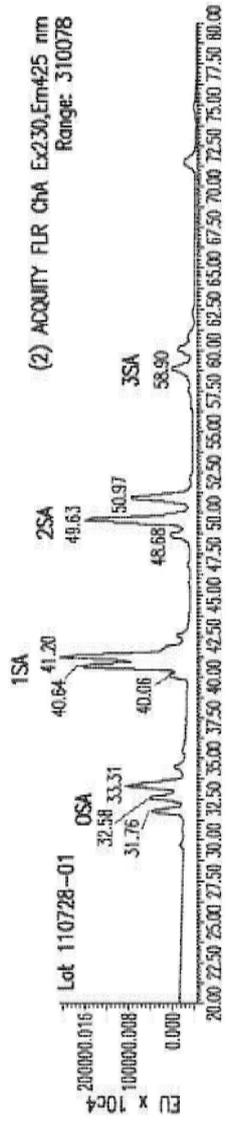
도면6a



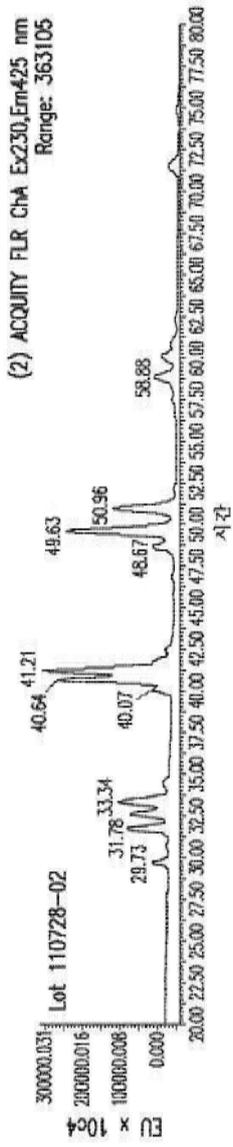
도면6b



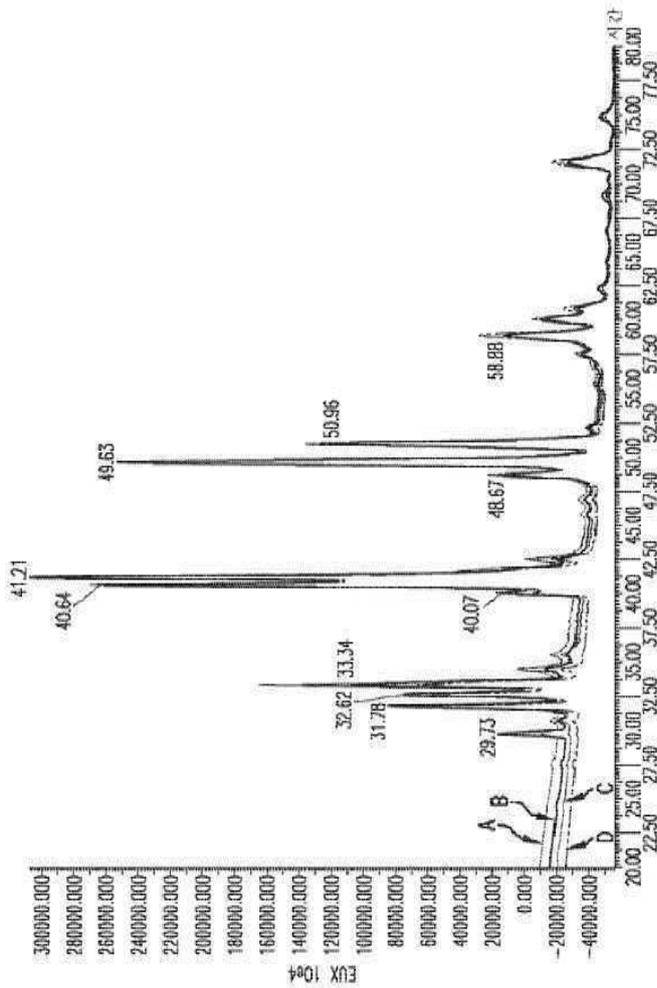
도면7c



도면7d



도면8



서열 목록

SEQUENCE LISTING

- <110> Regeneron Pharmaceuticals, Inc.
- Deshpande, Dipali
- Burakov, Darya
- Chen, Gang
- Fandl, James P.
  
- <120> EFFICIENT SELECTIVITY OF RECOMBINANT PROTEINS
- <130> 8700W0
- <150> US 62/039,416
- <151> 2014-08-19
- <160> 17
- <170> PatentIn version 3.5
- <210> 1

<211> 6964

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 1

```

aagcttatac tcgagctcta gattgggaac cgggtctct cgaattcgag atctagtta      60
aacacgcggc cgctaatcag ccataccaca ttgtagagg ttttacttgc tttaaaaaac      120

ctccccaccc tccccctgaa cctgaaacat aaaatgaatg caattgttgt tgtaacttg      180
tttattgcag cttataatgg ttacaaataa agcaatagca tcacaaatth cacaaataaa      240
gcattttttt cactgcattc tagttgttgt ttgtccaaac tcatcaatgt atcttatcat      300
gtctaccggt ataacttctg ataatgtata ctatacgaag ttagccggtg gggcccctct      360
cttcatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc      420
gtttttccat aggtccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag      480
gtggcgaaac ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctctg      540

gcgctctctt gttccgacct tgccgcttac cggatactg tccgctttc tccttcggg      600
aagcgtggcg ttttctcata gctcaagctg taggtatctc agttcgggtg aggtcgttcg      660
ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg cttatccgg      720
taactatcgt cttgagtcca acccgtaag acacgactta tcgccactgg cagcagccac      780
tggtaacagg attagcagag cgaggtatgt aggcgggtct acagagttct tgaagtggg      840
gcctaactac ggctacacta gaagaacagt atttggatc tgcgctctgc tgaagccagt      900
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg      960

tggttttttt gtttgaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc     1020
tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggtttt     1080
ggtcatgggc gcgctcata ctctgcagg catgagatta tcaaaaagga tcttcaccta     1140
gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg agtaaaactg     1200
gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg     1260
ttcatccata gttgctgac tccccgtctg gtagataact acgatcggg agggcttacc     1320
atctggcccc agtgctgcaa tgataccgag agaccacgc tcaccggctc cagatttacc     1380

agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc     1440
ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag     1500
tttgcgcaac gttgttgcca ttgctacagg catcgtgggtg tcacgctctg cgtttggtat     1560

```

ggcttcattc agctccggtt cccaacgata aaggcgagtt acatgatccc ccatgttgtg 1620  
 caaaaaagcg gttagctcct tcggctctcc gatcgttgtc agaagtaagt tggccgcagt 1680  
 gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag 1740  
 atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg 1800  
  
 accgagttgc tcttgcccgg cgtcaatacg ggataaact gcgccacata gcagaacttt 1860  
 aaaagtgtc atcattggaa aacgtttttc gggcgcaaaa ctctcaagga tcttaccgct 1920  
 gttgagatcc agttcgatgt aaccactcgc tgcaccaac tgatcttcag catcttttac 1980  
 tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggat 2040  
 aagggcgaca cggaaatgtt gaatactcat actcttctt tttcaatatt attgaagcat 2100  
 ttatcagggt tattgtctca tgagcggata catattgaa tgtatttaga aaaataaaca 2160  
 aataggggtt ccgcgcacat ttccccgaaa agtgccacct gacgtcaggt acacaacttc 2220  
  
 gtatagcata cattatacga agttatggt acaagcctag gcctcaaaa aagcctctc 2280  
 actacttctg gaatagctca gaggcagagg cggcctcggc ctctgcataa ataaaaaaaa 2340  
 ttagtcagcc atggggcgga gaatgggcgg aactgggcgg agttaggggc gggatgggcg 2400  
 gagttagggg cgggactatg gttgtgact aattgagatg catgctttgc atacttctgc 2460  
 ctgctgggga gcctggggac tttccacacc tggttgctga ctaattgaga tgcattctt 2520  
 gcatacttct gcctgctggg gagcctgggg actttccaca ccgatccac catgtgggcc 2580  
 ttccccgagt tgccgctgcc gctgctggtg aatttgttcg gctcctgct gggatttgtg 2640  
  
 gctactgtga cctcatccc tgcttccgt agccacttta tcgccgcgcg cctctgtggc 2700  
 caggacctca acaagctcag ccggcagcag atcccagaat cccaggaggt gatctgcggt 2760  
 gctgttttcc ttatcatcct cttctgctt atccctttcc ccttctgaa ctgctttgtg 2820  
 gaggagcagt gtaaggcatt cccccacat gaatttgtgg cctgatagg tgccctcctt 2880  
 gccatctgct gcatgatctt cctgggcttc gctgatgatg tactcaatct gcgctggcgc 2940  
 cataagctgc tgetgccac agctgcctct ctacctctcc tcatggttta cttactaac 3000  
 tttggcaata caaccattgt ggtaccaag ccttccgct ggattcttgg cctgcatttg 3060  
  
 gacttgggaa tcctatacta tgtctacatg ggactgcttg cgggttctg taccaatgcc 3120  
 atcaacatcc tagcaggaat taatggccta gaggtggtc agtcactagt catctctgct 3180  
 tctatcattg tctcaacct ggtagagctg gaaggtgatt atcgggatga tcatgcttt 3240  
 tccctctact tcattgatac atttttttt accaccttg gattgctata ccataactgg 3300  
 taccatcac aggtgtttgt gggagatac tctgttatt ttgctggcat gaccttggc 3360  
 gtggtgggaa tcttgggaca cttcagcaag accatgctac tcttctttat tccacaagtg 3420

ttcaatttc tctactcgct gcctcagctc cttcacgcca tccctgccc tcgacaccgc 3480

ataccagac tcaatccgaa gacgggcaaa ctggagatga gctattcaa gttcaagacc 3540

aagaacctct ctttcttggg cacctttatt ttaaaggtag cagagcgct ccagctagtg 3600

acagttcacc gaggcgagag tgaggatggt gccttctactg aatgtaacaa catgaccctc 3660

atcaacttgc tactcaaaa ctttgggccc atacatgaga gaaacctcac actgctcctg 3720

ctgcttttgc agatcctgag cagcgtgtc accttctcca ttcgatacca gettgtccga 3780

ctcttctatg atgtctgaac gcgtccccc tctccctccc cccccctaa cgttactggc 3840

cgaagccgct tggaataagg ccggtgtgcg tttgtctata tgttatttc caccatattg 3900

ccgtcttttg gcaatgtgag ggcccggaaa cctggccctg tcttcttgac gagcattcct 3960

aggggtcttt ccctctcgc caaaggaatg caaggtctgt tgaatgtcgt gaaggaagca 4020

gttctcttgg aagcttcttg aagacaaaaca acgtctgtag cgacccttg caggcagcgg 4080

aacccccac ctggcgacag gtgcctctgc ggccaaaagc cacgtgtata agatacact 4140

gcaaaggcgg cacaaccaca gtgccacgtt gtgagttgga tagttgtgga aagagtcaaa 4200

tggtctcct caagcgtatt caacaagggg ctgaaggatg cccagaaggt accccattgt 4260

atgggatctg atctggggcc tcggtgcaca tgctttacat gtgtttagtc gaggttaaaa 4320

aacgtctagg cccccgaac cacggggacg tggtttctt ttgaaaaaca cgattgctcg 4380

aatcaccatg gtgagcaagg gcgaggagct gttcacccggg gtggtgccca tcctggtcga 4440

gctggacggc gacgtaaagc gccacaagt cagcgtgtcc ggcgagggcg agggcgatgc 4500

cacctacggc aagctgacc tgaagttcat ctgcaccacc ggcaagctgc cegtgcctg 4560

gcccaccctc gtgaccacc tgacctacgg cgtgcagtgc ttcagccgct accccgacca 4620

catgaagcag cagacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac 4680

catcttcttc aaggacgacg gcaactaaa gaccgcgcc gaggtgaagt tcgagggcga 4740

cacctggtg aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct 4800

ggggcacaag ctggagtaca actacaacag ccacaacgtc tacatcatgg ccgacaagca 4860

gaagaacggc atcaaggta acttcaagat cgccacaac atcgaggacg gcagcgtgca 4920

gctcgccgac cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgccga 4980

caaccactac ctgagcacc agtccgccct gagcaaagac cccaacgaga agcgcgatca 5040

catggtcctg ctggagtctg tgaccgccgc cgggatcact ctcggcatgg acgagctgta 5100

caagtaatcg gccgctaac agccatacca cttttaga ggttttactt gctttaaaaa 5160

acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattggt gttgttaact 5220  
 tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaat ttcacaata 5280  
 aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttate 5340  
 atgtcggcgc gttgacattg attattgact agttattaat agtaatcaat tacgggggtca 5400  
 ttagttcata gcccatafat ggagttccgc gttacataac ttacggtaaa tggccccct 5460  
 ggctgaccgc ccaacgacc ccgcccattg acgtcaataa tgacgtatgt tcccatagta 5520  
 acgccaatag ggactttcca ttgacgtcaa tgggtggagt attiacggt aactgcccac 5580  
  
 ttggcagtac atcaagtga tcatatgcca agtacgcccc ctattgacgt caatgacggt 5640  
 aaatggcccc cctggcatta tgcccagtac atgacctat gggactttcc tacttggcag 5700  
 tacatctacg tattagtcac cgctattacc atggtgatgc ggttttgca gtacatcaat 5760  
 gggcgtggat agcggtttga ctcacgggga tttccaagtc tccacccat tgacgtcaat 5820  
 gggagttgt tttggcacca aaatcaacgg gactttccaa aatgctgaa caactccgcc 5880  
 ccattgacgc aaatggcgg taggcgtga cggggagg tctatataag cagagctctc 5940  
 cctatcagtg atagagatct ccctatcagt gatagagatc gtcgacgttt agtgaaccgt 6000  
  
 cagatgcct ggagacgcca tccacgctgt tttgacctcc atagaagaca cgggaccga 6060  
 tccagcctcc gcggccggga acggtgcatt ggaacgcgga tccccgtgc caagagtac 6120  
 gtaagtaccg cctatagagt ctataggecc accccttgg cttcttatgc atgctatact 6180  
 gtttttgct tgggtctat acacccccgc ttctcatgt tataggtgat ggtatagctt 6240  
 agcctatagg tgtgggttat tgaccattat tgaccactcc cctattggtg acgatacttt 6300  
 ccattactaa tccataacat ggctctttgc cacaactctc tttattggct atatgccaat 6360  
 aactgtcct tcagagactg acacggactc tgtattttta caggatgggg tctcatttat 6420  
  
 tatttaciaa ttcacatata caacaccacc gtccccagtg cccgcagttt ttattaaaca 6480  
 taacgtggga tctccacgcg aatctcgggt acgtgttccg gacatgggct cttctccggt 6540  
 agcggcggag cttctacatc cgagccctgc tccatgctt ccagcgactc atggtcgctc 6600  
 ggcagctcct tgetcctaac agtggaggcc agacttaggc acagcacgat gecccacc 6660  
 accagtgtgc cgcacaaggc cgtggcggta gggtatgtgt ctgaaaatga gctcggggag 6720  
 cgggcttga ccgtgacgc atttggaaga cttaggcag cggcagaaga agatgcaggc 6780  
 agctgagttg ttgtgtctg ataagagtca gaggtaactc ccgttgcggg gctgttaacg 6840  
  
 gttggaggca gtgtagtctg agcagtactc gttgctgccg cgcgcgccac cagacataat 6900  
 agctgacaga ctaacagact gttcctttcc atgggtcttt tctgcagtca ccgtccttga 6960  
 cacg 6964

<210> 2

<211> 1231

<212> DNA

<213> *Cricetulus griseus*

<400> 2

```

caccatgtgg gccttcccgg agttgccgct gccgctgctg gtgaatttgt tcggctcgct      60
gctgggattt gtggctactg tgaccctcat cctgccttc cgtagccact ttatcgccgc      120
gcgcctctgt ggccaggacc tcaacaagct cagccgcgag cagatcccag aatcccaggg      180

agtgatctgc ggtgctgttt tccttatcat cctcttctgc tteatccctt tccccttct      240
gaactgcttt gtggaggagc agtgaaggc attccccac catgaatttg tggccctgat      300
aggtgcctc cttgccatct gctgcatgat ctctctgggc ttcgctgatg atgtactcaa      360
tctgcctgg cgccataagc tgctgctgcc cacagctgcc tctctacctc tctcatggt      420
ttacttcaat aactttggca atacaacct tgtggtacc aagcccttc gctggattct      480
tggcctgcat ttggacttgg gaatcctata ctatgtctac atgggactgc ttgcggtgtt      540
ctgtaccaat gccatcaaca tcctagcagg aattaatggc ctagaggctg gtcagtcact      600

agtcactctt gcttctatca ttgtcttcaa cctggtagag ctggaagggtg attatcggga      660
tgatcatgct tttccctct acttcatgat accatttttt tttaccacct tgggattgct      720
ataccataac tggtagccat cacaggtgtt tgtgggagat accttctgtt attttctgg      780
catgaccttt gccgtggtgg gaatcttggg acacttcagc aagaccatgc tactcttctt      840
tattccaaa gtgttcaatt tcctctactc gctgcctcag ctccttcagc ccatccctg      900
ccctgcacac cgcatacca gactcaatcc gaagacgggc aaactggaga tgagctattc      960
caagttcaag accaagaacc tctctttctt gggcaccttt attttaaagg tagcagagcg      1020

cctccagcta gtgacagttc accgaggcga gactgaggat ggtgccttca ctgaatgtaa      1080
caacatgacc ctcatcaact tgctactcaa aatctttggg cccatacatg agagaaacct      1140
cacactgctc ctgctgcttt tgcagatcct gacgacgct gtcaccttct ccattcgata      1200
ccagcttgct cgactcttct atgatgtctg a                                1231

```

<210> 3

<211> 408

<212> PRT

<213> *Cricetulus griseus*

<400> 3

Met Trp Ala Phe Pro Glu Leu Pro Leu Pro Leu Leu Val Asn Leu Phe

1                    5                    10                    15  
 Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe  
                          20                    25                    30  
 Arg Ser His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys  
                          35                    40                    45  
 Leu Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Cys Gly Ala  
                          50                    55                    60  
 Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn  
 65                    70                    75                    80  
  
 Cys Phe Val Glu Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val  
                          85                    90                    95  
 Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly  
                          100                    105                    110  
 Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu  
                          115                    120                    125  
 Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe  
                          130                    135                    140  
  
 Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Trp Ile Leu Gly  
 145                    150                    155                    160  
 Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu  
                          165                    170                    175  
 Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly  
                          180                    185                    190  
 Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe  
                          195                    200                    205  
  
 Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Val Phe Ser  
                          210                    215                    220  
 Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr  
 225                    230                    235                    240  
 His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe Cys Tyr  
                          245                    250                    255

Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser  
 260 265 270

Lys Thr Met Leu Leu Phe Phe Ile Pro Gln Val Phe Asn Phe Leu Tyr  
 275 280 285

Ser Leu Pro Gln Leu Leu His Ala Ile Pro Cys Pro Arg His Arg Ile  
 290 295 300

Pro Arg Leu Asn Pro Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys  
 305 310 315 320

Phe Lys Thr Lys Asn Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val  
 325 330 335

Ala Glu Arg Leu Gln Leu Val Thr Val His Arg Gly Glu Ser Glu Asp  
 340 345 350

Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu  
 355 360 365

Lys Ile Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu  
 370 375 380

Leu Leu Gln Ile Leu Ser Ser Ala Val Thr Phe Ser Ile Arg Tyr Gln  
 385 390 395 400

Leu Val Arg Leu Phe Tyr Asp Val  
 405

<210> 4

<211> 408

<212> PRT

<213> Homo sapiens

<400> 4

Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu Leu Ile Asn Leu Ile  
 1 5 10 15

Val Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe  
 20 25 30

Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys  
 35 40 45

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

290 295 300

Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys

305 310 315 320

Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val

325 330 335

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

340 345 350

Gly Glu Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu

355 360 365

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

370 375 380

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

385 390 395 400

Leu Val Arg Leu Phe Tyr Asp Val

405

<210> 5

<211> 408

<212> PRT

<213> *Macaca mulatta*

<400> 5

Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu Leu Val Asn Leu Ile

1 5 10 15

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

20 25 30

Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys

35 40 45

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

50 55 60

Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn

65 70 75 80

Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val

	85	90	95
Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly			
	100	105	110
Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu			
	115	120	125
Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe			
	130	135	140
Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly			
145	150	155	160
Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu			
	165	170	175
Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly			
	180	185	190
Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe			
	195	200	205
Asn Leu Val Glu Leu Glu Gly Asp Cys Arg Asp Asp His Val Phe Ser			
	210	215	220
Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr			
225	230	235	240
His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe Cys Tyr			
	245	250	255
Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser			
	260	265	270
Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr			
	275	280	285
Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile			
	290	295	300
Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys			
305	310	315	320
Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val			
	325	330	335

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

Leu Val Arg Leu Phe Tyr Asp Val  
 405

<210> 6  
 <211> 408  
 <212> PRT  
 <213> Pan troglodytes  
 <400> 6

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

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

Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu  
 115 120 125  
 Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe

130                      135                      140  
 Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly  
 145                      150                      155                      160  
 Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu  
                          165                      170                      175  
  
 Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly  
                          180                      185                      190  
 Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe  
                          195                      200                      205  
 Asn Leu Val Glu Leu Glu Gly Asp Cys Arg Asp Asp His Val Phe Ser  
                          210                      215                      220  
 Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr  
 225                      230                      235                      240  
  
 His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe Cys Tyr  
                          245                      250                      255  
 Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser  
                          260                      265                      270  
 Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr  
                          275                      280                      285  
 Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile  
                          290                      295                      300  
  
 Pro Arg Leu Asn Ile Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys  
 305                      310                      315                      320  
 Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu Lys Val  
                          325                      330                      335  
 Ala Glu Ser Leu Gln Leu Val Thr Val His Gln Ser Glu Thr Glu Asp  
                          340                      345                      350  
 Gly Glu Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu Leu Leu  
                          355                      360                      365  
  
 Lys Ile Leu Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu Leu Leu  
                          370                      375                      380

Leu Leu Gln Ile Leu Gly Ser Ala Ile Thr Phe Ser Ile Arg Tyr Gln  
 385                      390                      395                      400

Leu Val Arg Leu Phe Tyr Asp Val  
 405

<210> 7

<211> 408

<212> PRT

<213> Canis familiaris

<400> 7

Met Trp Ala Phe Pro Glu Leu Pro Met Pro Leu Leu Val Asn Leu Val  
 1                      5                      10                      15

Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe  
 20                      25                      30

Arg Gly His Phe Ile Ala Ala His Leu Cys Gly Gln Asp Leu Asn Lys  
 35                      40                      45

Thr Gly Arg Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser Gly Ala  
 50                      55                      60

Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn  
 65                      70                      75                      80

Cys Phe Met Glu Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val  
 85                      90                      95

Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly  
 100                      105                      110

Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu  
 115                      120                      125

Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr Asn Phe  
 130                      135                      140

Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Ile Leu Gly  
 145                      150                      155                      160

Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu  
 165                      170                      175

Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly

180 185 190  
 Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe  
 195 200 205

Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Val Phe Ser  
 210 215 220

Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr  
 225 230 235 240

His Asn Trp Tyr Pro Ser Gln Val Phe Val Gly Asp Thr Phe Cys Tyr  
 245 250 255

Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His Phe Ser  
 260 265 270

Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe Leu Tyr  
 275 280 285

Ser Leu Pro Gln Leu Leu His Ile Ile Pro Cys Pro Arg His Arg Ile  
 290 295 300

Pro Arg Leu Asn Thr Lys Thr Gly Lys Leu Glu Met Ser Tyr Ser Lys  
 305 310 315 320

Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Asn Phe Ile Leu Lys Val  
 325 330 335

Ala Ala Ser Leu Gln Leu Val Thr Val His Gln Ser Glu Asn Glu Asp  
 340 345 350

Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Leu Asn Leu Leu Leu  
 355 360 365

Lys Val Leu Gly Pro Met His Glu Arg Asn Leu Thr Leu Leu Leu Leu  
 370 375 380

Leu Leu Gln Ile Leu Gly Ser Ala Val Thr Phe Ser Ile Arg Tyr Gln  
 385 390 395 400

Leu Val Arg Leu Phe Tyr Asp Val  
 405

<210> 8

<211> 408

<212> PRT

<213> *Cavia porcellus*

<400> 8

Met Trp Ala Phe Ser Glu Val Pro Ile Pro Leu Leu Val Asn Leu Ile

1                    5                    10                    15

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

                  20                    25                    30

Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys

                  35                    40                    45

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

50                    55                    60

Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe Leu Asn

65                    70                    75                    80

Cys Phe Val Lys Glu Gln Cys Lys Ala Phe Pro His His Glu Phe Val

                  85                    90                    95

Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe Leu Gly

                  100                    105                    110

Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu Leu Leu

115                    120                    125

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

130                    135                    140

Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Pro Val Leu Gly

145                    150                    155                    160

Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly Leu Leu

                  165                    170                    175

Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile Asn Gly

180                    185                    190

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

195                    200                    205

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

210                    215                    220

Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu Leu Tyr





Phe Ser Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe  
 275 280 285  
 Leu Tyr Ser Leu Pro Gln Leu Phe Gln Ile Ile Pro Cys Pro Arg His  
 290 295 300  
 Arg Met Pro Arg Leu Asn Thr Lys Thr Gly Lys Leu Glu Met Ser Tyr  
 305 310 315 320  
 Ser Lys Phe Lys Thr Lys Ser Leu Ser Phe Leu Gly Thr Phe Ile Leu  
 325 330 335  
  
 Lys Val Ala Glu Ser Leu Arg Leu Val Thr Val His Arg Gly Glu Ser  
 340 345 350  
 Glu Asp Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu  
 355 360 365  
 Leu Leu Lys Val Phe Gly Pro Thr His Glu Arg Asn Leu Thr Leu Phe  
 370 375 380  
 Leu Leu Leu Leu Gln Val Leu Ser Ser Ala Val Thr Phe Ser Ile Arg  
 385 390 395 400  
  
 Tyr Gln Leu Val Arg Leu Phe Tyr Asp Val  
 405 410  
  
 <210> 10  
 <211> 410  
 <212> PRT  
 <213> Mus musculus  
 <400> 10  
  
 Met Trp Ala Phe Pro Glu Leu Pro Leu Pro Leu Pro Leu Leu Val Asn  
 1 5 10 15  
 Leu Ile Gly Ser Leu Leu Gly Phe Val Ala Thr Val Thr Leu Ile Pro  
 20 25 30  
 Ala Phe Arg Ser His Phe Ile Ala Ala Arg Leu Cys Gly Gln Asp Leu  
 35 40 45  
  
 Asn Lys Leu Ser Gln Gln Gln Ile Pro Glu Ser Gln Gly Val Ile Ser  
 50 55 60  
 Gly Ala Val Phe Leu Ile Ile Leu Phe Cys Phe Ile Pro Phe Pro Phe

65                      70                      75                      80  
 Leu Asn Cys Phe Val Glu Glu Gln Cys Lys Ala Phe Pro His His Glu  
    85                      90                      95  
 Phe Val Ala Leu Ile Gly Ala Leu Leu Ala Ile Cys Cys Met Ile Phe  
    100                      105                      110  
  
 Leu Gly Phe Ala Asp Asp Val Leu Asn Leu Arg Trp Arg His Lys Leu  
    115                      120                      125  
 Leu Leu Pro Thr Ala Ala Ser Leu Pro Leu Leu Met Val Tyr Phe Thr  
    130                      135                      140  
 Asn Phe Gly Asn Thr Thr Ile Val Val Pro Lys Pro Phe Arg Trp Ile  
 145                      150                      155                      160  
 Leu Gly Leu His Leu Asp Leu Gly Ile Leu Tyr Tyr Val Tyr Met Gly  
    165                      170                      175  
  
 Leu Leu Ala Val Phe Cys Thr Asn Ala Ile Asn Ile Leu Ala Gly Ile  
    180                      185                      190  
 Asn Gly Leu Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile  
    195                      200                      205  
 Val Phe Asn Leu Val Glu Leu Glu Gly Asp Tyr Arg Asp Asp His Ile  
    210                      215                      220  
 Phe Ser Leu Tyr Phe Met Ile Pro Phe Phe Phe Thr Thr Leu Gly Leu  
 225                      230                      235                      240  
  
 Leu Tyr His Asn Trp Tyr Pro Ser Arg Val Phe Val Gly Asp Thr Phe  
    245                      250                      255  
 Cys Tyr Phe Ala Gly Met Thr Phe Ala Val Val Gly Ile Leu Gly His  
    260                      265                      270  
 Phe Ser Lys Thr Met Leu Leu Phe Phe Met Pro Gln Val Phe Asn Phe  
    275                      280                      285  
 Leu Tyr Ser Leu Pro Gln Leu Phe His Ile Ile Pro Cys Pro Arg His  
    290                      295                      300  
  
 Arg Met Pro Arg Leu Asn Ala Lys Thr Gly Lys Leu Glu Met Ser Tyr  
 305                      310                      315                      320

Ser Lys Phe Lys Thr Lys Asn Leu Ser Phe Leu Gly Thr Phe Ile Leu  
 325 330 335  
 Lys Val Ala Glu Asn Leu Arg Leu Val Thr Val His Gln Gly Glu Ser  
 340 345 350  
 Glu Asp Gly Ala Phe Thr Glu Cys Asn Asn Met Thr Leu Ile Asn Leu  
 355 360 365

Leu Leu Lys Val Phe Gly Pro Ile His Glu Arg Asn Leu Thr Leu Leu  
 370 375 380  
 Leu Leu Leu Leu Gln Val Leu Ser Ser Ala Ala Thr Phe Ser Ile Arg  
 385 390 395 400  
 Tyr Gln Leu Val Arg Leu Phe Tyr Asp Val  
 405 410

<210> 11

<211> 1920

<212> DNA

<213> Mus musculus

<400> 11

gttgcttct aagagcttct tgctggtcag gagggagggt caggctctag cgtcctagct 60  
 gggttttggt cccgctggcg ccggaatcct ctgccgggtg ggagccgcac tgccgctgc 120  
  
 cgaggccacg ggattgttcc tgcttacca gttagctgag taggcggcgg gcggcgcc 180  
 accggagggt caccatgtgg gccttcccgg agttgccct gccgtgccg ctgctggtga 240  
 atttgatcgg ctgctgttg ggattcgtgg ctacagtcac cctcatcct gccttcgta 300  
 gccactttat cgccgcgcgc ctctgtggcc aggacctcaa caagctcagc cagcagcaga 360  
 tcccagatc ccaggagtg atcagcgtg ctgttttct tatcctctc ttctgcttca 420  
 tcccttccc cttctgaac tgcttcgtgg aggagcagtg taaggcatt ccccaccatg 480  
 aatttgggc cctaataagt gccctccttg ccatctgctg catgatcttc ctggggttg 540  
  
 ctgatgatgt cctcaatctc cgctggcgcc acaagctgct gctgccaca gctgcctcac 600  
 tacctctct catggtctac ttacaaaact ttggcaatac aaccatcgtg gtgcccaagc 660  
 ccttccgctg gattctgggc ctgcatttgg acttggggat cctgtactac gtctacatgg 720  
 ggctgcttg agtgttctgt accaatgcca tcaacatcct ggccggcatt aatggcctag 780  
 aggccggtca gtcactagtc atctctgctt ctatcattgt cttcaacctg gtggaactgg 840  
 aagtgatta tcgagatgat catatcttt cctttactt catgatacca tttttctta 900

ccaccttggg actgctttac cacaactggt acccgteccg cgtgtttgtg ggagacacct 960

tctgttactt tgcgggcatg acttttgccg tgggtgggat cttgggacac ttcagcaaga 1020

ccatgetgct ctcttttatg ccacaagtat tcaatttctt ctaactactg cetcagctct 1080

tccatatcat ccctgcctt cgacaccgga tgcccagact caacgcaaag acaggcaaac 1140

tggaaatgag ctattccaag ttcaagacca agaacctctc tttcttgggc acctttattt 1200

taaaggtage agagaacctc cggttagtga cagttcacca aggtgagagt gaggacggtg 1260

ccttactga gtgtaacaac atgacctca tcaacttctt acttaagtc tttgggccta 1320

tacatgagag aaacctcacc ctgctcctgc tgcctcctgca ggtcctaagc agcgccgcca 1380

ccttctccat tcgttaccaa ctgcgccgac tcttctatga tgtctgagct ccctgacagc 1440

tgccctttac ctcacagtct ccattggacc tcagccagga ccagcctctg tctggtccga 1500

gatgaccctc tgggtccaggc ctgctgaca cttttgttct cagcttctgc catctgtgac 1560

tactgatatc ctggatggac accttctgg acttgaagtc cgctagtgg accttgccta 1620

gggctttcat ctgacctgc cctcccttc tgcctcctt gcagcctcac caggtgggct 1680

tgtagcctct attatgcaaa tattegtage tcagctttca gagcgctaac tetaaaggaa 1740

ttcacctgag ccttgagaga gaacctgggc tagggctaga gttagggcta catactcaa 1800

ggtgacctca catttgacta tcaaatgaag tgttgtgatt gggaagcgtg gaggcagggc 1860

catgtgctca gaacggtgac aataaaggac tgccttttac ttgttaaaaa aaaaaaaaaa 1920

<210> 12

<211> 2150

<212> DNA

<213> Homo sapiens

<400> 12

aagtatccgt tcttggctgc ctttctttaa ttgcgtttcc agtactctct cgggtattct 60

actcttgaac ataggatgaa atttggaatc acacttctct tccacttcca tccccacct 120

ctaatgccca tattaaaaat ggcggccgcc gccttccgca gtaatggttg ttcagegaac 180

aagatccggg cggaaacagt agataggcgg gtgcagcggg gcagaacata ggttgcctta 240

gagaggttcc ccggtgtccc gacggcggct caagtcagag ttgctgggtt ttgctcagat 300

tgggtgggga agagcctgcc tgtggggagc ggccactcca tactgctgag gcctcaggac 360

tgctgctcag ctgccccgtt acctgaagag gcggcggagc cgggccccctg accggtcacc 420

atgtgggctt tctcggaatt gcccatgccg ctgctgatca atttgatcgt ctgctgctg 480

ggatttgtgg ccacagtcac cctcatcccc gccttccggg gccacttcat tgctgcgcgc 540

ctctgtggtc aggacctcaa caaaaccagc cgacagcaga tcccagaatc ccagggagtg 600  
atcagcgggtg ctgttttctc tatcatctc tctgtctca tccctttccc cttcctgaac 660

tgctttgtga aggagcagtg taaggcattc ccccaccatg aatttgtggc cetgataggt 720  
gccctccttg ccatctgctg catgatcttc ctgggctttg cggatgatgt actgaatctg 780  
cgctggcgcc ataagctgct gctacctaca gctgcctcac tacctctcct catggtctat 840  
ttaccaact ttggcaacac gaccattgtg gtgcccaagc ccttccgcc gatacttggc 900  
ctgcatctgg acttgggaat cctgtactat gtctacatgg ggctgctggc agtgttctgt 960  
accaatgcca tcaatatcct agcaggaatt aacggcctag aggctggcca gtcactagtc 1020  
atctctgctt ccatcattgt cttcaacctg gtagagtgg aagtgattg tcgggatgat 1080

catgtctttt cctctactt catgatacc tttttttca ccactttggg attgctctac 1140  
cacaactggt acccatcac ggtgtttgtg ggagatacct tctgttactt tgctggcatg 1200  
acctttgccg tgggtggcat cttgggacac ttcagcaaga ccatgctact attcttcag 1260  
ccccaggtgt tcaacttct ctactcactg cctcagctcc tgcatatcat cccctgcct 1320  
cgccaccga taccagact caatatcaag acaggcaaac tggagatgag ctattccaag 1380  
ttcaagacca agagcctctc tttcttgggc acctttattt taaaggtggc agagagcctc 1440  
cagctggtga cagtacacca gagtgagact gaagatggtg aattcactga atgtaacaac 1500

atgaccctca tcaacttgct acttaaagtc cttgggcca tacatgagag aaacctcaca 1560  
ttgctcctgc tgcctgca gatcctgggc agtccatca ccttctccat tcgatatcag 1620  
ctcgttcgac tcttctatga tgtctgagtc ccttgatcat tgcctttac ctacagctct 1680  
ctaggattcc tgactcagc tgacctctc ccttgggcc agactgcctc cttgccagg 1740  
cctctctcac tctcact cctccagatt ttgttctcag cattttcctt tctctgtgat 1800  
cattggcate ctggcgctt cttgcctctc gctgactact gattggattt tacctatggc 1860  
tttctgcaac ttgctactc ctcctctcc atccatctt tgcagcctca tagggtggga 1920

tacagcagct tttttgcag ttatccacac tcacattca gactcctgac tctcaaggaa 1980  
ccactggttt ttggataga acttgggcca gggctaggaa cacaggctcc acggtgacat 2040  
gtcatttgat tgtaaattaa gtgttctgat tagtaagaac taagcagggg gccacatgct 2100  
ctcaatggag acaataaagt gttgtctttt tcttattgtt taaaaaaaaa 2150

<210> 13

<211> 1840

<212> DNA

<213> Rattus norvegicus

<400> 13

gggggctggc gccggaatcc tctgagtga gggagctgca ctgctggctg ccgaggcctc 60  
 tggtttgttc ctggcttacc aagttagctg agtaggcggc ggagcggcgg cccccggagg 120  
  
 gtcactatgt gggccttccc ggagttacct ctgccgtgc cgctgttggg gaatttgatc 180  
 ggatcgtgt tgggatttgg ggctaccgtc accctcatcc ctgccctccg tagccacttt 240  
 atcgccgcgc gtctctgtgg ccaggacctc aacaagetca gccggcagca gatcccagag 300  
 tcccaggagg tgatcagcgg tgctgttttc cttatcatcc tcttctgctt catccctttc 360  
 cctttctga actgctttgt ggaggagcag tgtaaggcat tccccacca tgaatttgtg 420  
 gccctgatag gtgccctcct tgctatctgc tgcatgatct tcttgggttt tgctgatgac 480  
 gtactcaatc tgcgatggcg tcataagctg ctgctcccca cagctgcctc actacctctc 540  
  
 ctcatggtct acttcaetaa ctttggcaat acaaccattg tggtgccgaa gcccttccgc 600  
 tggattcttg gcctgcattt ggatttgggc atcctgtact atgtctacat gggactgctt 660  
 gcagtgttct gtaccaatgc catcaacatc ctagcgggaa ttaatggcct agaggctggt 720  
 caatcactag teatctctgc ttctattatt gtcttcaacc tgggtggagct ggaagtgat 780  
 tatcgggacg atcatgtctt ttccctctac ttcatgatgc ctttttttt taccaccttg 840  
 ggattgctgt accataactg gtaccctctc cagggttttg tgggagacac cttctgttat 900  
 tttgctggca tgacctttgc cgtggtggga atcttgggac acttcagcaa gaccatgctg 960  
  
 ctcttcttta tgccacaagt attcaatttc ctctactcac tgccctcagct ctttcagatc 1020  
 atccctgcc ctgcacaccg tatgcccaaga ctcaatacga agacaggcaa actggagatg 1080  
 agctattcca agttcaagac caagagcctc tctttcttgg gcacgtttat tttaaaggta 1140  
 gcagagagcc tccggctggt gacagttcac cgaggggaga gtgaggatgg tgccttcact 1200  
 gagtgaaca acatgacct catcaacttg ctacttaag tctttgggcc tacacatgag 1260  
 agaaacctca cactgttctt gctgctcctg caggttctga gcagcgtgt caccttctcc 1320  
 attcgttacc agctcgtccg actcttctat gatgtctgag ctccctgacg actgcccttt 1380  
  
 accacacagt ctccattgga cctcagccag gaccacctc tgtccgctcc gaccgccttc 1440  
 tggccaggc tcagcttctg ccgtcatctg tgactactga catcctggat ggactcctta 1500  
 gtggacttga cgtccactag ttggacttgc ctatgcttcc ttgagtttgc tactccctcc 1560  
 ctttctgcag cctcaccagg tgggcctgia gcatctttta tgcaaatatt catggctcaa 1620  
 ctttcagaac cctaactcta aaggaatccc ctgggccttg agagagaacc tgggctaggg 1680  
 ctagagttag ggcaacatac tccaaggtaa cctcacatct gactatcaaa ttaagtgttc 1740

tgattaggaa gagcagaggc agggccatgt gctcagaatg gtgacaataa aggattgcct 1800  
  
 ttacttgcc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1840  
 <210> 14  
 <211> 5424  
 <212> DNA  
 <213> Macaca mulatta  
 <400> 14  
 tattaanaat ggcggctgcc gccctccgca gtaatagtgt ttcagcgaat aaaatccggg 60  
 cggaaacagt aggtaggctg gttcagtggg ggcagaaact aggttgcctt agagaggttc 120  
 ttcgaggctc cgagggcggc tcaagtcaga gttgttggat tctgctcaga ttggtgtggg 180  
 aagagcctgc ctgtggggag cggccactcc atactgctga ggcctcagga ctgctgctca 240  
 gcttgccagt tacctgaaga ggcggcggag ccgggccctt gaccggctac catgtgggcc 300  
  
 ttctcggaat tgcccatgcc gctgctggtc aatttgatcg tctcgtgct gggatttgtg 360  
 gccacagtca cctcatccc agccttccgg gccacttca ttgctgcgcg cctctgtggt 420  
 caggactca aaaaaccag ccgacaacag atgtgagcag cggcacacgg ttccgggcag 480  
 ggggcaaggg ctaaggaagg agtggttagg gcaggggcgg ggaccggggt gtttgaccac 540  
 acgtgaaaac tcagaactaa cccaggcagc ctggaactcg gagaggatgat gagcagaact 600  
 tattgcatt ggggaaagga tgggtaggga accttgggta tatcaggac tctagcagtg 660  
 gtgctttcct cctcggccc cctcaccac ttcccaaat aaaaaccag gaatgagaag 720  
  
 accgctttgg gtiattgtaa cacctgcaact agtgagtac cacacccctt ttcctctttc 780  
 ccctgcccc ctgctgctg gccacagcc cagaatcca gggagtgatc agcgggtctg 840  
 ttttcttat catcctcttc tgcttcatcc ctttccctt cctgaactgc tttgtgaagg 900  
 agcagtgtaa ggcattcccc caccatgaag taagtgggtt cgtgggggtt gttgcctgtg 960  
 gctgggacct gggagggtacc tgagagaatt ggtgttattt gggcttgtgg ggaggggcta 1020  
 agaaatgata agaaaagaca agaattctta aaaggtgaaa tgggagcagg cttgagtcac 1080  
 ggacctgccc tagcctcccc cagtttgtgg ccctgatagg tgcctcctt gccatctgct 1140  
  
 gcatgatctt cctgggcttt gcggatgatg tactgaatct gcgctggcgc cataagctgc 1200  
 tgctaccac agctgcctca ctacctctcc tcattgctta tttaccaac tttggcaaca 1260  
 cgaccattgt ggtgccgaag cccttccgcc cgattcttgg cctgcatctg gacttgggta 1320  
 ggtagtctg ccaactgctac tcctatggca cctacttcag ggcaccttc ctggtgcttc 1380  
 acattctct tcaagtgttc cttttctgtc tctgtgtctt ccagatcct tctggtagc 1440

ccttcacct atcctctgtc ctcaccactt ttctaaatcc tcctccccta ggtggcacta 1500  
 cttttcctac catctctccc ttcaggaatc ctgtactatg tctacatggg gctgctggca 1560  
  
 gtgttctgta ccaatgccat caatataccta gcaggaatta atggcctaga ggctggccag 1620  
 tcactagtaa ttctgtctc catcattgtc ttcaacctgg tagagttaga aggtaggtgg 1680  
 gattgggggt ggggagagag aagtctgaat gttaaagggt tggcctgata tatgactttg 1740  
 ggaaattcag ggaaaaaag caatatgcgt agtaattata gaagataagg gaggctactt 1800  
 actttgcaaa taatgcagat ttattgaaag tgagaaagaa aaatagcagc cgtgtcattt 1860  
 atagctggat tggcactaac agctaggcca tgatcttctc ccattgaata taaacaattt 1920  
 cacagaacct caacgttaca cagggtcatt ctgtgacat gatggaggaa gacccaaaact 1980  
  
 cgaccctcc ctctataatc ctgtttgagc acagataaaa ccacaaaaac actgagcaac 2040  
 ccacaaaatg gccaaagatc tctctctctg ttaacgtgag ccatgagcga ctgctcgggc 2100  
 tttcaataa caactcagtt cctaccacct tttatttgt tttttgagac agggctctccc 2160  
 tctgtcacc aggctggaga gcagtggcac gatcttagct cactgcatcc tctgactcaa 2220  
 acgatctcc tgcceccagc tcccaggtag ctggggetac aggcattgac caccacacct 2280  
 ggcaaat tttgtat tttgtagagacag ggtttcacca cgttgcttag gctggctttg 2340  
 aactcctggg ccgaagtgat ttgtcagcct tggcctccca aagtctggg attacactt 2400  
  
 tgagccactg tggccagcca gttcctacca cttcttagat aaacattaaa atgcttgatc 2460  
 agagaattat tgttgtttt ttttctttt ctttttttt ttttttgag acggagtttc 2520  
 actcttgccc aggtctgagt gcaatgggtc gatctcggct cactgcaacc tccacctccc 2580  
 aggttcaggc gattctcctg cctcagcctc cctagtagct gggattacag gcatgtgcca 2640  
 tcacgccag ctagtttgt atttttagta gagatgggat ttctccatgt tggtaagct 2700  
 ggtctcaaac tccagacctc aggtgatecg cccacctcgg cctcccaaag tgctgggatt 2760  
 acaggtgtga gccaccgac ccggccatga attacactg ctttctaaca gcaccaatc 2820  
  
 cagagcaaaa ctcttacttt cttttacct ctcccataat acccaaaact gcaagccct 2880  
 cctaactc tcttactgag acattccgtg gttcccaatg gtgtgtggtt tctgaagtct 2940  
 cccttttac aacaagtc taaacctagc ttcgaactat agatgtgttt ctagtgtct 3000  
 ttggctgatg gacatcaaca aatgtttatt aaagctaagt actttttaa catgatcgt 3060  
 tttaaatctt gtaatggttt tatgtggcac atgttataat cagccctgtt ttacagatga 3120  
 gtaaacagat ttagagaagt taaatgtgtc atgatcaaga tcaaggtcac aaagctaaga 3180

agtaaagttg ggtccaaac tgacatcaga atgggctaaa ccaaatttaa gacagtaact 3240  
  
 agtttgaag gctgcatgaa agaggtggaa tattgggaat tgccttgggt gacataaaag 3300  
 gggatttgag ttcttgaag tgacttgggt gaggtggatg atacagctgt aacagaact 3360  
 tagacaaaa taggacatg gtatgcagaa gaagtgggta ttaattttcc cttctttctt 3420  
 ccttgcttcc aaaggtgatt gtcgggatga tcatgtcttt tccctctact tcatgatacc 3480  
 ctttttttcc accactttgg gattgctgta ccacaactgg taagtaggcc tatggataag 3540  
 gggaaaagg gaaaactacc cgaacacatg gcaaagatgg cccttatcat aaccacctt 3600  
 gtggtggaga ggttaaact gtgcatacct ccatggaatt ttctgtgtct tcagttggtc 3660  
  
 gtattctgaa atttctcctt acccaacagt atttggggat gagtgcgtgg aggtcccagg 3720  
 aatagatgaa ttcagggcct tggatcctgc agagttgctg cacaactgga gtctcctcta 3780  
 agtcagaact agggtcaggg ctagtacagt gcccataggg tgtgatgtga gagaaaggat 3840  
 tggtaatgcc tcttgccact ggctcggatc ctctcccca cacaggtacc catcacgggt 3900  
 gtttgtggga gacaccttct gttactttgc tggcatgacc tttccctgg tgggcatctt 3960  
 gggacacttc agcaagacca tgctactatt ctctatgccc caggtgttca acttctctta 4020  
 ctactgcct cagctcctgc atatcatccc ctgccctcgc caccgcatc ccaggtagcc 4080  
  
 gctttggggc ttgaaatgga catcatagcc ttttacttgg ggatatctaa tgccagcctt 4140  
 tacattgctg tgcaaaggga gtgggcccga agaagggcta tttccatgtg agtaaccctt 4200  
 tataacttcc aaagcacatt tatttgcatc atctgatact cacagtgggt ctgataacag 4260  
 caagcagcag agccagaaat agatctcagg ttgactccac attcaatgct ctctctattg 4320  
 attagccaca gggaggaggg ttcaaatagt ggcccagtca catgaagctg tcttccccc 4380  
 gcagactcaa tatcaagaca ggcaactgg agatgagcta ttccaagttc aagaccaaga 4440  
 gcctctcttt cttgggcacc tttattttaa aggtaacagg gtaacaagga ggttaaggccc 4500  
  
 taggctgcca tctgacctt gaggaatggg gaacctagcc ctacatcaga tccaagggga 4560  
 acttgggaagc attaaataga tccacattcc taaagcatag gtattagctg aggttctctt 4620  
 cacctctggt cctccagggt ggcagagagc ctccggctgg tgacaataca ccaaagtgat 4680  
 actgaggatg gtgaattcac tgaatgtaac aacatgacc tcatcaactt gctacttaa 4740  
 atctttgggc ccatacatga gagaaacctc acattgctcc tgctgctgct gcaggtgagg 4800  
 atggggattg ggtttatacc tcttgtctc cctttctccg tgattcttat tccagtcct 4860  
 ttctccttgc agatcctggg cagtgccttc accttctcca ttcgatatca gctcgttcga 4920

ctcttctatg atgtctgagt cccctgatca ttgtccttta cctcacagtc tctaggattc 4980  
ctgactcagg ctgacctctc tctgggccca gactgcctcc ttgccaggc ctctctcact 5040  
cttcatactc ttccagatit tgttctcagc attttctttt ccctgtgac actggcatcc 5100  
tgggcgtttc ttgcccccta ctgtctactg attggatitit acttatgact ttctgcaact 5160  
tgctactctc cctctccatc ctgtctttgc agcctcacag ggtgggatac agaagtittt 5220  
tttttgacgt tatccacagt cacatttcag agtctgact ctcaaggaac tactggtttt 5280  
tgggatagaa ctggggccag ggctagggac acaggctcca cagtacctg ttatttgatt 5340

gtaaattaag tgttctgatt agtaagaagt aagcaggggg ccacatgctc tcaatggaga 5400  
caataaagtg ttgtctatit cttat 5424

<210> 15  
<211> 5894  
<212> DNA  
<213> Pan troglodytes  
<220><221> misc\_feature  
<222> (3408)..(3408)  
<223> n is a, c, g, or t  
<400> 15

aaaccgtacg tgcgtttccg ggaactgagt tgtgtttacc ttggcttccg actatgttgg 60  
caacaggttt cctgcaagaa actggcgcgt ctccacacc tcgtccctcc tccccaccc 120  
ctgcctttca atagccatct tcttggagcc ggaggcatcc cagattaagg gagaggtacg 180

ggccctttaa gcttgaccta tggaggcgga cggagctaaa actgacgtgg aaccggaatg 240  
tgagcgggtg cagacacgtg gtacaaggag gcattcatct tggaaccggg caattggcat 300  
ttccgctctg ggiagtacat ctttaacata atgttaggga agtatccgtt ctgggctgcc 360  
tttctttaat tgcgtttcca gtactctctc ggtgattcta ctcttgaaca taggatgaaa 420  
tttggaatca cacttctctt gcacttccat ccccaccctc taatgccat attaaaaatg 480  
gcgcccgccg ccttccgcag taatggttgt tcagcgaaca agatccgggc ggaaacagta 540  
gatagcgggg tgcagcgggg cagaacatag gttgccttag agaggttctc cggtgtctcg 600

agggcggctc aagttagagt tgttgggttt tgetcagatt ggtgtgggaa gacctgcct 660  
gtggggagcg gccactccat actgctgagg cctcaggact gctgctcagc ttgccgtta 720  
cctgaagagg cggcggagcc gggcccctga ccggtcacca tgtgggcctt ctcggaattg 780  
cccatgccgc tgetgatcaa tttgatcgtc tcgctgctgg gatttgtggc cacagtcacc 840  
ctcatcccgg ccttccgggg ccaattcatt gctgcgcgcc tctgtgttca ggacctcaac 900

aaaaccagcc gacagcagat gtgagcagcg gcacacgggt ctgggcaggg ggcaaggct 960  
 aaggaaggag tggctagggc aggggcgggg accgggtgc ttgaccacac gtgaagactc 1020  
  
 agaactaacc caggcagcct ggaactcgga gaggtgatga gcagaactta ctgcattgg 1080  
 ggaaaggatg ggtlagggacc cttgggtata tctgggactc tggcagtggg gctttctcc 1140  
 ctccgcccc ctcaccactt accagaataa aaaaccggga atgagaagac cactttgggt 1200  
 tattgtaaca cctgcactag tgagtgaaca cgccccctt gctcttcccc ctgccccct 1260  
 tgctgctggg ccacagccca gaatcccagg gagtgatcag cggctctgtt ttcctatca 1320  
 tcctctctg cttcatcctt tcccccttc tgaactgctt tgtgaaggag cagtgaagg 1380  
 cattccccca ccatgaagta agtgggttcg tgggggtgat tgcctgtggc tgggactgg 1440  
  
 gaggtacctg agagaattgg ggttatttgg gcttgtgggg aggggctaag aaattatcag 1500  
 aaaagacagg aattcttaaa aggtggaatg ggagcaggct tgagtcattg acctgccccga 1560  
 gccccccca gtttgtggcc ctgatagggt cctccttgc catctgctgc atgatcttc 1620  
 tgggctttgc ggatgatgta ctgaatctgc gctggcgcca taaactgctg ctacctacag 1680  
 ctgcctcact acctctctc atggtctatt tcaccaactt tggcaacacg accattgtgg 1740  
 tgccaagcc ctccgcccc atacttggcc tgcatttga cttgggtagg tagtcctacc 1800  
 actgctgccc ctatggcacc tacttcaggg aaccttctt ggtgctccac attctctcc 1860  
  
 aagtgttctt ttctgtctc tgtgtcttc cagatccttt ctggtagccc ttcactctat 1920  
 cgtccgtcct caccactttt ctaaaaattc ttaaatctc ctcccctagg tggcactact 1980  
 tcttttcta ceatttctc ccgcaggaat cctgtactat gtctacatgg ggctgctggc 2040  
 agtgttctgt accaatgcca tcaatctct agcaggaatt aacggcctag aggctggcca 2100  
 gtcactagtc atttctgctt ccatcattgt ctcaacctg gtagagttgg aaggtaggtg 2160  
 ggattggggg tggggagaga gaagtctgag cattaaaggt gtggcctgat atatgacttt 2220  
 gggaaattca gggaaaaaaa gcaatatgtg tagtaattat agaagataag ggaagctact 2280  
  
 tactttgcaa ataacaatgc agatttatta aaagtgagaa agaaaaatag cagccctgtc 2340  
 atttatagct ggattggcac taatagctag gccatgatct tctccattg aatataaaca 2400  
 gtttcacaga accccaactg tacacagggt cattctgtga ccatgatgga gcaagactaa 2460  
 aactagacc ctccctctgt aatcatgttt gagcacagge aaaaccacaa gaacctgag 2520  
 caaccacaa aatggccaag atcccctctc tcggctaaca tgagcgactg ctgctgctct 2580  
 ccaataacaa ctcagttcct accacttctt ttttttttt ttgagacagg gtctccctct 2640

gtcatgcagg ctggagagca gtggcgcaat cttagctcac tgcacccct gactcaaacg 2700  
  
 atcctcctgc cccagcctcc caagtagctg gggctacagg catgtgccac cacacctggc 2760  
 aaatTTTTgt atTTTTgtA gagacagggt ttcacatgt tccctagct ggtcttgaac 2820  
 tcttgactc aagtgatctg ccaggcctcc caaagtgtg ggattcactc ttgagccact 2880  
 gtgccagcc agttcctacc atttcttaa taaacattaa aatgcttgat catagaatta 2940  
 ctcttgcttt cttttcttt cttttcttt tttttgaga cggagttttg ttcttgccca 3000  
 ggccggagta caatggtgcg atctcggctc accgcaacct ccgcctcca ggttcaagcg 3060  
 attctcctgc ctcagcctcc ctatgtagctg ggattacagg cacgtgccac cacgcccagc 3120  
  
 taatTTTTgtA ttcttagtag agacgggggt tctccatgtt ggtcaggctg gtctcgaact 3180  
 cctgacctca ggtgatctgc ctgcttcagc ctcccaaagt gctgggatta caggcgtgag 3240  
 tcaccgacc cggccatgaa ttactcctgc ttctaacag caccagctcc agagcaaac 3300  
 tactttctt caccctctcc caaaatacc aaaacaaacg ctactacaag ccctcctaa 3360  
 cacccttta ctgagacatt ccgtggttcc caatggtgtg tggtttnga agtctccctt 3420  
 tttaacaaca gtcattaaac ctatctttga gctatagatg tgtttctgat ggtcttgct 3480  
 gatgaacatc aacaagtgtt tattaagct aagtactttt taaacactat cttatttaa 3540  
  
 tcttgaatg gttttatgtg gcagatgta taatcagccc tgttttacag atgagaaaac 3600  
 aggcttagag aagtcaaatg tgtcatgatc aagatgaagg tcacaaagct aagaagtaaa 3660  
 gttggtatcc aaacttacat cagaatgggc taaaccaaat ttaagatagt aactagtttg 3720  
 gaaggetgca cgaagagggt ggaatattgg gaattgcctt gggtagacata aaaggagtat 3780  
 tgagtctta aaagtgactt gggtaggtg gatgataaca gctgtaaca gaacttagac 3840  
 aaaaatagga ccaaggttg cagaggaagt gggatataac tttccttct ttcttcttg 3900  
 ctccaagg tgattgtcgg gatgatcatg tcttttcct ctacttcag atacctttt 3960  
  
 tttcaccac ttgggattg ctctaccaca actggttaagt aggctgtgg ataaggggac 4020  
 aactacctga acacatggca aagatggccc ttatcataac ccacctgtg gtggtgaagc 4080  
 taaacctgcg catacctta tggagtttc tgcgtcttca gttgtagta ttctgaaatt 4140  
 tctctacc cagtagtagt taggatgag tgcgtggagg cccaggaat agttgaattc 4200  
 agggccttg atctgcaga gttgctgca aactggagtc tctctgagt cagaactagg 4260  
 gtcagggcta gtccagtcc catagggtgt gatgtgagag aagggattgg taatgcctct 4320  
 tgccactgce tcggatctc tccccaca caggtacca tcacgggtgt ttgtgggaga 4380

taccttctgt tactttgctg gcatgacctt tgccgtggg ggcactcttg gacacttcag 4440  
 caagaccatg ctactattct tcatgcccc ggtgttcaac ttcctctact cactgcctca 4500  
 gctcctgcat atcatccctt gccctcgcca ccgcataccc aggtagccgc tttggggctt 4560  
 gaaatggaca tcatagcctt ttcacttggg atatctaag ccagcctata catttctgt 4620  
 gcaaaggag tgggccccaa gaaggctat ttccatgtga gtagccctt ataacttaca 4680  
 aagcacattt atttgataa tctgctacag tggttctgat aacagtaagc agcagagcca 4740  
 gaaatagatc tcaggttgac tccacattca atgctcttcc tattagccac agggaggagg 4800

gttcaaatag tggcccagtc acatgaagct atcttcccc cgcagactca atatcaagac 4860  
 aggcaaaactg gagatgagct attccaagtt caagaccaag agcctctctt tcttgggcac 4920  
 ctttatttta aaggtaacag ggtaacaagg aggtaaggcc ctaggctgcc atcctgacct 4980  
 tgaggaatgg ggaacctagt cctacatcag atccaagggg aacttgaaag cattaatag 5040  
 atccacattc ctaaagcata ggtattagct gaggttctct tcacctctgg tcctccagg 5100  
 tggcagagag cctccagctg gtgacagtac accagagtga gactgaagat ggtgaattca 5160  
 ctgaatgtaa caacatgacc ctcatcaact tgctacttaa aatccttggg cccatacatg 5220

agagaaacct cacattgctc ctgctgctgc tgcaggtag gatgggaatc gagtttatac 5280  
 ctccgtgtct cctttctgc gtgattctta ctccagtcca tttctcttg cagatcctgg 5340  
 gcagtccat caccttctcc attcgatac agctcgttcg actcttctat gatgtctgag 5400  
 tccttgatc attgtcctt acctcacagt ctctaggatt cctgactcag gctgacctct 5460  
 ctctctggtc ccagactgcc tccttgcca ggctctctc actcttata ctctccaga 5520  
 tttgttctc agcattttcc tttctctgtg atcattggca tcctgggctt tcttgcct 5580  
 ctactgacta ctgattgat ttacctatg gctttctgcg acttgctact ctctcctct 5640

ccatccate tttgcagcct catagggtag gatacagcag cttttttgc agttatccac 5700  
 actcacattt cagagtctg actctcaagg aacctagg ttttgggata gaacttgggc 5760  
 cagggctagg aacacagct ccacggtag atgtcattg attgtaaatt aagtgtctg 5820  
 attagtaaga actaagcagg gggccacatg ctctcaatgg agacaataaa gtgttctct 5880  
 tttctattg tta 5894

<210> 16

<211> 4557

<212> DNA

<213> Canis familiaris

<400> 16

gtgaggagc aagtgcggcg ggggacagcc gaggggtgcmc gctggaggct cgcgggagtc 60

ctgggggcmc ctcaattcag agttgggttt tgctcaggcc gctgtgggag gatccagcct 120

gtgccgagcg gctgctctc cccgcggggg gctccgggct accgcccagc tcgcccatta 180

gccgaggcgg cggcagagcg gggccctgg ctggtcatca tgtggcctt cccggagtgg 240

ccgatcccg tctggtgaa tttggtcggc tcgctgctgg gatttgtggc cacggtcacc 300

ctcatccccg ctttccgtgg ccacttcac gccgcgcacc tctgtggcca ggacctcaac 360

aaaaccggcc ggcagcagat gtgagcggtg gcacccgggt cgggggaggg ggccggcagg 420

gcaagggcgg gacctggggt gcctgacccc gcggacacgc agcgcctaacc ccgacagacg 480

ctgcgggctc tgggagacga agggcagcmc tggccaactc tgggaaggga tgttgagta 540

caggggacct tcgggtgat cagggactcc agcgcctggtg cccttcacc cccctcccc 600

gtagatcgt gtaatgctg ctctagttag tgaccacgcc ccctctctc tcccccgccc 660

cctccctttg ctgctgggcc acagcccaga gtcccaggga gtgatcagcg gtgctgtttt 720

ccttatcacc ctcttctgct tcacccctt ccccttctg aactgtttta tggaggagca 780

gtgtaaagcc ttccccacc acgaagtaag tgggtgagtt gggggcggtt gcttggggct 840

ggggcctggg agctacctgg gagagtttg gttattaggg tttgggtgga ggggctgagg 900

aaggagcga gagacgggtg tttttgcaag atgatgtggg cataggcttg agcggtgacc 960

tgcccgagcc tccccagtt cgtggccctg atagggtgcmc tccttgccat ctgctgcatg 1020

atcttctgg gctttcgga cgatgtactg aatctgctg ggcgccacaa gctgctgctg 1080

cctacagctg cctcgtacc tcttctatg gtctattca ccaactttgg caacacgacc 1140

attgtggtgc ccaagccttt ccggccgatt cttggcctgc atctggactt gggtagtag 1200

ccctgtgact gacgtccctg tggcccttac tttggggcac ccttaccctg ggagataatc 1260

tagcagagca tcattcctgg tgctecagat cctcttccaa gtgtcccat cttgttctg 1320

tgtcttctca gatccgttct gttggtcctt cgtccaatcc tctgtctca ccactttct 1380

cagaagaata ttcttaagtc ctcatctca tggatggcac acttcttact ctcttctcc 1440

cccagggac ctgtattatg tttacatggg gctgctggca gtgttctgta ccaatgcat 1500

caataccta gcaggaatta atggcctaga ggcaggccag tcgctagtta tttctgctc 1560

catcatcgtc ttcaatctgg tggagctgga aggtaggtga gagtgggagt ctgagtatta 1620

aggaaactgc ctgatactg gctttgggga attcaggaaa aaataaaagc aatatattaa 1680

gattaaatgt aaagaaaaac agctctgtca ttgacagctg aattggcact aatagtagg 1740

ccatggtcctt ctgctgaaca taaacaattt cacagaactt cacaatcaga cgaggctact 1800  
 ctatgtccat gatagagtaa agcaaacca gattcctcca taaacatgtc tgagtatagc 1860  
 cagaactgca ttttgtgcat cccacaaaaa tgactaggat ctcctcttcc tggctaaggt 1920  
 gagcaattgc ttctttctga taacttgggt ctacttagag aaaactaaga tgctcataga 1980  
 attacttcca ctgacagcac ccagcttgg gcaaaacttt gctccttcc tttctcccc 2040  
 aaattactca aaacaatcct ataacacatc ttcttaatac ttcctactg aggcatcccc 2100  
 tggttaccta tggtgctggg tctacagtgt ctctcttgtt acacgtcagt aaaccagct 2160  
  
 ttgactgcag gtgtgtttct ggtggtcctt ggctgatgga tatcagtgtc tattaaaaca 2220  
 aaatactctt aaagcattta aactttgtaa tgggcaagt gttctcatga accatatttt 2280  
 acagttgagg aaacagaggg cgagagaatt taagtgtgtc atgatcaagg tcacacagtt 2340  
 agaaaagtaa gctagtattc aaacctgggc tgaatgatct aaaccaaatt gaagacagca 2400  
 acttgattta ggaagggttc atgaatgagg tggaatatta ggaattgcct gaggacaca 2460  
 aaagaagtag tgagttctgg aatgggactt ggaagaggtg gaaagtacag ctggggacag 2520  
 aacttgagac agaaatagga cccagttatg cagggggaag taccttatca actcctcctt 2580  
  
 ctttctttt ctcttcccc tgcttccaaa ggtgattatc gggatgatca cgtctttcc 2640  
 ctctacttta tgataccctt tttttcacc accttgggat tgctctacca taactggtaa 2700  
 gtgggcatg tgaacatgta gcaagtatgg tctgttggg cctgacccaa ctctgttgg 2760  
 agaggctaag cctgcgcaca cctgtattga gtgtttctg gatgcctagt tggtaatatt 2820  
 ctcaattac tctctacca gttgcagtta gagacaagtg ctgtggagcc cccaagaaga 2880  
 gatgaattca gggctttggg ttctggaggc ttgttgaag atctggagtt tctccgggc 2940  
 caggactaga gtcagggcta gtccagggtt cagggcgtgt aatatgagag aaaagactga 3000  
  
 tagtgcctcc tgccactggc tcagatctc tccccacac aggtacccat cacagtggtt 3060  
 tgtgggagat accttctgtt actttgctgg catgacctt gccgtgggtg gcatcttggg 3120  
 aacttcage aagaccatgc tactcttctt catgccccag gtgttcaact tctctactc 3180  
 actgcctcag ctctgcata tcataacctg ccctcgcac cgcattcca ggtagccact 3240  
 ttggggctta aaaggacat cttagctttt tcaactggga tgcataaagc cagccttctg 3300  
 catctgtgt gtaagggaa tgggccc aaa ggaggctct ttcctgcaa ttagccctta 3360  
 taaatgacag agcacattca cccacataat ctgatcagct ctgatcacac agtggtaagc 3420  
  
 agagccgaa acagatctc aggttgtctg attccactt cggfactctt cctattaatt 3480  
 gaccgcagtg tggagagttc ttggagtagt ggcccagtca cataaagctc tcttccccct 3540  
 gcagactcaa taccaagaca ggcaaactgg agatgagcta ttccaagttc aagaccaaga 3600

gcctctcttt ctgggcaac tttattctaa aggtaacagg gtaacgaggt aaggctctag 3660  
gccaccatcc ggaattcagg gcctggggac cctcggcttg catcagatcc aaggggagcc 3720  
tggaagcatg aagcagatcc cccattgctg aagcagagtt gaagtctct ccacctctgg 3780  
cccctccagg tagcagcgag cctgcagctia gtgacagtgc accagagtga gaatgaggat 3840

ggatgccttca cggagtgtaa caacatgacg ctctcaact tgctccttaa gtttctcggg 3900  
cccatgcatg agagaaacct gactctgctc ctgctgctgc tccaggtgtg gtcagggag 3960  
ggctttgctg gctctggtct cctttctcc atggctctga ctctggtgtg tttctttctc 4020  
ctcacagatc cttggcagtg ctgtcacctt ctccatccgg taccagcttg tccggtctt 4080  
ctacgatgct tgagtcccc aatccttgcc cttcactgca tagtctgcag ggttctgac 4140  
tcaggcctgc ctctttctgg gccaggcacg cttccgggcc caggcctctc tcaccttta 4200  
cttttctcca gatthttgtac ttagegattc cgttccgctg tgatcgacat cctgggctg 4260

tcttgcctg tactgactgt tgattggact ttgcctgtgg ctttcttcaa cttgctgctc 4320  
tcctctcta tccatccct gcggcctccc aaagtgggat actgtgcttt ttatgcagtt 4380  
atccaccact cggactctcg aggaatatgt tggcctggg gatagaacce tggctgggga 4440  
gaggacaca ggctcgaaga tcacttgatt atttgacct aaattaagta ttctgattcg 4500  
taagagcaga ttggggggcc aggtgctccc agtggtgaca ataaagtgtt gtctttt 4557

<210> 17  
<211> 5374  
<212> DNA  
<213> *Cricetulus griseus*  
<400> 17

caagcagag cctaggttgc tttataaac ctcttgggga agcccagagg cggttcaaat 60

taagagtgt tgggttttgc cccgctcgc atgtgaggag cggacactgc tcacggctga 120  
gacctcgggg ctgcttccca ccagttagct gagaaggctg cggagctgga acctctggcc 180  
actcgccatg tgggccttct ctgaggtacc gattccgctg ctggtgaatt tgateggctc 240  
gctgctggga tttgtggcca cgctcacct catcccggcc tttcgtggcc actttatcgc 300  
tgccgcctc tgtggccagg acctcaaca aaccaaccg cagcagatgt gagcagtggc 360  
acacgggtgt cccgggcagg ggccaggggt gggcaaggca caggcgagct ctgaggtgct 420  
taaatgtgct tacgaacaa atctaactgg agttgtccgg gaccctggga ctcgatggcc 480

agaagtgggt agcactgggg aatgctaagg aaggggaccc ttgagtgaga acatccagcg 540  
gctgctgctt cccccgcc cccactgcc tcccgtcca ctgctcccc gcctcactcc 600

tgggaagatc ttttgggtca catggttttt gcactaacca cgcccatttc ttcttccttc 660  
 tccaccccct tgctgcgggg ccacagccca gaatcccagg gagtgatcag cggtgccgtt 720  
 ttctttatca tctgtttctg cttcatcccc ttccccttct tgaactgctt tgtgaaggag 780  
 cagtgtaaag ctttccccca ccatgaagta agtgggttcg tgggggcggt tgcctggggc 840  
 ctgggaggtt cccgagagag ttggggttgt gtggatttga ggaggaggga ctgaggacct 900  
  
 agtggaaaag acagaaattt ttgaaagctt gaatggcagt aggcttgagt catgacctgc 960  
 ccgagcctcc ccagtttgt ggccctcata ggtgcccttc ttgccatctg ctgcatgatt 1020  
 ttcttgggct tcgcggacga tgtcctgaat ctacgctggc gccataagct gctgctgccc 1080  
 acagctgcat cactacctct tcttatggtc tattttacca actttggcaa cacaaccatt 1140  
 gtggtacca agcccttccg ccagtttctt ggcttgcata tggatttggg tgagtatccc 1200  
 tgctgctaca gccctgtgg cacttatttc aagtcacct cccccaaag gtgccagca 1260  
 gagcacctt ctgatgttc cacactcccc tgtttttgtt ccgtccctgt gaatgctcag 1320  
  
 gttctctctt gtgccctgtc attgtgtgtt ctgttttcag aataccgtta gatcctttcc 1380  
 tagctgtcac tgetttttat actatgtctt gcagggatcc tgtactatgt ctacatgggg 1440  
 ctctggcag tgttctgtac caatgccatc aatatactag caggaattaa cggcctagag 1500  
 gccggccagt cattggtcat ctctgcttcc atcattgtct tcaacctggt ggagctgcaa 1560  
 ggttggtggg aagagagaga tctcagtgtt cagagaattg cctgatatat agctttgaga 1620  
 aaaggggggc ttatagaaga tagggaaagc tatttacttt gcaaataaca atgaagagtt 1680  
 acttgagtag gaggaagaaa aatagcagtc tgtcatttat agctggattg gcatgagtag 1740  
  
 ctagaccatg acigtttcct attggacata aatagtttca tagaaccca gcatgagaga 1800  
 ggggcgccct gaccgtggtg aaacaagaca gaaaccagac ttctcccact gtaatcatgt 1860  
 ctgaacaccg acaaaagcac aggaacaaaag tcagcccaaa catcccttct tgtctaatgt 1920  
 gagagggtat agcttctttg cagtaacaac tcagttctctg ctacttctta agttgttcag 1980  
 tcagaaaatt acttctgctt tctgacatca ggcagtccag agcacaactt ttcttgcag 2040  
 cctccccaga accacttaaa gtgaatccta ttgtaagtcc cttctaaca cctttagagt 2100  
 acctgcccag catgaggcct tgggtacagt cccagtatc tctgtttgca tgcattgaca 2160  
  
 catacccaca tgcacacact gaacttactt attgaaggta agcaatattt atttgattt 2220  
 tttgtgtgtg tgacaaaatt tcactatgia attcagaata gccttgaatt cactatgtag 2280  
 cctaggccgt cctcgaactt acagtataa tctgcctca gcttccctag tgctaagatt 2340  
 caaggtatgc actaccaggc cagctaagaa agcaattttt aaactaggta tgggtgcaca 2400  
 catcactaat tcaaacctc tgggagactt gggcaggaag atcatgagtt tgagctcagc 2460

ctgggcactt ggtaagtctc tgtttctaga aataaaacat ggagtggatga tacacacctg 2520  
 taatcccagc attcattgagg ctggggcagg aggatcacca caaggtcaag acctgcctgg 2580

gttacataag caagttcaag gccagcgtga actacgtagt gagaccctgc ctcaaacaaa 2640  
 caaataaata aataaacatg atcctgagtt tggttcccag tactcccccc aataaatgaa 2700  
 atgaaatgaa agagctgggg aggacagctta gtgctaaggt ccaggacccc atgtgaaggc 2760  
 agctgcgtgt atgtgttacc agcctgttt catacactag ataataaac tggttttcaa 2820  
 acttaagtca gcatgtctgg acaaagtga gactttaact tgtttttgac ggtttcatga 2880  
 cagtagtgag ctattgggaa ctgcctgggt accatcaaag gaataatgag gggctgggga 2940  
 tttagctcag cggcataagc gcctgccttg caagcaggca gtcatgagtt tgatccccgg 3000

taccgataaa aaggaaaaag acaaaaaaaaa aaggaataat gagtttttga aggtggctcg 3060  
 ggtgagggga ggtggcaaca gagacaggga tgcgacagac aaaatgaaga gcaggggaca 3120  
 taggggagat ggggtgtcac ttttcttct ttgtcctttg tttccaagg tgattaccgg 3180  
 gatgatcatg tcttttcct ctacttcatg ataccgttt tttcaccac ctgggattg 3240  
 ctgtatcata actggtaagg aggctgtggc tcagggaaaa ggaaaacaac taactggtea 3300  
 ttggacaaag atggctctga tcttaacca gctcctgaaa gacaggctga acttgcgat 3360  
 acttttgctc agtgttttct gggattcag ttggtggatt gcctcccccc gccccgttt 3420

tttttgaga cacctgtggc ctctcgagtg ctaggccagt gctttactac tgagtcttgc 3480  
 cctctagat tctcaggttt gttcttttct cagcagttgg agacaagtgc tatggagccc 3540  
 caggaataat tatggggact tgcgttctgc agacttgcta gacctcctg tccgaactag 3600  
 gatcagggga gcatgtgtgg tggctcacac ctgtaatccc agcactcagg agactgagc 3660  
 aggaggatta ccatgagatc gagggcacc ctagctcaca tagtgacttt gaggccagcc 3720  
 tggactacat agcgagactc ttgtctcaa aagaaaaaaaa aagaagaat aaaagaacag 3780  
 gggctctgagt tcgtccagta cctagcctgt gtgatgtgag agaaaagact gtgatgcctc 3840

ttggcactgg cttggatcct tccccca caggtacca tctcaggtgt ttgtgggaga 3900  
 caccttctgt tactttgctg gcatgacctt tgccgtggta ggcatcttgg gacacttcag 3960  
 caagaccatg ctgctcttct tcatgccaca ggtgttcaac ttccttact cgctgcctca 4020  
 gctctgcat atcatcctct gtectgccca ccgtataccc aggtagctgt ttgggggctg 4080  
 gaaagccttt ctactgggat gtctaacacc aggctctaca tttgctgtgc aaagaatgtg 4140  
 ggcccatagg aaggctaact ttttcatgt aggtggcct ttaagtttac acagcacgtt 4200

tacttccata atctcattta atactcacag tagttctgat catagagtag taagcagcag 4260

agccagaaat agatctcact ccatgatcag tgtttttctt agtcattaac ggaagaaagt 4320

tttttgagta gtgaccagct cacacgaagc tgcctttccc ctacagactc aataccaaga 4380

caggcaaact ggagatgagc tattccaagt tcaagaccaa cagcctttct ttcttgggca 4440

cctttatfff aaaggaatac aggtaacgag gaggaagggc cccaggccac catcctgaac 4500

ttgggacatg ggggaccagc gcctacatta gatctagagg gagcttggaa gcattaagca 4560

gagccctgtt ccigacatac aggtattggc tgaagttttt ctgtctgtct ctggctcttc 4620

taggtagcag agagactcca gctagtgaca gtgcaccgga gtgagggtga ggacggggcc 4680

ttcactgagt gtaacaacat gaccctcatic aacttgctgc ttaaaatctt tgggcccata 4740

catgagagga acctcacatt gctcttgctg ctgctacagg tgagcctggg gtgagtttgt 4800

gcctctcat gtctttttct ctatggttct tattctagtc cttttctctt tgcagatcgt 4860

gggcagtgtc gtcaccttct ccattcgata ccagcttgtc cgactcttct atgacgtttg 4920

agttctgaa gattgcctc tgccacactg tctccagggg tctgctcag gccagccagt 4980

ctggttctgt gggcctctcc caatctcag tctcttcag atttattccc agcatttttc 5040

ataacctatg attatcaaca tctgagcca tttttgcct ccagcaacta ctaactggac 5100

tttgccatg gctccttca acttgccact ctcctacc atcacagcca gaggttggat 5160

gtagcagctt ttatgcagat atccacaact cagctttcag agtcctcact ctcaaagaac 5220

atgctgggcc ttgagataga acctgagcta gggctagggc cactggtgca agggtgattt 5280

gatatttggat tataaattaa gtgttctgat tagtaagaca gaaggggagc ctgggtctcc 5340

caacggtgac aataaagtgt tacctttttc ttgt 5374