



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년07월21일

(11) 등록번호 10-2135745

(24) 등록일자 2020년07월14일

(51) 국제특허분류(Int. Cl.)
C12N 15/85 (2006.01) C07K 16/00 (2006.01)

(21) 출원번호 10-2014-7018672

(22) 출원일자(국제) 2012년12월05일

심사청구일자 2017년12월04일

(85) 번역문제출일자 2014년07월04일

(65) 공개번호 10-2014-0113666

(43) 공개일자 2014년09월24일

(86) 국제출원번호 PCT/IB2012/056977

(87) 국제공개번호 WO 2013/084157

국제공개일자 2013년06월13일

(30) 우선권주장

61/567,675 2011년12월07일 미국(US)

(56) 선행기술조사문헌

EP1308510 A

(73) 특허권자

아이크노스 사이언스 에스. 아.

스위스, 2300 라 쇼드폰, 슈명 드 라 콤포타 5

(72) 발명자

로셰, 테니얼

스위스, 2300 라 쇼드폰, 슈명 드 라 콤포타 5

에비셔-거미, 크리스텔

스위스, 2300 라 쇼드폰, 슈명 드 라 콤포타 5

(뒷면에 계속)

(74) 대리인

이원희

전체 청구항 수 : 총 30 항

심사관 : 이효진

(54) 발명의 명칭 발현 카세트

(57) 요약

본 발명은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열의 발현에 유용한 발현 카세트에 관한 것이다.

(72) 발명자

모레띠, 뎀에르

스위스, 에프-2300 라 쇼드퐁, 슈명 드 라 콤베타
5

버츠친거, 마틴

스위스, 2300 라 쇼드퐁, 슈명 드 라 콤베타 5

명세서

청구범위

청구항 1

프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 포유동물(eukaryotic) 글리세르알데하이드 3-포스페이트 디하이드로제네즈(Glyceraldehyde 3-phosphate dehydrogenase(GAPDH)) 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하는 발현 증강 요소를 포함하는 발현 카세트(expression cassette)로서,

여기서, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호 8인 발현 카세트.

청구항 2

제 1항에 있어서, 상기 발현 카세트는 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림(non-translated genomic DNA sequence upstream)을 추가적으로 포함하고,

여기서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호 7인 발현 카세트.

청구항 3

프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열 및 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 포함하는 발현 카세트로서,

여기서, 상기 발현 카세트는 포유동물 GAPDH 프로모터 또는 이의 단편을 포함하지 않는다는 것에 기반하여, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호 7인 발현 카세트.

청구항 4

제 3항에 있어서, 상기 발현 카세트는 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 추가적으로 포함하고,

여기서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호 8인 발현 카세트.

청구항 5

제 1항 내지 제 4항 중 어느 한 항에 있어서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열에 실시가능하게(operably) 연결되지 않는 발현 카세트.

청구항 6

제 1항 내지 제 4항 중 어느 한 항에 있어서, 상기 발현 카세트는 폴리아데닐레이션 자리(polyadenylation site)를 포함하는 발현 카세트.

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

제 1항 내지 제 4항 중 어느 한 항에 있어서, 상기 프로모터 및 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열은 작동가능하게 연결된(operatively linked) 발현 카세트.

청구항 22

제 1항 또는 제 4항에 있어서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 동일한 방향으로 지향되는(orientated) 발현 카세트.

청구항 23

제 1항 또는 제 4항에 있어서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 반대 방향으로 지향되는 발현 카세트.

청구항 24

제 2항 또는 제 3항에 있어서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 동일한 방향으로 지향되는 발현 카세트.

청구항 25

제 2항 또는 제 3항에 있어서, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 반대 방향으로 지향되는 발현 카세트.

청구항 26

제 1항 내지 제 4항 중 어느 한 항에 있어서, 상기 프로모터는 SV40 프로모터, MPSV 프로모터, 마우스 CMV, 인간 tk, 인간 CMV, 래트 CMV, 인간 EF1알파, 차이니즈 햄스터 EF1알파, 인간 GAPDH, MYC, HYK 및 CX 프로모터를 포함하는 하이브리드(hybrid) 프로모터로 구성된 군으로부터 선택되는 발현 카세트.

청구항 27

제 1항 내지 제 4항 중 어느 한 항에 있어서, 상기 폴리펩티드는 항체, 항체 단편(antibody fragments) 또는 항체 유도체(antibody derivates)로 구성된 군으로부터 선택되는 발현 카세트.

청구항 28

제 6항에 있어서, 상기 폴리아데닐레이션 자리는 BGH 폴리(A) 및 SV40 폴리(A)로 구성된 군으로부터 선택되는 발현 카세트.

청구항 29

제 1항 내지 제 4항 중 어느 한 항에 있어서, 추가 프로모터(additional promoter), 인핸서(enhancer), 전사조절 요소(transcriptional control elements), 및 선별가능한 마커(selectable marker)로 구성된 군으로부터 선택되는 유전적 요소(genetic element)를 추가적으로 포함하는 발현 카세트.

청구항 30

제 29항에 있어서, 상기 유전적 요소는 선별가능한 마커를 암호화하는 폴리뉴클레오티드 서열에 포함된 CpG 자리의 함량(content)이 45 또는 그 미만인 선별가능한 마커인 발현 카세트.

청구항 31

제 1항 내지 제 4항 중 어느 한 항의 발현 카세트를 포함하는 발현 벡터.

청구항 32

하기의 순서를 포함하는, 발현 벡터로서:

- a) 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- b) 프로모터
- c) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- d) 폴리아데닐레이션 자리
- e) 인핸서
- f) 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림, 또는
- a) 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- b) 인핸서
- c) 프로모터
- d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- e) 폴리아데닐레이션 자리
- f) 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림, 또는
- a) 인핸서
- b) 포유동물 GAPDH의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- c) 프로모터
- d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- e) 폴리아데닐레이션 자리
- f) 포유동물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림,

만약 a) 또는 b)가 포유동물 GAPDH의 비번역된 게놈 DNA 서열 업스트림이면 f)는 포유동물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림이고 만약 a) 또는 b)가 포유동물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림이면 f)는 포유동물 GAPDH의 비번역된 게놈 DNA 서열 업스트림인 조건으로, 상기 인핸서의 포함은 선택적이고, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호 8이고 상기 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호 7인 발현 벡터.

청구항 33

제 31항에 있어서, 상기 발현 벡터는 추가 프로모터, 인핸서, 전사 조절 요소, 복제의 기원(origin of replication) 및 선별가능한 마커로 구성된 군으로부터 선택되는 유전적 요소를 추가적으로 포함하는 발현 벡터.

청구항 34

제 31항에 있어서, 상기 발현 벡터는 복제의 기원 및 선별가능한 마커를 암호화하는 발현 벡터의 폴리뉴클레오티드 서열에 포함된 CpG 자리의 함량이 200 또는 미만인 복제의 기원 및 선별가능한 마커를 추가적으로 포함하는 발현 벡터.

청구항 35

삭제

청구항 36

제 31항의 발현 벡터를 포함하는 숙주 세포.

청구항 37

삭제

청구항 38

제 1항 내지 제 4항 중 어느 한 항의 발현 카세트 또는 제 31항 내지 제 34항 중 어느 한 항의 발현 벡터로 숙주 세포를 형질전환 및 폴리펩티드를 회수하는 것을 포함하는, 폴리펩티드 발현을 위한 시험관 내(*in vitro*) 방법.

청구항 39

제 38항에 있어서, 상기 발현 카세트는 안정적으로 형질전환되는 방법.

청구항 40

제 38항에 있어서, 상기 발현 카세트는 일시적으로 형질전환되는 방법.

청구항 41

제 1항 내지 제 4항 중 어느 한 항의 발현 카세트를 포함하는 포유동물 숙주 세포(mammalian host cell)로부터 이종폴리펩티드(heterologous polypeptide)의 발현을 위한 조성물.

청구항 42

유전자 치료를 위한 제 31항의 발현 벡터.

청구항 43

- 1) 제 31항의 발현 벡터를 가진 숙주세포를 형질전환 시키는 단계; 및
- 2) 폴리펩티드를 회수하는 단계; 를 포함하는 폴리펩티드의 발현을 위한 시험관 내 방법(*in vitro* method).

청구항 44

제 43항의 방법에 있어서, 상기 발현 벡터는 안정적으로 형질전환되는 방법.

청구항 45

제 43항의 방법에 있어서, 상기 발현 벡터는 일시적으로 형질전환되는 방법.

청구항 46

제 31항의 발현 벡터를 포함하는 포유동물 숙주 세포(mammalian host cell)로부터 이종폴리펩티드(heterologous polypeptide)의 발현을 위한 조성물.

발명의 설명

기술 분야

[0001] 본 발명은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열의 발현에 유용한 발현 카세트(expression cassette)에 관한 것이다. 또한, 본 발명은 숙주 세포로부터 폴리펩티드의 생성을 위한 발현 카세트 및 발현 카세트의 용도를 포함하는 벡터 및 숙주 세포에 관한 것이다.

[0002]

배경 기술

[0003] 재조합 폴리펩티드의 생성을 위한 발현 시스템은 당업계에서 잘 알려져 있고, 예를 들면, Marino MH (1989) Biopharm, 2: 18-33; Goeddel DV *et al.*, (1990) Methods Enzymol 185: 3-7; Wurm F & Bernard A (1999) Curr Opin Biotechnol 10: 156-159에 의해 설명되어 있다. 일반적으로, 약학적 적용에서 용도를 위한 폴리펩티드는 바람직하게는 CHO 세포, NSO 세포, SP2/0 세포, COS 세포, HEK 세포, BHK 세포 또는 등과 같은 포유동물 세포에서 생성된다. 상기 목적을 위해 사용된 발현 벡터의 필수적 요소는 복제의 원핵생물 기원(origin) 및 원핵생물 선별가능한 마커(selectable marker), 선택적으로 진핵생물 선별가능한 마커, 및 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 선택적으로 폴리아데닐레이션 신호를 포함하는 전사 종결자를 포함하는 각각의 관심의 구조적 유전자의 발현을 위한 하나 또는 그 이상의 발현 카세트를 포함하는 원핵생물 플라스미드 증식 유닛(propagation unit), 예를 들면, *E.coli*로부터 선택된다. 포유동물 세포에서 일시적 발현을 위하여, SV40 Ori 또는 OriP와 같은 복제의 포유동물 기원이 포함될 수 있다. 프로모터로서 구조적(constitutive) 또는 유발하는(inducible) 프로모터가 선택될 수 있다. 최적화된 전사를 위해서, 코작(Kozak) 서열이 5' 비번역 영역에 포함될 수 있다. mRNA 진행, 특히 mRNA 스플라이싱(splicing) 및 전사 종료(transcription termination)를 위해서, 폴리아데닐레이션 신호뿐만 아니라 구조적 유전자(엑손/인트론 조합)의 조합에 따른 mRNA 스플라이싱 신호가 포함될 수 있다. 유전자의 발현은 일시적 또는 안정 세포주를 사용하여 수행될 수 있다. 생성 세포주에서 안정 수준 및 폴리펩티드의 고발현은 재조합 폴리펩티드의 생성의 전체 진행에 중대하다. 단백질과 같은 생물 분자 및 특정 항체 또는 항체단편을 위한 요구는 지난 수년에 걸쳐 현저히 증가하고 있다. 고비용과 낮은 수율은 생물 분자의 활용성에 있어서 제한요인이 되어왔고 산업적 규모에서 원하는 생물 분자의 수율을 증가시키는 강건한 진행을 발달시키는데 주요 도전이 되어왔다. 따라서, 재조합 폴리펩티드 생성에서 고발현을 얻기 위하여 발현 벡터의 효율을 개선시키기 위한 필요성이 아직 있다.

[0004]

발명의 내용

해결하려는 과제

과제의 해결 수단

발명의 효과

도면의 간단한 설명

[0005] 도 1은 마우스 거대세포바이러스(cytomegalovirus) 프로모터(mCMV), 첫번째 인트론을 포함하는 Ig 기증 수용체 단편(IgDA), IgG1 항체 경쇄(IgG1 LC), 뇌심근염바이러스(Encephalomyocarditis)로부터 유래된 내부 리보솜 입구부위(Internal Ribosomal Entry Sites)(IRES), IgG1 항체 중쇄(IgG1 HC), 녹색 형광 단백질(GFP) 및 시미안

바이러스 40 폴리아데닐레이션 신호(poly (A))를 구성하는 리포터 발현 구조체(REP)를 나타낸다.

도 2는 형질전환 후 5일째에 CHO-S 세포에서 IgG1 항체의 일시적 발현을 나타낸다(IgG 적정량(titers)의 평균이 두 개 별개의 형질전환에 대해 플롯됨). 세포를 GAPDH_A 및 GAPDH_B 벡터(GAPDH_A 및 GAPDH_B), GAPDH 업스트림 및 다운스트림 요소 (A 및 B)가 없는 동일한 벡터 및 대조군(pGLEX41)으로서 pGLEX41 벡터를 사용하여 형질 전환하였다. 상층액에 축적된 IgG1 항체의 농도를 옥테트(Octet) 기구(Fortebio, Menlo, CA, USA)를 사용하여 결정하였다.

도 3은 HEK293 EBNA 세포에서 IgG1 항체의 발현을 나타낸다. 세포를 GAPDH_A 및 GAPDH_B 벡터(GAPDH_A 및 GAPDH_B) 및 대조군(pGLEX41)으로서 pGLEX41 벡터를 사용하여 형질전환하였다. 상층액을 옥테트(Octet) 기구를 사용하여 형질전환 후 10일째에 수확하고 분석하였다. 데이터는 벡터 당 튜브스핀에서 N = 3 독립 형질전환을 나타낸다.

도 4는 세포 풀(pools)을 사용하여 일괄(batch) 생성에 대한 발현 수준 연구를 나타낸다. 세포를 형질전환하고 안정 세포의 풀을 GAPDH_A 및 GAPDH_B 벡터 (GAPDH_A(1), GAPDH_A(2), GAPDH_B(1) 및 GAPDH_B(2)), GAPDH 업스트림 및 다운스트림 요소(A(1) 및 A(2))가 없는 동일한 벡터 및 대조군(pGLEX41)으로서 pGLEX41 벡터를 사용하여 제조하였다. 배양 7일 후에 상층액을 상층액에 축적된 항체에 대해 옥테트 기구를 사용하여 분석하였다. IgG 적정량의 평균이 각각 풀에 대해 ($\mu\text{g}/\text{ml}$) 주어졌다. 데이터는 풀 당 N = 2 일괄(batches)을 나타낸다.

도 5는 안정 형질전환 및 제한 희석에 의해 생성된 집단(populations) 상에서 발현 수준 연구를 나타낸 도이다. 세포를 GAPDH_A 및 GAPDH_B 벡터(GAPDH_A 및 GAPDH_B), GAPDH 업스트림 및 다운스트림 요소가 없는 동일한 벡터(A 및 B) 및 대조군(pGLEX41)으로서 pGLEX41 벡터를 사용하여 형질전환하였다. 안정 형질전환으로부터 클론 및 미니풀에 의해 발현된 GFP 형광의 평균 값을 형질전환 후에 14일째에 측정하였다. 세포를 96-웰 플레이트에서 선별 압력(selection pressure)하에서 배양하였다. 데이터는 벡터당 N = 48 클론 또는 미니풀을 나타낸다.

도 6은 상층액에서 IgG1 항체의 발현 상에 대한 배지 첨가제인 인슐린 및 PMA (포르볼 12-미리스테이트 13-아세테이트, 포르볼 에스테르)(phorbol 12-myristate 13-acetate, a phorbol ester)의 효과를 나타낸다. GAPDH_A 벡터(GAPDH_A)와 대조군(pGLEX41)로서 pGLEX41 벡터로 형질전환 후에 세포를 PowerCHO2 배지, 4mM Gln, +/- 인슐린 및 PowerCHO2, 4mM Gln, PMA +/- 인슐린에서 희석하였다. pGLEX41(채워진 바(filled bars)) 또는 GAPDH_A(공개 바(open bars))에 대하여 표준 배지와 비교하여 발현에서 차이가 없음을 관찰하였다.

도 7은 인간 GAPDH 유전자 자리(locus)의 개요를 나타낸다. GAPDH 유전자는 유전자 NCAPD2 및 IFFO1 옆에 있다(flanked).

도 8은 인간 GAPDH 유전자, GAPDH 업- 및 다운스트림 요소 및 GAPDH 업스트림 단편화 연구 분석을 위해 제조된 단편의 세부 사항을 나타낸다. NruI 제한 자리는 클로닝 단계를 용이하게 하기 위하여 도입되었고 게놈 5' GAPDH 업스트림 서열(이는 별표(asterisk)를 사용하여 강조됨)의 일부가 아니다. 단편의 크기는 하기와 같다: 단편 1 (서열번호: 9): 511 bps, 단편 2 (서열번호: 10): 2653 bps, 단편 3 (서열번호: 11): 1966 bps, 단편 4 (서열번호: 12): 1198 bps, 단편 8 (서열번호: 13): 259 bps, 단편 9 (서열번호: 14): 1947 bps, 단편 11 (서열번호: 15): 1436 bps, 및 단편 17 (서열번호: 16): 1177 bps.

도 9는 GAPDH 업스트림 및 다운스트림 요소의 단편화의 발현 결과를 나타낸다. 발현 결과를 형질전환 후에 10 일째에 CHO 세포에서 일시적 형질전환으로 얻었다. 정량화를 옥테트 기구를 사용하여 수행하였다. 벡터 pGLEX41은 음성 대조군으로서 이용되었다. 또한, pGLEX41-ampiA는 GAPDH 인접 요소가 없는 벡터의 기저 발현을 나타내는 음성 대조군이다. pGLEX41-업/다운은 전장 인접(업스트림 및 다운스트림) 영역을 포함하고 양성 대조군h으로서 이용되었다. pGLEX41-업은 단지 업스트림 인접 영역을 포함하고 pGLEX41-다운은 단지 다운스트림 인접 영역을 포함한다. 모든 다른 구조체는 도 8에 설명된 단편을 포함한다. 단편 2 및 3은 IgG1 LC 및 IgG1 HC로서 동일한 방향으로 클론되거나 또는 IgG1 LC 및 IgG1 HC(AS)와 관련하여 반대 방향으로 클론되었다.

도 10은 형질전환 후 8일째에 CHO-S 세포에서 IgG1 항체의 일시적 발현을 나타낸다(IgG의 적정량의 평균은 세 개의 독립 형질전환에 대하여 플롯됨; 에러바: SD +/-). 세포를 마우스 CMV(A_GAPDH_UP) 또는 차이니즈 햄스터 GAPDH 프로모터 (A_GAPDH_UP_PR)와 조합된 차이니즈 햄스터 GAPDH 업스트림 요소를 갖는 벡터를 사용하여 형질 전환하였다. 단지 마우스 CMV (A) 또는 차이니즈 햄스터 GAPDH 프로모터 (A_PR)를 갖는 플라스미드는 대조군으로서 형질전환되었다. 상층액에 축적된 IgG1 항체의 농도는 옥테트 QK 기구를 사용하여 결정되었다(Fortebio, Menlo, CA, USA).

발명을 실시하기 위한 구체적인 내용

[0006]

발명요약

[0007]

본 발명은 일반적으로 재조합 폴리펩티드 생성에서 증가된 발현을 얻기 위하여 사용될 수 있는 발현 카세트 및 발현 벡터와 같은 발현 시스템에 관한 것이다. 하나의 측면에서, 본 개시는 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 진핵생물 글리세르알데하이드 3-포스페이트 디하드로제네이스(Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)) 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하는 발현 카세트를 제공하고, 여기서 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치로부터 +7000 주위의 뉴클레오티드 위치까지 영역거리(region spanning)내에서 시작하고, 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 100 주위부터 15000 주위의 뉴클레오티드까지이다.

[0008]

추가적 측면에서, 본 개시는 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 제공하고, 여기서 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터 5'말단 주위로부터 -3500 주위의 뉴클레오티드 위치까지 영역 거리내에서 시작하고, 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 발현 카세트는 진핵생물 GAPDH 프로모터 또는 이의 단편을 포함하지 않는다는 조건으로 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100부터 15000 주위의 뉴클레오티드까지이다.

[0009]

추가적 측면에서, 본 개시는 발현 카세트를 포함하는 발현 벡터 및 발현 카세트 또는 발현 카세트를 포함하는 발현 벡터를 포함하는 숙주세포를 제공한다

[0010]

더욱 추가적 측면에서, 본 개시는 발현 카세트 또는 발현 벡터를 갖는 숙주세포 형질전환 및 폴리펩티드 회수를 포함하는, 폴리펩티드 발현을 위한 시험관 내 방법 및 포유동물 숙주 세포로부터 이중 폴리펩티드의 발현을 위한 발현 카세트 또는 발현 벡터의 용도를 제공한다.

[0011]

본 발명의 상세한 설명

[0012]

본 개시는 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열 및 진핵생물 글리세르알데하이드 3-포스페이트 디하드로제네이스(Glyceraldehyde 3-phosphate dehydrogenase(GAPDH)) 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하는 발현 카세트 및 발현 벡터에 관한 것이고, 여기서 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치로부터 +7000 주위의 뉴클레오티드 위치까지 영역거리(region spanning)내에서 시작하고, 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 100 주위부터 15000 주위의 뉴클레오티드까지이다.

[0013]

본 개시는 추가적으로 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 포함하는 발현 카세트에 관한 것이고, 여기서 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고,

[0014]

진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 5'말단 주위로부터 -3500 주위의 뉴클레오티드 위치까지 영역 거리내에서 시작하고, 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 발현 카세트는 진핵생물 GAPDH 프로모터 또는 이의 단편을 포함하지 않는다는 조건으로 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100부터 15000 주위 뉴클레오티드까지이다.

[0015]

본 명세서에서 사용된 용어 "발현 카세트"는 시스-액팅 전사 조절 요소(cis-acting transcriptional control

elements)의 임의 조합을 포함하는 발현되도록 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열 및 프로모터 및 선택적으로 인핸서(enhancer) 서열과 같은 자체의 발현을 조절하는 서열을 포함한다. 유전자 발현을 조절하는 서열은, 예를 들면, 전사 생성물의 자체의 전사 및 번역은 조절유닛(regulatory unit)으로서 일반적으로 언급된다. 조절유닛의 대부분은 유전자의 코딩 서열의 업스트림에 위치해 있고, 실시가능하도록 거기에 연결되어 있다. 발현카세트는 폴리아데닐레이션 자리를 포함하는 다운스트림 3' 비번역 영역을 포함할 수 있다. 본 발명의 조절 유닛은 전사유닛과 같은 발현되는 유전자에 실시가능하도록 연결되거나 또는 예를 들면, 이중 유전자의 5' -비번역된 영역과 같은 간섭 DNA에 의해 거기로부터 분리된다. 바람직하게는 발현 카세트는 벡터로 발현 카세트의 삽입 및/또는 벡터로부터 자체의 절제를 가능하도록 하기 위하여 하나 또는 그 이상의 적절한 제한효소에 의해 옆에 있다. 따라서, 본 발명에 따른 발현 카세트는 발현 벡터, 특히, 포유동물 발현 벡터의 구조체에 사용될 수 있다. 본 발명의 발현 카세트는 하나 또는 그 이상, 예를 들면, 진핵생물 GAPDH 프로모터 또는 이의 단편의 2개, 3개 또는 그 이상의 비번역된 게놈 DNA 서열 다운스트림, 및/또는 하나 또는 그 이상, 예를 들면, 진핵생물 GAPDH 프로모터 또는 이의 단편의 2개, 3개 또는 그 이상의 비번역된 게놈 DNA 서열 업스트림을 포함할 수 있다. 만약 본 발명의 발현 카세트가 진핵생물 GAPDH 프로모터 또는 이의 단편의 하나 이상의 DNA 서열 다운스트림 및/또는 업스트림을 포함한다면, 상기 DNA 서열은 직접적으로 연결되는, 즉, 예를 들면, 5' 및 3' 말단에 부착되고 서열 또는 이의 단편의 일련의 클로닝을 수월하게 하는 제한효소 자리를 포함하는 링커서열을 포함할 수 있다. 진핵생물 GAPDH 프로모터 또는 이의 단편은 DNA 서열 다운스트림 및/또는 업스트림은 직접적으로 연결되지 않을 수 있다, 즉, 간섭 DNA 서열로 클론될 수 있다.

[0016] 본 명세서에서 사용된 용어 "폴리펩티드를 암호화하는 폴리뉴클레오티드 서열"은 유전자, 바람직하게는 폴리펩티드를 발현하는 이중 유전자를 위한 DNA 코딩을 포함한다.

[0017] 용어 "이중 코딩 서열", "이중 유전자 서열", "이중 유전자", "재조합 유전자", 또는 "유전자"는 교체사용될 수 있다. 상기 용어는 재조합, 특히 숙주세포, 바람직하게는 포유동물 세포에서 발현되고 수확되어야만 하는 재조합 이중 단백질 생성물에 대해 코드하는 DNA 서열을 의미한다. 상기 유전자의 생성물은 폴리펩티드일 수 있다. 상기 이중 유전자 서열은 숙주 세포에 자연적으로 존재하지 않고 동일한 또는 상이한 종으로부터 유래되고 유전적으로 변형될 수 있다.

[0018] 상기 용어 "단백질" 및 "폴리펩티드"는 붙어있는 잔기의 알파-아미노와 카르복시기 사이의 펩티드 결합에 의해 다른 것과 연결된 일련의 아미노산 잔기를 포함하도록 교체될 수 있다.

[0019] 본 명세서에서 사용된 상기 용어 "비번역 게놈 DNA 서열"은 유기체의 유전자 정보를 구성하는 DNA를 포함한다. 대부분의 모든 유기체의 게놈은 DNA이고, 단지 RNA 게놈을 갖는 일부 바이러스는 예외이다. 대부분의 유기체에서 게놈 DNA 분자는 염색체라고 불리는 DNA-단백질 복합체로 구성체로 조직화되어 있다. 크기, 염색체의 수, 게놈 DNA의 성질은 다른 유기체 사이에서 다양하다. 바이러스 DNA 게놈은 단일- 또는 이중-가닥, 직선 또는 원형일 수 있다. 박테리아는 단일, 원형 염색체를 가진다. 진핵생물에서, 대부분의 게놈 DNA는 다른 크기의 다수의 직선 염색체로서 핵(핵의 DNA) 내에 위치해 있다. 진핵생물 세포는 미토콘드리아 및, 식물 및 저등 진핵생물, 클로로플라스트(chloroplast)에서 추가적으로 게놈 DNA를 포함한다. 상기 DNA는 원형분자이고 상기 세포기관 내에 다수의 복사본으로 존재한다. 비번역된 게놈 DNA 서열은 일반적으로 프로모터에 실시가능하도록 연결되어 있지 않으므로 번역되지 않는다. 상기는 예를 들면, 발현되지 않는 단백질을 암호화하는 유전자, 따라서, 번역되지 않는 유전자를 포함할 수 있다.

[0020] 본 명세서에서 사용된 상기 용어 "진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림"은 진핵생물 GAPDH 프로모터의 비번역된 진핵생물 게놈 DNA 3'에 해당한다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치, 바람직하게는 +1 뉴클레오티드 위치에서 시작하고, 여기서, 뉴클레오티드 위치는 GAPDH mRNA의 전사시작에 비례, 즉, GAPDH에 대한 진핵생물 유전자 코딩의 전사 시작의 기원(origin)에 비례한다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 항상 진핵생물 GAPDH 프로모터로서 동일한 기원이고, 예를 들면, 만약 GAPDH 프로모터가 인간 기원이면, 인간 GAPDH 프로모터

의 비번역된 게놈 DNA 서열 다운스트림은 더욱이 인간 기원이고 자연적으로 발생하는 인간 GAPDH 프로모터의 인간 게놈 DNA 서열 다운스트림에 해당함으로써, 항상 동일한 기원이다.

[0021] 본 명세서에서 사용된 용어 "진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 비번역된 진핵생물 게놈 DNA 5'에 해당한다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 일반적으로 진핵생물 GAPDH 프로모터의 5' 주위의 뉴클레오티드 위치, 바람직하게는 진핵생물 GAPDH 프로모터 5' 말단 후 즉시의 뉴클레오티드 위치에서 시작한다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터, 예를 들면, 만약 GAPDH 프로모터가 인간 기원이면, 인간 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 더욱이 인간 기원이고 자연적으로 발생하는 인간 GAPDH 프로모터의 인간 게놈 DNA 서열 업스트림에 해당함으로써, 항상 동일한 기원이다.

[0022] 본 명세서에서 나타난 진핵생물 GAPDH 프로모터의 위치, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 또는 업스트림 및 다른 DNA 서열은 GAPDH mRNA의 전사 시작에 비례, 예를 들면, 특정적으로 지적하지 않는 한, 진핵생물 GAPDH의 전사 시작의 기원에 비례한다.

[0023] 상기 용어 "진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 확장한다" 또는 "진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 확장한다"는 시작으로부터 특정 유전적 요소까지 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림의 길이 확장(extension), 예를 들면, 인트론(intron)으로 확장을 정의하는데 사용된다. 상기 확장은 유전적 요소를 암호화하는 전장의 DNA 서열, 예를 들면, 인트론 또는 이의 일부를 포함한다.

[0024] 진핵생물 GAPDH 프로모터 및 GAPDH 프로모터의 진핵생물 게놈 DNA 업스트림 and/or 다운스트림은 NCBI 공동 데이터은행(public databank)에 있는 인간, 래트 및 마우스에 대하여 발견될 수 있고(인간, 마우스, 래트 및 차이니즈 햄스터 GAPDH 유전자 가입은 각각 유전자 IDs 2597(mRNA: NM_002046.3), 14433(mRNA: NM_008084.2), 24383(mRNA: NM_017008.3) 및 100736557(mRNA: NM_001244854.2); 생물기술 정보를 위한 국가센터(NCBI): <http://www.ncbi.nlm.nih.gov/>), 인간 GAPDH 유전자에 대하여 도 7 및 8에서 예로써 나타난다.

[0025] 진핵생물 GAPDH 프로모터는 GAPDH mRNA의 전사 시작에 비례하여 bps -500 주위로부터 +50 주위까지 스트레치 되도록 항상 고려된다. 인간 GAPDH 프로모터는 염색체 12 상에 위치해 있다. 인간 GAPDH 프로모터는 단편화(fragmentation) 연구에 기반한 GAPDH mRNA의 전사 시작에 비례하여 bps -488 까지 +20 스트레치(stretch) 되도록 Graven *et al.*(Graven *et al.*, (1999) Biochimica et Biophysica Acta, 147: 203-218)에 의해 고려된다. NCBI 공동 데이터은행에 따라 인간 GAPDH 프로모터는 NCBI 공동 데이터은행에 의해 정의된 GAPDH mRNA의 전사 시작에 비례하여 bps -462부터 +46까지 스트레치한다. 특정적으로 달리 지적되지 않는 한, 본 명세서에서 언급된 인간 GAPDH 프로모터는 4071부터 서열번호: 17의 4578까지 스트레칭하는 서열에 해당하는 GAPDH mRNA의 전사 시작에 비례하여 -462부터 +46 위치까지 스트레치한다.

[0026] 본 명세서에서 언급된 인간, 마우스 및 래트 기원의 GAPDH 유전자, IFF01 유전자 및 NCAPD2 유전자의 DNA에 대해 사용된 숫자는 NCBI 공동 데이터 은행(<http://www.ncbi.nlm.nih.gov/>)에 있는 상기 유전자에 대해 사용된 숫자에 해당한다.

[0027] 본 명세서에서 사용된 용어 "프로모터"는 DNA에 결합하는 RNA 중합효소를 감속하거나 RNA 합성을 개시하여 전사의 개시를 매개하는 유전자의 업스트림에 일반적으로 위치한 조절 DNA 서열을 정의한다.

[0028] 본 명세서에서 사용된 용어 "인핸서(enhancer)"는 유전자의 정체성(identity), 유전자에 관련한 서열 위치 또는 서열의 방향(orientation)에 독립적으로 유전자의 전사를 강력하게 행하는 뉴클레오티드 서열을 정의한다.

[0029]

[0030]

상기 용어 "기능적으로 연결된(functionally linked)" 및 "실시가능하게 연결된(operably linked)"은 교체적으로 사용되고 두 개 이상의 DNA 시그먼트(segments), 특히, 발현되는 유전자 서열과 자체의 발현을 조절하는 상기 서열 사이에서의 기능적 관계를 언급한다. 예를 들면, 만약 프로모터 및/또는 인핸서 서열이 적절한 숙주 세포 또는 다른 발현 시스템에서 코딩 서열의 전사를 촉진 또는 조절하면 시스-액팅(cis-acting) 전사 조절요소의 임의 조합을 포함하는, 프로모터 및/또는 인핸서 서열은 코딩 서열에 실시가능하게 연결된다. 전사된 유전자 서열에 실시가능하게 연결된 프로모터 조절 서열은 전사된 서열에 물리적으로 근접하다.

[0031]

"방향(Orientation)"은 주어진 DNA 서열에서 뉴클레오티드 순서를 언급한다. 예를 들면, 얻어진 서열로부터의 DNA에서 참조 포인트와 비교하였을 때 다른 DNA 서열과 관련하여 반대 방향으로 DNA 서열의 방향은 다른 서열이 역방향된 것과 관련하는 서열의 5'에서 3'으로의 순서인 것이다. 상기 참조 포인트는 원천 DNA에 있는 다른 특정된 DNA 서열의 전사 방향 및/또는 서열을 포함하는 반복가능한 벡터의 복제의 기원을 포함한다.

[0032]

본 명세서에서 사용된 용어 "발현 벡터"는 적절한 숙주 세포로 형질전환이 숙주 세포 내에 재조합 유전자 생성물의 고수준의 발현을 제공하는 분리되고 정제된 DNA 분자를 포함한다. 재조합 또는 유전자 생성물에 대한 DNA 서열 코딩에 추가하여, 발현 벡터는 mRNA로 DNA 코딩 서열의 효율적 전사 및 숙주 세포주의 단백질의 mRNA의 효율적 번역에 대해 요구되는 조절 DNA 서열을 포함한다.

[0033]

본 명세서에서 사용된 용어 "숙주 세포" 또는 "숙주 세포주"는 배양에서 자라게 하고 원하는 재조합 생성 단백질을 발현하는 것이 가능한 임의의 세포, 특히 포유동물 세포를 포함한다.

[0034]

본 명세서에서 사용된 용어 "단편(fragment)"은 각각의 뉴클레오티드 서열, 예를 들면, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림의 일부 또는 프로모터와 같은 특정 유전적 요소를 암호화하는 뉴클레오티드 서열의 일부를 포함한다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림의 단편은 생물 활성을 유지할 수 있고 따라서 변경, 예를 들면, 프로모터에 실시가능하게 연결된 코딩 서열의 발현 패턴을 증가할 수 있다. 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림의 단편은 적어도 약 100부터 약 3000 bp까지, 바람직하게는 약 200부터 약 2800 bp까지, 더욱 바람직하게는 약 300부터 약 2000 bp까지 뉴클레오티드, 특히 약 500부터 약 1500까지 bp 뉴클레오티드의 범위일 수 있다. 본 발명의 발현 카세트에서 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림의 단편을 클론하기 위하여, 클로닝을 수월하게 하는 제한효소 자리를 포함하는 링커 서열은 단편의 5' 및 3' 말단에 항상 부착된다.

[0035]

본 명세서에서 사용된 용어 "뉴클레오티드 서열 정체성" 또는 "동일한 뉴클레오티드 서열"은 가능하면, 최대 퍼센트 서열 정체성(identity)을 얻기 위하여 서열 및 도입 갭(introducing gap)을 정렬시킨 후에, 예를 들면, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림의 뉴클레오티드 서열과 동일한 후보서열에서 뉴클레오티드의 퍼센트를 포함한다. 따라서, 서열의 정체성은 두개의 뉴클레오티드 서열의 뉴클레오티드 위치에 있는 유사성을 비교하기 위하여 사용되는 표준 방법에 의해 결정될 수 있다. 항상 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림에 대한 후보 서열의 뉴클레오티드 서열의 정체성은 예를 들면, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 및 100%를 포함하여, 적어도 80%, 바람직하게는 적어도 85%, 더욱 바람직하게는 적어도 90%, 및 가장 바람직하게는 적어도 95%, 특히 96%, 더욱 특히 97%, 더욱 더 특히 98%, 가장 특히 99%에 있다.

[0036]

본 명세서에서 사용된 용어 "CpG 자리"는 시토신(cytosine) 뉴클레오티드가 자체의 길이를 따라 베이스(bases)의 직선서열에서 구아닌(guanine) 뉴클레오티드 옆에서 일어나는 DNA 영역을 포함한다. "CpG"는 "-C-포스페이트-G-"에 대해 함축되고, 즉, 시토신 및 구아닌이 단지 하나의 포스페이트에 의해 분리된 것; 포스페이트는 DNA

에서 임의의 두 개의 뉴클레오타이드와 함께 연결된다. "CpG" 표기법은 시토신 및 구아닌의 CG 베이스-쌍짓기로 부터 상기 직선 서열을 구별하도록 사용된다.

[0037] 본 명세서에서 사용된 용어 "선택적 코돈 사용(alternative codon usage)"은 CpG 서열 모티프를 피하기 위하여 동일한 아미노산을 위한 선택적 코돈 코딩의 사용을 포함한다. 상기는 새로운 CpG 자리를 유도하는 두 개의 코돈 결합을 회피하는 것뿐만 아니라 내부의 CpG 자리를 가지지 않는 바람직한 코돈 사용을 포함한다(예를 들면, 알라닌(Alanine)을 위한 GCG 코딩 및 CpG 자리를 포함하는 것은 GCT, GCC 또는 GCA에 의해 교체될 수 있다)

[0038] GAPDH mRNA의 전사시작에 비례한, 예를 들면, 진핵생물의 전사 시작의 기원에 비례하는 DNA 서열의 길이와 관련 및 뉴클레오타이드 위치와 관련한 본 명세서에서 사용된 용어 "주위(around)"는 최대 ± 50 % 편차 값을 포함하고, 기술된 값, 예를 들면, "3000 뉴클레오타이드 주위"의 항상 최대 ± 10 %는 2700 내지 3300 뉴클레오타이드, 바람직하게는 2900 내지 3100 뉴클레오타이드, 더욱 바람직하게는 2995 내지 3005 뉴클레오타이드를 포함하고, "100 뉴클레오타이드 주위"는 50 내지 150 뉴클레오타이드, 바람직하게는 90 내지 110 뉴클레오타이드, 더욱 바람직하게는 95 내지 105 뉴클레오타이드의 값을 포함하고, "15000 뉴클레오타이드 주위"는 13500 내지 16500 뉴클레오타이드, 바람직하게는 14500 내지 15500 뉴클레오타이드, 더욱 바람직하게는 14990 내지 15010 뉴클레오타이드, 가장 바람직하게는 14995 내지 15005 뉴클레오타이드 값을 포함하고, "200 뉴클레오타이드 주위"는 50 to 250 뉴클레오타이드s, 바람직하게는 190 to 210 뉴클레오타이드s, 더욱 바람직하게는 195 to 205 뉴클레오타이드 값을 포함하고, "8000 뉴클레오타이드 주위"는 7200 내지 8800, 바람직하게는 7500 내지 8500 뉴클레오타이드, 더욱 바람직하게는 7990 내지 8010 뉴클레오타이드, 가장 바람직하게는 7995 내지 8005 뉴클레오타이드의 값을 포함하고, "500 뉴클레오타이드 주위"는 50 내지 550 뉴클레오타이드, 바람직하게는 475 내지 525, 더욱 바람직하게는 490 내지 510, 가장 바람직하게는 495 내지 505 뉴클레오타이드의 값을 포함하고, "5000 뉴클레오타이드 주위"는 4500 내지 5500 뉴클레오타이드, 바람직하게는 4750 내지 5250, 더욱 바람직하게는 4990 내지 5010, 가장 바람직하게는 4995 내지 5005 뉴클레오타이드의 값을 포함하고, "1000 뉴클레오타이드 주위"는 900 내지 1100 뉴클레오타이드, 바람직하게는 950 내지 1050, 더욱 바람직하게는 990 내지 1010, 가장 바람직하게는 995 내지 1005 뉴클레오타이드의 값을 포함하고, "4500 뉴클레오타이드 주위"는 4050 내지 4950 뉴클레오타이드, 바람직하게는 4250 내지 4750, 더욱 바람직하게는 4490 내지 4510, 가장 바람직하게는 4495 내지 4505 뉴클레오타이드의 값을 포함하고, "1500 뉴클레오타이드 주위"는 1350 내지 1650 뉴클레오타이드, 바람직하게는 1450 내지 1550, 더욱 바람직하게는 1490 내지 1510, 가장 바람직하게는 1495 내지 1505 뉴클레오타이드의 값을 포함하고, "4000 뉴클레오타이드 주위"는 3600 내지 4400 뉴클레오타이드, 바람직하게는 3800 내지 4200, 더욱 바람직하게는 3990 내지 4010, 더욱 바람직하게는 3995 내지 4005 뉴클레오타이드의 값을 포함하고, "2000 뉴클레오타이드 주위"는 1800 내지 2200 뉴클레오타이드, 바람직하게는 1900 내지 2100, 더욱 바람직하게는 1990 내지 2010, 가장 바람직하게는 1995 내지 2005 뉴클레오타이드의 값을 포함하고, "2700 뉴클레오타이드 주위"는 2430 내지 2970 뉴클레오타이드, 바람직하게는 2600 내지 2800, 더욱 바람직하게는 2690 내지 2710, 가장 바람직하게는 2695 내지 2705 뉴클레오타이드의 값을 포함하고, "3300 뉴클레오타이드 주위"는 2970 내지 3630 뉴클레오타이드s 바람직하게는 3100 내지 3500, 더욱 바람직하게는 3290 내지 3310, 가장 바람직하게는 3295 내지 3305 뉴클레오타이드의 값을 포함하고, "3200 뉴클레오타이드 주위"는 2880 내지 3520 뉴클레오타이드s, 바람직하게는 3000 내지 3400, 더욱 바람직하게는 3190 내지 3210, 가장 바람직하게는 3195 내지 3205 뉴클레오타이드의 값을 포함하고, +7000 주위 또는 위치 +7000 주위는 위치 +6300 내지 +7700, 바람직하게는 위치 +6700 내지 +7300, 더욱 바람직하게는 위치 +6990 내지 +7010, 가장 바람직하게는 위치 +6995 내지 +7005를 포함하고, +1 주위 또는 위치 +1 주위는 위치 -10 내지 +10, 바람직하게는 위치 -5 내지 +5, 더욱 바람직하게는 위치 -1 내지 +2를 포함하고, -3500 주위 또는 위치 -3500 주위는 위치 -3150 내지 -3850, 바람직하게는 위치 -3300 내지 -3700, 더욱 바람직하게는 위치 -3490 내지 -5010, 가장 바람직하게는 위치 -3495 내지 -3505를 포함한다.

[0039] SEQ ID 번호의 서열에서 위치와 관련하여 본 명세서에서 언급되거나 사용된 인간, 마우스 및 래트 기원의 GAPDH 유전자, IFFO1 유전자 및 NCAPD2 유전자의 DNA에 사용된 번호와 관련하여 본 명세서에서 사용된 용어 "주위(round)"는 최대 ± 500 bps, 바람직하게는 ± 100 bps, 더욱 바람직하게는 ± 10 bps, 가장 바람직하게는 ± 5 bps의 편차 값을 포함한다.

[0040] 일 실시형태에서, 본 개시는 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열, 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하는 발현 카세트를 제공하고, 여기서, 폴리뉴클레오타이드

서열에 의해 암호화되는 폴리뉴클레오티드 서열은 GAPDH가 아니고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치부터 +7000 주위의 뉴클레오티드 위치까지의 영역거리 내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 100 주위부터 15000 주위 뉴클레오티드까지이다.

[0041] 일 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 적어도 100 뉴클레오티드 주위이고 자체 최대에서 IFF01 유전자의 두 번째 인트론 또는 이의 일부까지 연장한다. 일 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 적어도 100 뉴클레오티드 주위이고 자체 최대에서 IFF01 유전자의 마지막 인트론까지 연장한다.

[0042] 인간 IFF01 유전자는 염색체 12 (NCBI 유전자 ID: 25900)의 bps 6665249 내지 6648694 주위의 인간 DNA에 위치해 있다. 일 실시형태에서, 인간의 IFF01 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 인간의 IFF01 유전자(위치 +7021)에 대한 염색체 12 코딩의 bps 6650677 주위까지 자체의 최대에서 스트레치한다. 일 실시형태에서, 인간의 IFF01 유전자의 두 번째 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 인간(위치 + 13574)의 IFF01 유전자에 대한 염색체 12 코딩의 bps 6657230 주위까지 자체의 최대에서 스트레치한다. 각각 인간의 IFF01 유전자의 마지막 인트론 및 IFF01 유전자의 두 번째 마지막까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 염색체 12 (NCBI 유전자 ID: 25900)의 bps 6657230 내지 6639125를 나타내는 서열번호: 17에 포함된다. 상기 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 17에 의해 나타낸 뉴클레오티드 서열의 bps 11553 주위까지 스트레치하고 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 17에 의해 나타낸 뉴클레오티드 서열의 bps 18106 주위까지 스트레치한다.

[0043] 마우스 IFF01 유전자 (NCBI 유전자 ID: 320678)는 염색체 6의 bps 125095259 내지 125111800 주위의 마우스 DNA에 위치해 있다. 일 실시형태에서, 마우스의 IFF01 유전자의 마지막 인트론까지 자체 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 마우스의 IFF01 유전자 (위치 + 6391)에 대한 염색체 6 코딩의 bps 125109211 주위의 자체 최대에서 스트레치한다. 일 실시형태에서, 마우스의 IFF01 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 마우스 IFF01 유전자(위치 +12081)에 대한 염색체 6 코딩의 bps 125103521 주위까지 자체의 최대에서 한다. 마우스의 IFF01 유전자의 마지막 인트론 및 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 염색체 6(NCBI 유전자 ID: 320678)의 bps 125103521 내지 125119832를 나타내는 서열번호: 18에서 각각 포함된다. 마우스의 IFF01 유전자의 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 18에 의해 나타내는 뉴클레오티드 서열의 bps 10622 주위까지 스트레치하고 마우스의 IFF01 유전자의 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 18에 의해 나타내는 뉴클레오티드 서열의 bps 16312 주위까지 스트레치한다.

[0044] 래트 IFF01 유전자(NCBI 유전자 ID: 362437)는 염색체 4의 bps 161264966 내지 161282150 주위의 래트 DNA에 위치해 있다. 일 실시형태에서, 래트의 IFF01 유전자의 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 래트의 IFF01 유전자(위치 + 5154)에 대한 염색체 4 코딩의 bps 161280937 주위까지 자체의 최대에서 스트레치한다. 일 실시형태에서, 래트의 IFF01 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 래트의 IFF01 유전자(위치 +6640)에 대한 염색체 4 코딩의 bps 161279451 주위까지 자체의 최대에서 스트레치한다.

[0045] 래트의 IFF01 유전자의 마지막 인트론 및 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH

프로모터의 비번역된 게놈 DNA 서열 다운스트림은 염색체 4(NCBI 유전자 ID: 362437)의 bps 161279451 내지 161290508를 나타내는 서열번호: 19에서 각각 포함된다. IFF01 유전자의 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 19에 의해 나타내는 뉴클레오티드 서열의 bps 9572 주위까지 스트레치하고 IFF01 유전자의 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 19에 의해 나타내는 뉴클레오티드 서열의 bps 11058까지 스트레치한다.

[0046] 차이니즈 햄스터 IFF01 유전자(NCBI 유전자 ID: 100753382)는 bps 3577293 내지 3593683 주위의 차이니즈 햄스터 DNA에 위치해 있다. 일 실시형태에서, 차이니즈 햄스터에서 IFF01 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 차이니즈 햄스터 (위치 +6883)에서 IFF01 유전자에 대한 bps 3579014 코딩 주위까지 자체의 최대에서 스트레치한다. 일 실시형태에서, 차이니즈 햄스터의 IFF01 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 차이니즈 햄스터의 IFF01 유전자(위치 +12930)에 대한 bps 3585061 코딩 주위에 자체의 최대에서 스트레치한다. 염색체의 위치는 염색체의 데이터뱅크에서 아직 주석이 달리지 않았고 현재 서열정보는 많은 알려지지 않은 베이스를 포함한다. 그러므로 한계의 정확한 표기법은 더 많은 정확한 서열 정보의 활용가능성으로 변할 수 있다.

[0047] 차이니즈 햄스터의 IFF01 유전자의 마지막 및 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 bps 3567932 내지 3585061를 나타내는 서열번호: 29에 각각 포함된다. IFF01 유전자의 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 29에 의해 나타낸 바와 같이 뉴클레오티드 서열의 bps 11083 주위까지 스트레치하고 IFF01 유전자의 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 29에 의해 나타낸 바와 같이 뉴클레오티드 서열의 bps 17130 주위까지 스트레치한다.

[0048] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 진핵생물 GAPDH 폴리아데닐레이션 자리에서 시작하고, 예를 들면, 진핵생물 GAPDH 폴리아데닐레이션 자리를 암호화하는 첫 번째 뉴클레오티드에서 시작한다. 바람직하게는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 진핵생물 GAPDH 폴리아데닐레이션 자리의 다운스트림에서 시작하고, 예를 들면, 진핵생물 GAPDH 폴리아데닐레이션 자리를 암호화하는 마지막 뉴클레오티드 후에 즉시 시작한다. 더욱 더 바람직하게는, 진핵생물 GAPDH 프로모터의 진핵생물 GAPDH 프로모터는 진핵생물 GAPDH 폴리아데닐레이션 자리의 다운스트림에서 시작하고 진핵생물 GAPDH 프로모터는 비번역된 게놈 DNA 서열 다운스트림의 길이는 적어도 100 뉴클레오티드 주위이고 IFF01 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장한다.

[0049] 일 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +3881 주위의 뉴클레오티드 위치로부터 +5000 주위의 뉴클레오티드 위치까지 내에서 시작하고, 바람직하게는 +3931 주위의 뉴클레오티드 위치로부터 +5000 주위의 뉴클레오티드 위치까지 영역거리 내에서, 더욱 바람직하게는 +4070 주위의 뉴클레오티드 위치로부터 +5000 주위의 뉴클레오티드 위치까지 영역거리 내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례한다.

[0050] 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은, 예를 들면, 본 발명에서 사용된 진핵생물 GAPDH 폴리아데닐레이션 자리의 다운스트림은 항상 위치 +3931 주위의 뉴클레오티드 위치에서 시작하고, 바람직하게는 +4070 주위의 뉴클레오티드 위치에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례한다.

[0051] 인간 GAPDH 폴리아데닐레이션 자리의 인간 비번역된 게놈 DNA 서열 다운스트림은 뉴클레오티드 위치 +3931(서열 번호: 17에서 나타낸 바와 같이 bp 8463에 해당하는 GAPDH mRNA의 전사 시작에 비례하여) 주위에서 시작한다. 바람직하게는, 만약 GAPDH 폴리아데닐레이션 자리의 비번역된 게놈 DNA 서열 다운스트림은 인간으로부터 이면,

GAPDH 폴리아데닐레이션 자리의 비번역된 게놈 DNA 서열 다운스트림은 +3931 주위에서 시작하고(서열번호: 17에 나타낸 바와 같이 bp 8463에 해당하는 GAPDH mRNA의 전사 시작에 비례하여) 자체의 길이는 서열번호: 17에서 나타낸 바와 같이 bps 8463 주위로부터 11819 주위까지의 서열에 해당하는 3357 bps 주위이고 더욱 바람직하게는 이것은 +4070 주위에서 시작하고(서열번호: 17에서 나타낸 바와 같이 bps 8602에 해당하는 GAPDH mRNA의 전사 시작에 비례하여) 자체의 길이는 서열번호: 17에 나타낸 바와 같이 bps 8602 주위부터 11819 주위까지의 서열에 해당하는 3218 bps 주위이다.

[0052] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 8 및 21 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열을 포함한다.

[0053] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 8 및 21 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상보적인 뉴클레오티드 서열을 포함한다.

[0054] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 8 및 21 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열을 포함한다.

[0055] 일부 실시형태에서, 서열번호: 8 및 21 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열은 5 개 또는 미만, 4개 또는 미만, 더욱 바람직하게는 3개 또는 미만, 가장 바람직하게는 2개 또는 미만, 특히, 1 개 핵산 변형, 상기 핵산 변형은 바람직하게는 핵산 치환이다.

[0056] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 바람직하게는 200 주위부터 8000 주위까지 뉴클레오티드이고, 더욱 바람직하게는 500 주위부터 5000 주위까지 뉴클레오티드이고, 더욱 더 바람직하게는 1000 주위부터 4500 주위까지 뉴클레오티드이고, 가장 바람직하게는 1500 주위부터 4000 주위까지 뉴클레오티드이고, 특히 2000 주위부터 3500 주위까지 뉴클레오티드, 더욱 특히 2700 주위부터 3300 주위까지, 더욱 더 특히 3200 주위, 가장 특히 3218 뉴클레오티드이다. 본 명세서에서 정의된 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 비번역된 게놈 DNA 서열에 첨가된 임의 링커 서열을 포함한다.

[0057] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향되었다.

[0058] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열에 관하여 반대 방향으로 지향되었다.

[0059] 일부 실시형태에서, 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하는 발현 카세트는 추가적으로 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 포함하고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 5' 말단 주위부터 -3500 주위의 뉴클레오티드 위치까지 영역내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100 주위부터 15000 주위까지 뉴클레오티드이다.

[0060] 다른 실시형태에서, 발현 카세트는 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 진핵생물

GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하고, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 5' 말단 주위부터 -3500 주위의 뉴클레오티드 위치까지 영역거리 내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 발현 카세트가 진핵생물 GAPDH 프로모터 또는 이의 단편을 포함하지 않는 것에 기반하여 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100부터 15000 주위의 뉴클레오티드까지이다.

[0061] 일부 실시형태에서, 상기 발현 카세트는 추가적으로 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림을 포함하고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치부터 +7000 주위의 뉴클레오티드 위치까지의 영역거리 내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사 시작에 비례하고 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 100 주위부터 15000 주위까지의 뉴클레오티드이다. 상기 실시형태에서 사용된 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은, 예를 들면, 상기에 설명되어 있다.

[0062] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 바람직하게는 200 주위부터 8000주위까지 뉴클레오티드이고, 더욱 바람직하게는 500 주위부터 5000 주위까지 뉴클레오티드이고, 더욱 더 바람직하게는

[0063] 1000 주위부터 4500 주위의 뉴클레오티드, 가장 바람직하게는 1500 주위부터 4000 주위의 뉴클레오티드, 특히 길이에서 2000 주위부터 3500 주위의 뉴클레오티드, 더욱 특히 2700 주위부터 3300 주위까지, 더욱 더 특히 3200 주위, 가장 특히 3158 뉴클레오티드이다. 본 명세서에서 정의된 바와 같이 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 비번역된 게놈 DNA 서열에 첨가된 임의 링커 서열을 포함하지 않는다.

[0064] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 적어도 100 뉴클레오티드 주위이고 NCAPD2 유전자의 시작 코돈까지 자체의 최대에서 확장한다. 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 적어도 100 뉴클레오티드 주위이고 NCAPD2 유전자의 세 번째 마지막 인트론까지 자체의 최대에서 확장한다. 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 적어도 100 뉴클레오티드 주위이고 NCAPD2 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장한다. 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 적어도 100 뉴클레오티드 주위이고 NCAPD2 유전자의 마지막 인트론까지 자체의 최대에서 확장한다.

[0065] 인간 NCAPD2 유전자 (NCBI 유전자 ID: 9918)는 염색체 12의 bps 6603298 내지 6641132 주위의 인간 DNA에 위치해 있다. 일실시형태에서, 인간 NCAPD2 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 인간의 NCAPD2 유전자에 대한 염색체 12 코딩의 6640243 bps 주위까지 자체의 최대에서 스트레치한다(서열번호: 17에서 bp 1119에 해당하는 GAPDH 유전자의 전사 시작에 비례하여 -3414위치).

[0066] 일실시형태에서, 인간의 NCAPD2 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림이 길이는 인간의 NCAPD2 유전자에 대한 염색체 12 코딩의 6639984 bps 주위까지 자체의 최대에서 스트레치한다(서열번호: 17에서 bp 860에 해당하는 GAPDH 유전자의 전사 시작에 비례하여 -3673 위치).

[0067] 일실시형태에서, 인간의 NCAPD2 유전자의 세 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 인간의 NCAPD2 유전자에 대한 염색체 12 코딩의 6639125 bps 주위까지 자체의 최대에서 스트레치한다(서열번호: 17에서 bp 1에 해당하는 GAPDH 유전자의 전사 시작에 비

레하여 -4532 위치).

- [0068] 인간의 NCAPD2 유전자의 마지막 인트론, 두 번째 마지막 인트론 및 세 번째 마지막 인트론까지, 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 17에 각각 포함되고, 이는 염색체 12의 bps 6657230 내지 6639125(NCBI 유전자 ID: 9918)를 나타낸다.
- [0069] 마우스 NCAPD2 유전자(Gene ID: 68298)는 염색체 6의 125118025 내지 125141604 위치 주위의 마우스 DNA에 위치해 있다. 일 실시형태에서, 마우스의 NCAPD2 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는(전사 시작의 500 bps 업스트림의 길이를 가지는 것으로 추정됨) 마우스의 NCAPD2 유전자에 대한 염색체 6 코딩의 bps 125118607 주위까지 자체의 최대에서 스트레치한다.
- [0070] 일 실시형태에서, 마우스의 NCAPD2 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 마우스의 NCAPD2 유전자에 대한 염색체 6 코딩의 125118880 bps 주위까지 자체의 최대에서 스트레치한다.
- [0071] 일 실시형태에서, 마우스의 NCAPD2 유전자의 세 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 마우스의 NCAPD2 유전자에 대한 염색체 6 코딩의 125119832 bps 주위까지 자체의 최대에서 확장한다.
- [0072] 마우스의 NCAPD2 유전자의 마지막 인트론, 두 번째 마지막 인트론 및 세 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 18에 각각 포함되고, 이는 염색체 6의 bps 125103521 내지 125119832(NCBI 유전자 ID: 68298)를 나타낸다. 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 18(마우스 GAPDH mRNA의 전사 시작에 비례하여 -3006)에 의해 나타낸 바와 같이 뉴클레오타이드 서열의 bp1226 주위까지 스트레치한다. 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 18에 의해 나타낸 바와 같이 뉴클레오타이드 서열의 bps 953 주위까지 스트레치한다(마우스 GAPDH mRNA의 전사 시작에 비례하여 -3279). 세 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 18에 나타낸 바와 같이 뉴클레오타이드 서열의 bp 1 주위까지(마우스 GAPDH mRNA의 전사 시작에 비례하여 -4231)스트레치한다.
- [0073] 래트 NCAPD2 유전자(Gene ID: 362438)는 염색체 4의 161288671 내지 161310417위치 주위의 진핵생물 DNA에 위치해 있다. 일 실시형태에서, 래트의 NCAPD2 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 래트의 NCAPD2 유전자에 대한 염색체 4 코딩의 161289191 bps 주위까지 자체의 최대에서 스트레치한다. 일 실시형태에서, 래트의 NCAPD2 유전자의 두 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 래트의 NCAPD2 유전자에 대한 염색체 4 코딩의 161289446 bps 주위까지 자체의 최대에서 스트레치한다. 일 실시형태에서, 래트의 NCAPD2 유전자의 세 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 래트의 NCAPD2 유전자에 대한 염색체 4 코딩의 161290508 bps 주위까지 자체의 최대에서 스트레치한다.
- [0074] 래트의 NCAPD2 유전자의 마지막 인트론, 두 번째 마지막 인트론 및 세 번째 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 19에 각각 포함되고, 염색체 4의 bps 161279451 내지 161290508(NCBI 유전자 ID: 362438)를 나타낸다. 마지막 인트론을 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 19에 나타낸 바와 같이 뉴클레오타이드 서

열의 bps 1318 주위까지 스트레치한다(랫 GAPDH mRNA의 전사 시작에 비례하여 -3101). 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 19에 의해 나타낸 바와 같은 뉴클레오티드 서열의 bps 1063 주위까지 스트레치한다(랫 GAPDH mRNA의 전사시작에 비례하여 -3356 위치). 세 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 19에 의해 나타낸 바와 같은 뉴클레오티드 서열의 bp 1 주위까지 스트레치한다(랫 GAPDH mRNA의 전사 시작에 비례하여 -4418 위치).

[0075] 차이니즈 햄스터 NCAPD2 유전자(Gene ID: 100753087)는 3544184 내지 3569879 위치 주위의 진핵생물 DNA에 위치해 있다. 염색체의 위치는 NCBI 데이터베이스 상에서 활용되지 않는다. 일 실시형태에서, 차이니즈 햄스터의 NCAPD2 유전자의 마지막 인트론까지 자체의 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 차이니즈 햄스터에서 3569380 bps 주위에서 자체 최대에서 스트레치한다. 일 실시형태에서, 차이니즈 햄스터의 NCAPD2 유전자의 두 번째 마지막 인트론까지 자체 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 차이니즈 햄스터의 3569131 bps 주위까지 자체 최대에서 스트레치한다. 일 실시형태에서, 차이니즈 햄스터의 NCAPD2 유전자의 세 번째 마지막 인트론까지 자체 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 차이니즈 햄스터의 3567932 bps 주위까지 자체 최대에서 스트레치한다.

[0076] 차이니즈 햄스터의 NCAPD2 유전자의 마지막 인트론, 두 번째 마지막 인트론 및 세 번째 마지막 인트론까지 자체 최대에서 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 각각 서열번호: 29에 포함되고, 이는 bps 3567932 내지 3585061를 나타낸다. 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 29에 의해 나타낸 바와 같이 뉴클레오티드 서열의 bps 1449 주위까지 스트레치한다(차이니즈 햄스터 GAPDH mRNA의 전사시작에 비례하여 -2752). 두 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 29(차이니즈 햄스터 GAPDH mRNA의 전사시작에 비례하여 -3001 위치)에 의해 나타낸 바와 같은 뉴클레오티드 서열의 bps 1200 주위까지 스트레치한다. 세 번째 마지막 인트론까지 확장하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 서열번호: 29에 의해 나타낸 바와 같은 뉴클레오티드 서열의 bp 1까지 스트레치한다(차이니즈 햄스터 GAPDH mRNA의 전사시작에 비례하여 -4200 위치).

[0077] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 -500 주위 위치의 뉴클레오티드부터 -3500주위 위치의 뉴클레오티드까지 영역거리 내, 바람직하게는 -576 주위 위치의 뉴클레오티드부터 -3500 주위 위치의 뉴클레오티드까지 영역거리 내에서 항상 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사시작에 비례한다.

[0078] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 -500 주위 뉴클레오티드 위치에서, 바람직하게는 -576 주위 뉴클레오티드 위치에서 항상 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사시작에 비례한다.

[0079] 인간에서 인간 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 뉴클레오티드 위치 -463 주위에서 시작한다(서열번호: 17에 나타낸 바와 같이 bp 4533에 해당하는 GAPDH mRNA의 전사시작에 비례하여). 바람직하게는, 만약 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림이 인간으로부터이면, GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 -500 주위에서 시작한다(서열번호: 17에 나타낸 바와 같이 bp 4533에 해당하는; GAPDH mRNA의 전사시작에 비례하여). 더욱 바람직하게는, 만약 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림이 인간으로부터이면, GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 -576 주위에서 시작하고(GAPDH mRNA의 전사시작에 비례하여; 서열번호: 17에 나타낸 바와 같이 bp 4533에 해당하는) 자체의 길이는 서열번호: 17에 나타낸 바와 같이 bps 800 주위부터 3957주위까지 서열에 해당하는 3158 bps 주위이다.

- [0080] 추가 실시형태에서, 진행생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 및 28 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열, 바람직하게는 서열번호: 7, 9, 10, 11, 12, 13, 14, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열, 또는 서열번호: 20, 22, 23, 24, 25, 26, 27, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열을 포함한다. 더욱 바람직한 것은 서열번호: 10, 12, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열, 더욱 바람직하게는 서열번호: 10, 12, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열이고, 상기 서열번호: 10 및/또는 16을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 관련하여 반대 방향으로 지향되고, 서열번호: 12 및/또는 15를 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향된다. 마찬가지로 더욱 바람직한 것은 서열번호: 23, 25, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열, 더욱 바람직하게는 서열번호: 23, 25, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열이고, 상기 서열번호: 23 및/또는 16을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 관련하여 반대 방향으로 지향되고 서열번호: 25 및/또는 28을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향된다.
- [0081] 추가 실시형태에서, 진행생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 및 28 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상호보완라는 뉴클레오티드 서열, 바람직하게는 서열번호: 7, 9, 10, 11, 12, 13, 14, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상호보완적인 뉴클레오티드 서열, 또는 서열번호: 20, 22, 23, 24, 25, 26, 27, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상호보완적인 뉴클레오티드 서열을 포함한다. 더욱 바람직한 것은 서열번호: 10, 12, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상호보완되는 뉴클레오티드 서열이다. 마찬가지로 더욱 바람직한 것은 서열번호: 23, 25, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 상호보완적인 뉴클레오티드 서열이다.
- [0082] 추가 실시형태에서, 진행생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 서열번호: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 및 28 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열을 포함하고, 바람직하게는 서열번호: 7, 9, 10, 11, 12, 13, 14, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열, 또는 서열번호: 20, 22, 23, 24, 25, 26, 27, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열을 포함한다. 더욱 바람직한 것은 서열번호: 10, 12, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 80% 일치하는 뉴클레오티드 서열, 더욱 바람직하게는 서열번호: 10, 12, 15 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 80% 일치하는 뉴클레오티드 서열이고, 상기 서열번호: 10 및/또는 16을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 관련하여 반대 방향으로 지향되고, 서열번호: 12 및/또는 15를 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향된다. 마찬가지로 더욱 바람직한 것은 서열번호: 23, 25, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열이고, 더욱 바람직하게는 서열번호: 23, 25, 28 및 16 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열에 적어도 80% 일치하는 뉴클레오티드 서열이고, 상기 서열번호: 23 및/또는 16을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열과 관련하여 반대 방향으로 지향되고, 서열번호: 25 및/또는 28을 포함하는 뉴클레오티드 서열은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향된다.
- [0083] 일부 실시형태에서, 서열번호: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 및 28 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열은 5개 또는 미만, 바람직하게는 4개 또는 미만, 더욱 바람직하게는 3개 또는 미만, 가장 바람직하게는 2개 또는 미만, 특히 핵산 변형이고, 상기 핵산 변형은 바람직하게는 핵산 치환이다.

- [0084] 일부 실시형태에서, 서열번호: 7, 9, 11, 14, 20, 22, 24 및 27 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열은 5개 또는 미만, 바람직하게는 4개 또는 미만, 더욱 바람직하게는 3개 또는 미만, 가장 바람직하게는 2개 또는 미만, 특히 핵산 변형을 포함하고, 상기 핵산 변형은 바람직하게는 핵산 치환이다.
- [0085] 일부 실시형태에서, 서열번호: 7, 9, 11, 14, 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열은 서열번호: 7, 9, 11, 14의 뉴클레오티드 서열의 시작에 비례하여 16 위치에서 하나의 핵산 치환을 포함한다. 바람직하게는 뉴클레오티드 서열의 시작에 비례하는 16 위치에서 G는 T로 대체될 수 있다.
- [0086] 일부 실시형태에서, 서열번호: 20, 22, 24 및 27 또는 이의 단편으로 구성된 군으로부터 선택되는 뉴클레오티드 서열은 서열번호: 20, 22, 24 및 27의 뉴클레오티드 서열의 시작에 비례하여 13 위치에서 하나의 핵산 치환을 포함한다. 바람직하게는 뉴클레오티드 서열의 전사에 비례하여 13 위치의 G는 T로 대체된다.
- [0087] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열로서 동일한 방향으로 지향된다.
- [0088] 추가 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열에 비례하여 반대 방향으로 지향된다.
- [0089] 바람직한 실시형태에서, 발현 카세트는 상기에서 설명된 바와 같이 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 포함한다. 바람직하게는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 기원은 동일하고, 예를 들면, 동일한 종(species)이다. 더욱 바람직하게는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림, 진핵생물 GAPDH 프로모터 및 숙주 세포의 비번역된 게놈 DNA 서열 업스트림의 기원은 동일하고, 예를 들면, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및 숙주 세포의 기원은 동일한 포유동물이고, 예를 들면, 인간으로부터이다.
- [0090] 일부 실시형태에서, 만약 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림이 하나의 종으로부터 비번역된 게놈 DNA 서열이면, 발현 카세트의 프로모터는 동일한 동일한 종으로부터의 GAPDH 프로모터가 아니다.
- [0091] 일부 실시형태에서, 만약 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림이 인간 기원의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림이면, 발현 카세트의 프로모터는 인간 GAPDH 프로모터가 아니다.
- [0092] 일부 실시형태에서, 발현 카세트의 프로모터는 GAPDH 프로모터가 아니다.
- [0093] 일 실시형태에서, 만약 발현 카세트가 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 및 진핵생물 글리세르알데하이드 3-포스페이트 디하이드로제네스(Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)) 프로모터를 포함한다면, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치부터 +7000 주위의 뉴클레오티드 위치까지 영역거리 내에서 시작하고, 뉴클레오티드 위치는 GAPDH mRNA의 전사시작에 비례하고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 100 주위부터 15000 주위의 뉴클레오티드까지이

고, 상기 발현 카세트는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림을 추가적으로 포함하고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 5' 말단 주위부터 -3500 주위의 뉴클레오타이드 위치까지의 영역거리 내에서 시작하고, 상기 뉴클레오타이드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100 주위부터 15000 주위의 뉴클레오타이드까지이면, 발현 카세트의 프로모터는 진핵생물 GAPDH 프로모터, 바람직하게는 포유동물 GAPDH 프로모터, 더욱 바람직하게는 설치류(설치류) 또는 인간 GAPDH 프로모터일 수 있다. 상기 실시형태에서, 진핵생물 GAPDH 프로모터의 5' 말단 주위부터 -3500 주위의 뉴클레오타이드 위치까지 영역거리 내에서 시작하는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 바람직하게는 진핵생물 GAPDH 프로모터의 업스트림에 직접적으로 위치해 있고, 더욱 바람직하게는 상기 실시형태에서 발현 카세트는 진핵생물 GAPDH 프로모터를 포함하는 자연적으로 발생하는 게놈 DNA 서열 및 -3500 주위의 뉴클레오타이드 위치까지 확장을 포함하고, 상기 뉴클레오타이드 위치는 GAPDH mRNA의 전사 시작에 비례한다

[0094] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 포유동물 기원이고, 예를 들면, 진핵생물 GAPDH 프로모터는 포유동물 GAPDH 프로모터이고 포유동물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 본 명세서에서 설명된 바와 같이 사용될 수 있다.

[0095] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 설치류 또는 인간 기원이고, 예를 들면, 진핵생물 GAPDH 프로모터는 설치류 또는 인간 GAPDH 프로모터이고 설치류 또는 인간 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 본 명세서에서 설명된 바와 같이 사용된다.

[0096] 바람직하게는 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 인간, 래트 또는 마우스 기원, 더욱 바람직하게는 인간 또는 마우스 기원으로부터, 가장 바람직하게는 인간 기원으로부터 선택된다.

[0097] 일부 실시형태에서, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림은 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열에 실시가능하게 연결되지 않는다.

[0098] 일부 실시형태에서, 발현 카세트는 폴리아데닐레이션 자리를 포함한다. 바람직하게는 폴리아데닐레이션 자리는 SV40 poly(A) 및 BGH(Bovine Growth Hormone) poly(A)로 구성된 군으로부터 선택된다.

[0099] 일부 실시형태에서, 발현 카세트의 프로모터 및 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열은 실시가능하도록 연결된다.

[0100] 일부 실시형태에서, 발현 카세트의 프로모터는 SV40 프로모터, 인간 tk 프로모터, MPSV 프로모터, 마우스 CMV, 인간 CMV, 래트 CMV, 인간 EF1알파, 차이니스 햄스터 EF1알파, 인간 GAPDH, MYC, HYK 및 CX 프로모터를 포함하는 하이브리드 프로모터로 구성된 군으로부터 선택된다.

[0101] 일부 실시형태에서, 발현 카세트에 의해 암호화되는 폴리펩티드는 비당화(non-glycosylated) 및 당화된(glycosylated) 폴리펩티드일 수 있다. 당화된 폴리펩티드는 적어도 하나의 올리고사카라이드(oligosaccharide) 사슬을 갖는 폴리펩티드를 의미한다.

[0102] 비당화된 단백질의 예로는, 예를 들면, 비당화된 호르몬; 비당화된 효소; 신경 성장 요인(NGF) 페밀리, 상피 성

장 요인(EGF) 및 섬유모세포 성장요인(fibroblast growth factor(FGF))페밀리의 비당화된 성장 요인 및 호르몬 및 성장요인에 대한 비당화된 수용체이다.

[0103] 당화된 단백질의 예로는 호르몬 및 호르몬 방출요인, 응고요인(clotting factors), 항응고 요인, 호르몬 또는 성장요인에 대한 수용체, 신경영양성 요인 사이토카인(neurotrophic factors cytokines) 및 이의 수용체, T-세포 수용체, 표면 메브린 단백질, 운반 단백질, 귀소수용체(homing receptors), 어드레신(addressins), 조절 단백질, 항체, 면역접합체(immunoadhesins)와 같은 키메릭 단백질, 및 당화된 단백질의 임의 단편이다. 바람직하게는 항체, 항체 단편 또는 항체 유도체(예를 들면, Fc 융합 단백질 및 양특이성 항체(bispecific antibodies)와 같은 특정 항체 포맷)로 구성된 군으로부터 선택된다. 본 명세서에서 사용된 항체 단편은 (i) 도메인, (ii) Fab' 및 Fab'-SH를 포함하여, VL, VH, CL 또는 CK 및 CH1 도메인을 구성하는 Fab 단편, (iii) VH 및 CH1 도메인을 구성하는 Fd 단편, (iv) 단일 가변 도메인(variable domain)으로 구성된 dAb 단편(ard ES *et al.*, (1989) Nature, 341(6242): 544-6) (v) F(ab')₂ 단편, 2 개 연결된 Fab 단편을 포함하는 이가단편(bivalent fragment) (vi) 단일 사슬 Fv 분자(scFv), 상기 VH 도메인 및 VL 도메인은 항원결합 자리를 형성하기 위해 연결된 두 개의 도메인을 가능하게 하는 결합 링커에 의해 연결된다(Bird RE *et al.*, (1988) Science, 242(4877): 423-6; Huston JS *et al.*, (1988) Proc Natl Acad Sci U S A, 85(16): 5879-83), (vii) "이중체(diabodies)" 또는 "삼중체(triabodies)", 다가(multivalent) 또는 유전자 융합에 의한 다중특정 단편 구조체(Holliger P *et al.*, (1993) Proc Natl Acad Sci U S A, 90(14): 6444-8; Holliger P *et al.*, (2000) Methods Enzymol, 326: 461-79), (viii) scFv, 이중체 또는 Fc 영역에 융합된 도메인 항체 및 (ix) 동일한 또는 상이한 항체에 융합된 scFv를 포함하나, 이에 한정하지 않는다.

[0104] 일부 실시형태에서 발현 카세트는 프로모터, 인핸서, 전사 대조군 요소, 및 선별가능한 마커, 바람직하게는 동물세포에서 발현되는 선별가능한 마커로 구성된 군으로부터 선택되는 유전적 요소를 추가적으로 포함한다. 전사 대조군 요소, 예를 들면, 코작(Kozak) 서열 또는 전사 종결자 요소(transcriptional terminator element)이다.

[0105] 일 실시형태에서, 유전적 요소는 선별가능한 마커를 암호화하는 폴리뉴클레오티드 서열에 포함된 CpG 자리의 내용이 45 또는 미만, 바람직하게는 40 또는 미만, 더욱 바람직하게는 20 또는 미만, 특히 10 또는 미만, 더욱 특히 5 또는 미만, 가장 특히 0(CpG 자리가 완전히 제거됨)인 선별가능한 마커이다.

[0106] 추가적 측면에서, 본 개시는 발현 벡터, 바람직하게는 상기에서 설명된 바와 같이 발현 카세트를 포함하는 포유동물 발현 벡터를 제공한다. 일부 실시형태에서, 발현 벡터는 적어도 두 개의 별도의 전사 유닛을 포함한다. 두 개의 별도의 전사 유닛을 갖는 발현 벡터는 또한 이중-유전자 벡터로서 언급된다. 이의 실시예로는 첫 번째 전사 유닛은 항체 또는 이의 단편의 중쇄를 암호화하고 두 번째 전사 유닛은 항체의 경쇄를 암호화하는 벡터이다. 다른 실시예로는 두 개의 전사 유닛이 효소와 같은 단백질의 두 개의 상이한 서브유닛을 암호화하는 이중-유전자 벡터이다. 그러나, 본 발명의 발현 벡터는 두 개 이상의 별도의 전사 유닛, 예를 들면, 3개, 4개 또는 그 이상의 별도의 상이한 폴리펩티드 사슬을 암호화하는 다른 뉴클레오티드 서열을 포함하는 각각의 전사유닛을 포함하는 것이 또한 가능하다. 따라서, 실시예는 각각은 4개의 상이한 서브유닛으로 구성된 효소의 하나의 서브유닛을 암호화하는 상이한 뉴클레오티드 서열을 포함하는 4 개의 별도의 전사 유닛을 갖는 벡터이다.

[0107] 일부 실시형태에서, 발현 벡터는 추가 프로모터, 인핸서, 전사 대조군 요소, 복제 기원 및 선별가능한 마커로 구성된 군으로부터 선택되는 유전적 요소를 추가적으로 포함한다.

[0108] 일부 실시형태에서, 발현 벡터는 복제 기원 및 선별가능한 마커를 암호화하는 발현 벡터의 폴리뉴클레오티드 서열에 포함된 CpG 자리의 내용이 200 또는 미만, 바람직하게는 150 또는 미만, 특히 100 또는 미만, 더욱 특히 50 또는 미만, 가장 특히 30 또는 미만인 복제 기원 및 선별가능한 마커를 추가적으로 포함한다.

- [0109] 티미딘 키나제(thymidine kinase(tk)), 디하이드로폴레이트 리덕타제(dihydrofolate reductase (DHFR)), 푸로마이신, 네오마이신 또는 글루타민 합성효소(glutamine synthetase(GS))와 같은 일반적으로 사용되는 임의 선별 가능한 마커는 본 발명의 발현 카세트 또는 발현 벡터에 사용될 수 있다. 바람직하게는, 본 발명의 발현 벡터는 본 발명의 이중 단백질의 배출(secretion)에 대한 발현 카세트의 삽입체를 위한 유용한 제한효소 자리의 제한된 수를 또한 구성한다. 단지 일시적/유전자 부체(episomal) 발현에 특히 사용되는 경우에, 본 발명의 발현 벡터는 진핵생물 숙주 세포에서 자율복제/유전자 부체 유지를 위한 엡스타인-바바이러스(Epstein Barr Virus(EBV)) 또는 SV40 바이러스의 oriP 기원과 같은 복제의 기원이 추가적으로 포함될 수 있으나 선별가능한 마커가 결핍될 수 있다. 벡터의 복제를 용이하게 하는 관련 요인이 부족한 세포에서 일시적 발현이 또한 가능하다.
- [0110] 발현 카세트를 정박하는 발현 벡터는 형광마커, ncRNA를 코딩하는 발현 카세트, 항세포사멸 단백질을 코딩하는 발현 카세트, 또는 배출 경로의 용량을 증가시키는 단백질을 코딩하는 발현 카세트를 추가적으로 포함할 수 있다.
- [0111]
- [0112] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0113] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- [0114] b) 프로모터
- [0115] c) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- [0116] d) 폴리아데닐레이션 자리
- [0117] e) 인핸서
- [0118] f) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림, 또는
- [0119] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- [0120] b) 인핸서
- [0121] c) 프로모터
- [0122] d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- [0123] e) 폴리아데닐레이션 자리
- [0124] f) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림, 또는
- [0125] a) 인핸서
- [0126] b) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림 및/또는 다운스트림
- [0127] c) 프로모터
- [0128] d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- [0129] e) 폴리아데닐레이션 자리
- [0130] f) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림,
- [0131] 만약 a) 또는 b)가 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림이면 f)는 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림이고 만약 a) 또는 b)가 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림이면 f)는 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림이라는 조건으로, 상기 인핸서의 포함은 선택적이고, 상기 폴리뉴클레오티드 서열에 의해 암호화되는 폴리펩티드는 GAPDH가 아니고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은 +1 주위의 뉴클레오티드 위치부터 +7000 주위의 뉴클레오티드 위치까지의 영역거리 내에서 시작하고, 상기 뉴클레오티드 위치는 GAPDH mRNA의 전사시작에 비례하고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림의 길이는 100 주위부터 15000 주위 뉴클레오티드까지이고, 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림은 진핵생물 GAPDH 프로모터의 5' 말단 주위부터 -3500

주위의 뉴클레오타이드 위치까지 영역거리 내에서 시작하고, 상기 뉴클레오타이드 위치는 GAPDH mRNA의 전사 시작에 비례하고, 상기 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림의 길이는 100 주위부터 15000 주위의 뉴클레오타이드까지이다.

- [0132] 일부 실시형태에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0133] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림
- [0134] b) 프로모터
- [0135] c) 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열
- [0136] d) 폴리아데닐레이션 자리
- [0137] e) 인핸서
- [0138] f) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림, 상기 인핸서의 포함은 선택적이다.
- [0139] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0140] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림
- [0141] b) 인핸서
- [0142] c) 프로모터
- [0143] d) 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열
- [0144] e) 폴리아데닐레이션 자리
- [0145] f) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림, 상기 인핸서의 포함은 선택적이다.
- [0146] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0147] a) 인핸서
- [0148] b) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림
- [0149] c) 프로모터
- [0150] d) 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열
- [0151] e) 폴리아데닐레이션 자리
- [0152] f) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림, 상기 인핸서의 포함은 선택적이다.
- [0153] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0154] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림
- [0155] b) 프로모터
- [0156] c) 폴리펩티드를 암호화하는 폴리뉴클레오타이드 서열
- [0157] d) 폴리아데닐레이션 자리
- [0158] e) 인핸서
- [0159] f) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 업스트림, 상기 인핸서의 포함은 선택적이다.

- [0160] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0161] a) 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림
- [0162] b) 인헨서
- [0163] c) 프로모터
- [0164] d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- [0165] e) 폴리아데닐레이션 자리
- [0166] f) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림, 상기 인헨서의 포함은 선택적이다.
- [0167] 추가적 측면에서, 본 개시는 하기의 순서를 포함하는 발현 벡터를 제공한다:
- [0168] a) 인헨서
- [0169] b) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 다운스트림
- [0170] c) 프로모터
- [0171] d) 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열
- [0172] e) 폴리아데닐레이션 자리
- [0173] f) 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림, 상기 인헨서의 포함은 선택적이다.
- [0174] 진핵생물 GAPDH의 비번역된 게놈 DNA 서열 업스트림, 인헨서, 프로모터, 폴리펩티드를 암호화하는 폴리뉴클레오티드 서열, 폴리아데닐레이션 자리 및 발현 벡터의 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림은, 예를 들면, 상기에서 설명된 바와 같다.
- [0175] 추가적 측면에서, 본 개시는 상기에서 설명된 바와 같이 발현 카세트 또는 발현 벡터를 포함하는 숙주 세포를 제공한다. 숙주 세포는 인간 또는 비인간 세포일 수 있다. 바람직한 숙주 세포는 포유동물 세포이다. 포유동물 숙주 세포의 바람직한 실시예는, 제한 없이, 인간 배아 신장(human embryonic kidney) 세포(Graham FL et al., J. Gen. Virol. 36: 59-74), MRC5 인간 섬유모세포(fibroblasts), 983M 인간 흑색종(melanoma) 세포, MDCK 카닌 신장(canine kidney)세포, 스프라그-다우리 래트로부터 분리된 RF 배양된 래트 폐 섬유모세포, B16BL6 마우스 흑색종(murine melanoma) 세포, P815 마우스 비만 세포, MT1 A2 마우스 유방 샘암종(murine mammary adenocarcinoma) 세포, PER:C6 세포(Leiden, Netherlands) 및 차이니스 햄스터 난소(CHO) 세포 또는 세포주(Puck et al., 1958, J. Exp. Med. 108: 945-955)를 포함한다.
- [0176] 특히 바람직한 실시형태에서 숙주 세포는 차이니스 햄스터 난소(CHO) 세포 또는 세포주이다. 적합한 CHO 세포주는, 예를 들면, CHO-S(Invitrogen, Carlsbad, CA, USA), CHO K1(ATCC CCL-61), CHO pro3-, CHO DG44, CHO P12 또는 dhfr- CHO 세포주 DUK-BII(Chasin et al., PNAS 77, 1980, 4216-4220), DUXBI 1(Simonsen et al., PNAS 80, 1983, 2495-2499), 또는 CHO-K1SV(Lonza, 베이스1, Switzerland)를 포함한다.
- [0177] 추가적 측면에서, 본 개시는 상기에서 설명된 바와 같이 발현 카세트 또는 발현 벡터를 갖는 숙주 세포를 형질 전환시키는 것 및 폴리펩티드를 회수하는 것을 포함하는, 폴리펩티드의 발현을 위한 시험관 내 방법을 제공한다. 상기 폴리펩티드는 바람직하게는 이종(heterologous), 더욱 바람직하게는 인간 폴리펩티드이다.
- [0178] 발현 카세트 또는 발현 벡터를 본 발명에 따른 숙주세포로 형질전환을 위하여, 당업계에서 잘 알려진 기술, 예를 들면, 전기천공법(electroporation), 칼슘포스페이트 공용침전법(calcium phosphate co-precipitation), DEAE-덱스트란 형질전환, 리포펙틴과 같은 임의 형질전환 기술은 만약 주어진 숙주 세포 형태가 적절하다면 사

용될 수 있다. 본 발명의 발현 카세트 또는 발현 벡터를 갖는 형질전환된 숙주 세포는 일시적 또는 안정적으로 형질전환된 세포주로서 해석되는 것을 알려져 있다. 따라서, 본 발명에 따라 본 발명의 발현 카세트 또는 발현 벡터는 유전자 부체(episomally)로 유지될 수 있고, 예를 들면, 일시적으로 형질전환될 수 있거나 숙주 세포의 게놈에서 안정적으로 통합(integration)될 수 있고, 예를 들면, 안정적으로 형질전환될 수 있다.

[0179] 일시적 형질전환은 벡터 운반 선별가능한 마커를 위한 임의 선별 압력(selection pressure)의 비장치(non-appliance)에 의해 특징될 수 있다. 형질전환 후 통상적으로 마지막 2일부터 10일 까지 지속하는 일시적 발현 실험에서, 형질전환된 발현 카세트 또는 발현 벡터는 유전자 부체 요소로서 유지되고 게놈으로 통합되지 않는다. 즉 형질전환된 DNA는 항상 숙주 세포 게놈으로 통합되지 않는다. 숙주 세포는 형질전환된 DNA를 손실하는 경향이 있고 일시적으로 형질전환된 세포 풀의 배양에 따른 집단에서 형질전환된 세포를 과성장시키는 경향이 있다. 따라서 발현은 형질전환 직후에 가장 강하고 시간이 지나면서 감소한다. 바람직하게는, 본 발명에 따른 일시적 형질전환체는 형질전환 후 2 내지 10일까지 선별 압력의 부재에서 세포 배양에서 유지되는 세포로서 이해된다.

[0180] 본 발명의 바람직한 실시형태에서 숙주 세포 예를 들면, CHO 숙주 세포는 본 발명의 발현 카세트 또는 발현 벡터를 같이 안정적으로 형질전환된다. 안정적 형질전환은 벡터 DNA와 같은 새로이 도입된 외부 DNA가 게놈 DNA로 항상 무작위, 비상동재조합 사건에 의해 혼합되는 것이다. 벡터 DNA의 복제수(copy number) 및 동시에 유전자 생성의 수는 벡터 서열이 숙주 세포의 DNA로 통합 후에 증폭되어 왔다. 따라서, 상기 안정적 통합(integration)은, 유전자 증폭을 위한 선별 압력에서 추가적으로 증가하는 유출에 따라, CHO 세포에서 이중분염색체로 발생한다. 추가적으로, 안정적 형질전환은 예를 들면, 박테리아 복제수 조절 영역이 게놈 통합에 따른 과잉(superfluous)을 표현하는 것과 같은 재조합 유전자 생성의 발현과 직접적으로 관련되지 않은 벡터 서열 부분의 손실을 초래할 수 있다. 따라서, 형질전환된 숙주 세포는 게놈으로 발현 카세트 또는 발현 벡터의 적어도 일부 또는 상이한 일부를 통합한다.

[0181] 추가적 측면에서, 본 개시는 포유동물 숙주 세포로부터 이중 폴리펩티드의 발현을 위한, 특히 포유동물 숙주 세포로부터 이중 폴리펩티드의 시험관 내 발현을 위한 상기에서 설명된 발현 카세트 또는 발현 벡터의 용도를 제공한다.

[0182] 단백질의 발현 및 회수는 당업자에 알려진 방법에 따라 수행될 수 있다.

[0183] 폴리펩티드의 발현을 위하여, 상기에서 설명된 발현 카세트 또는 발현 벡터의 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림 및 상기에서 설명된 숙주 세포사 사용되고 항상 동일한 기원이다. 놀랍게도 발현의 증가는 만약 발현 카세트 또는 발현 벡터 및 숙주 세포의 진핵생물 GAPDH 프로모터의 비번역된 게놈 DNA 서열 다운스트림 및/또는 업스트림이 상이한 기원이라면, 예를 들면, 진핵생물 GAPDH 프로모터의 인간 DNA 서열 다운스트림 및/또는 업스트림이 CHO 세포에서 사용되면 얻어진다.

[0184] 추가적 측면에서, 본 개시는 장애의 치료를 위한 의약 제조를 위한 상기에서 설명된 발현 카세트 또는 발현 벡터의 용도를 제공한다.

[0185] 추가적 측면에서, 본 개시는 장애의 치료를 위한 의약용으로서 상기에서 설명된 발현 카세트 또는 발현 벡터를 제공한다.

[0186] 추가적 측면에서, 본 개시는 유전자 치료에서 용도를 위한 상기에서 언급된 발현 카세트 또는 발현 벡터를 제공한다.

[0187] 실시예

[0188] 실시예 1: 발현 벡터의 클로닝:

[0189] I. 재료 및 방법

[0190] I.1 플라스미드 구조체

[0191] I.1.1. LB 배양 플레이트

[0192] 500 ml의 물을 16 g의 LB 아가(Invitrogen, Carlsbad, CA, USA)와 혼합하고 끓였다(1 litre의 LB는 10 g 트립톤, 5 g 이스트 추출물 및 10 g NaCl을 포함한다). 냉각 후, 각각의 항생물질을 플레이트 될 용액(100 µg/ml 엠퍼실린 플레이트 및 50 µg/ml 카나마이신 플레이트)으로 첨가하였다.

[0193] I.1.2. 중합효소연쇄반응 (PCR)

[0194] 모든 PCR은 50µl 최종 부피에서 1 µl의 dNTPs(각각의 dNTP에 대해 10 mM; Invitrogen, Carlsbad, CA, USA), 2 유닛의 Phusion®DNA 중합효소 (Finnzymes Oy, Espoo, Finland), 25 nmol의 프라이머 A(Mycrosynth, Balgach, Switzerland), 25 nmol의 프라이머 B (Mycrosynth, Balgach, Switzerland), 1.5 µl의 디메틸설폭사이드(DMSO, Finnzymes, Espoo, Finland) 및 1-3 µl의 템플레이트(1-2 µg)를 사용하여 수행하였다. 모든 프라이머는 표1에 나타났다.

[0195] 상기 PCR은 98℃에서 3분 동안 초기 변성(denaturation)에 의해 시작되었고, 98℃에서 30초 변성, 프라이머-특정 온도(CG 함량에 따라)에서 30초 풀림(annealing) 및 72℃에서 2 분 연장(elongation)의 35 사이클을 하였다. 10 분 동안 72℃에서 최종 연장을 냉각 전에 수행하였고 4℃에 보관하였다.

[0196] 표 1: PCRs에 사용된 프라이머의 요약 GAPDH: 글리세르알데하이드 3-포스페이트 디하이드로제네즈 (glyceraldehyde 3-phosphate dehydrogenase)서열, 5': 업스트림 서열, 3': 다운스트림 서열 프라이머 GlnPr1172에 있는 "T" (밑줄침)은 프라이머 다이머의 형성을 피하기 위하여 도입되었다.

표 1

[0197]

프라이머	프라이머 서열	서열 중폭	Seq ID
GlnPr 1171	ATTATTCGCGATGGCTCCTGGCATCTCTGGGACCGAGGC	5' GAPDH	서열번호: 1
GlnPr 1172	ATCGTCGCGAAGCTTGAGATTG <u>T</u> CCAAGCAGGTAGCCAG		서열번호: 2
GlnPr 1173	AGCAAGTACTTCTGAGCCTTCAGTAATGGCTGCCTG	3' GAPDH	서열번호: 3
GlnPr 1174	TGGCAGTACTAAGCTGGCACCACTACTTCAGAGAACAAG		서열번호: 4

[0198] I.1.3. 제한 분해(Restriction digest)

[0199] 모든 제한 분해를 위하여, 약 1 µg의 플라스미드 DNA(나노드롭으로 정량화, ND-1000 분광광도계(Thermo Scientific, Wilmington, DE, USA))를 각각의 10-20 유닛효소, 4 µl의 해당 10 X NE버퍼(NEB, Ipswich, MA, USA)와 혼합하였고 부피는 물로 40µl로 채웠다. 추가 지시 없이, 분해를 37℃ 1 시간 배양하였다.

[0200] 각각의 백본(backbone)의 예비 분해 후에, 1 유닛의 송아지 장 알칼린 포스파타제(CIP; NEB, Ipswich, MA, USA)를 첨가하고 37℃에서 30 분 배양하였다.

[0201] 만약 분해가 NE버퍼 3(NEB, Ipswich, MA, USA)에서 완성되면, 상기 효소는 상기 버퍼에서 강한 활성을 가지고 외부 말단의 뉴클레오타이드의 일부를 또한 분해할 수 있기 때문에 상기 버퍼를 CIP 첨가전에 NEB 버퍼 4로 변경

하였다

[0202] I.1.4. PCR 정제 및 아가로우즈 젤 전기이동(agarose gel electrophoresis)

[0203] I.1.4.1. PCR 클린업(clean up)

[0204] 분해를 하기 위해서 모든 PCR 단편은 맥커리 나겔 추출 II 키트(Macherey Nagel Extract II kit)(Macherey Nagel, Oensingen, Switzerland)를 사용하여 40 μ l의 용출버퍼를 사용하여 제조사의 매뉴얼을 따라 제한 분해 전에 클린되었다. 상기 프로토콜은 DNA 시료의 변경버퍼(changing buffer)를 위해 또한 사용되었다.

[0205] I.1.4.2. DNA 추출

[0206] 젤 전기이동을 위하여, 1% 젤을 울트라퓨어™ 아가로우즈(UltraPure™ Agarose)(Invitrogen, Carlsbad, CA, USA) 및 50X 트리시아세트산(Tris Acetic Acid)EDTA 버퍼(TAE, pH 8.3; Bio RAD, Munich, Germany)을 사용하여 제조하였다. DNA 염색을 위하여 1 μ l의 젤 레드 다이(Gel Red Dye (Biotum, Hayward, CA, USA))를 100 ml의 아가로우즈 젤에 첨가하였다. 사이즈 마커로서 2 μ g의 1 kb DNA 래더(NEB, Ipswich, MA, USA)를 사용하였다. 전기이동은 125 볼트에서 약 1시간 동안 실행하였다. 관심 밴드는 키트 추출 II(kit Extract II)(Macherey-Nagel, Oensingen, Switzerland)에 이어서 40 μ l의 용출버퍼를 사용하여 제조사의 매뉴얼을 사용하여 아가로우즈 젤로부터 잘랐다.

[0207] I.1.5. 결찰(Ligation)

[0208] 각각의 결찰을 위하여, 4 μ l의 삽입체를 10 μ l 부피에서 1 μ l의 벡터, 400 유닛의 라이게이즈(T4 DNA 라이게이즈, NEB, Ipswich, MA, USA), 1 μ l의 10X 라이게이즈 버퍼(T4 DNA ligase 버퍼; NEB, Ipswich, MA, USA)에 혼합하였다. 혼합은 RT에서 1-2 시간 배양하였다.

[0209] I.1.6. 컴피턴트 박테리아로 결찰 생성물의 변형

[0210] pGLEX41-[REP]의 클로닝 및 복제의 표준 기원을 포함하는 pCR-블런트 벡터로 만들어진 구조체를 위하여, TOP 10(One Shot®TOP 10 컴피턴트 *E. coli*; Invitrogen, Carlsbad, CA, USA)을 사용하였다.

[0211] 복제의 R6K 기원을 포함하는 플라스미드의 복제 개시를 위하여, pir 서열에 의한 코드된 π 단백질의 발현이 요구된다. π 단백질은 One Shot®PIR1 컴피턴트 *E. coli*(Invitrogen, Carlsbad, CA, USA)에 발현된다. 상기 박테리아는 R6K 서열을 포함하는 모든 벡터에 사용된다.

[0212] 결찰 생성물을 갖는 컴피턴트 생성물을 갖는 컴피턴트 박테리아를 변형하기 위하여, 25-50 μ l의 박테리아는 5분 동안 얼음 상에서 녹였다. 이후 3-5 μ l의 결찰 생성물을 컴피턴트 박테리아에 첨가하고 42°C에서 1분 동안 열 충격(thermic shock) 전에 얼음 상에서 20-30 분 동안 배양하였다. 이 후, 500 μ l의 S.O.C 배지 (Invitrogen, Carlsbad, CA, USA)를 튜브 당 첨가하였고 교반하에서 37°C에서 1 시간 동안 배양하였다. 최종적으로, 박테리아를 엠포실린이 있는 LB 플레이트(Sigma-Aldrich, St. Louis, MO, USA) 상에 넣고 37°C에서 밤사이 배양하였다. pCR-Blunt 벡터에서 클로닝을 위하여, 카나마이신이 있는 플레이트(Sigma-Aldrich, St. Louis, MO, USA)를 사용하였다.

[0213] I.1.7. 소량 (미니(mini)) 및 중급 규모 (미디(midi))의 플라스미드 제조

[0214] I.1.7.1. 미니제조(Minipreparation)

[0215] 미니 제조를 위하여, 형질전환 박테리아의 콜로니를 37°C, 200 rpm에서 LB 및 엠포실린 또는 카나마이신의 2.5 ml에서 6-16 시간 동안 자라게 하였다. DNA를 *E. coli* (QuickPure, Macherey Nagel, Oensingen, Switzerland)

를 위한 플라스미드 정제 키트에 이어서 제공된 매뉴얼로 추출하였다.

- [0216] 미니 제조로부터 플라스미드 DNA를 260 nm에서 흡광도를 측정하여 나노드롭 ND-1000 분광광도계 (Thermo Scientific, Wilmington, DE, USA)로 일단 정량화하고 1.8과 2사이이어야만 하는 OD260 nm/OD280 nm의 비율을 평가하였다. 조절 분해는 서열 확인을 위해 Fasteris SA(Geneva, Switzerland)로 시료를 보내기 전에 수행되었다.
- [0217] BAC 추출을 위해, 퀵퓨어 키트(QuickPure kit)(Macherey Nagel, Oensingen, Switzerland)를 프로토콜의 하기의 변형으로 사용하였다: LB 및 클로람페니콜(12.5 $\mu\text{g/ml}$)(Sigma-Aldrich, St. Louis, MO, USA)은 pBACe3.6 벡터를 포함하는 박테리아로 분주하였다. 37°C에서 밤사이 동안 흔들기 플랫폼 상에서 배양한 후에 배양은 500 μl 의 A1 버퍼에서 재현탁하기 전에 배양은 13300 rpm에서 5 분 동안 원심분리하였다. 500 μl 의 A2 용해 버퍼를 첨가하였고 용액을 상온에서 5 분 동안 배양하였다. 이 후, 600 μl 의 A3 버퍼로 중성화하고 13300 rpm에서 10 분 원심분리하였다. 상층액을 컬럼상에 로드하고 상기 단계로부터 퀵퓨어 미니프랩 키트의 표준 프로토콜에 따라서 사용하였다.
- [0218] **I.1.7.2. 미디제조(Midipreparation)**
- [0219] 미디제조를 위하여, 형질전환된 박테리아를 LB 및 엠포실린(또는 kanamycin)의 200 내지 400 ml에서 밤사이 동안 37°C 밤사이 동안 자라게 하였다. 이 후, 배양은 20 분 동안 725 g에서 원심분리하였고 플라스미드를 상업적 키트(NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Switzerland)에 이어서 제조사의 매뉴얼에 제공된 낮은 플라스미드 프로토콜을 사용하여 정제하였다.
- [0220] 미디제조로부터 플라스미드-DNA를 나노드롭 ND-1000 분광광도계로 3회 정량화하고 제한 분해로 확인하고 최종적으로 시퀀싱을 보냈다(Fasteris SA, 유전자va, Switzerland).
- [0221] **II. 결과 및 토론**
- [0222] **II.1. GAPDH 발현카세트(5' 및 3' GAPDH)의 DNA 영역 업스트림 및 다운스트림의 클로닝**
- [0223] BAC 클론 RPCIB753F11841Q을 이메진Imagene (Berlin, Germany)에 주문하였다. 상기 클론은 클로람페니콜(chloramphenicol) 저항 유전자를 포함하는 pBACe3.6 벡터 백본에서 인간 GAPDH 서열을 포함한다. 미디제조(midipreparation)에 의해 DNA 추출 후에, 벡터 농도를 나노드롭에 의해 27 ng/ μl 으로 결정하였다.
- [0224] 프로모터의 GAPDH 발현 카세트 업스트림 및 폴리-아데닐레이션 자리의 다운스트림을 둘러싼 직후의 DNA 서열은 템플레이트로서 27 ng의 정제된 클론 RPCIB753F11841Q 을 사용하여 증폭되었다. 프로모터의 3 kb 단편 업스트림은 서열번호. 5을 갖는 앰플리콘으로 유도하는 프라이머 GlnPr1171(서열번호: 1) 및 GlnPr1172(서열번호: 2)으로 증폭되었다. 프라이머 GlnPr1172(서열번호: 2)는 템플레이트 서열에 비례하여 베이스(base) 변경(G에서 T로)을 옮기기 때문에, 상기 PCR 반응으로부터 유래된 모든 서열은 상기 베이스 변경을 역시 옮길 것이다. 상기 변경은 GAPDH 유전자(서열번호: 17의 bp 812, 서열번호: 5의 시작에 비례하여 위치 23)의 전사시작에 비례하여 위치 -3721에 위치해 있다. 폴리아데닐레이션 자리의 3kb 단편 다운스트림은 서열번호: 6(표 1)을 갖는 앰플리콘으로 유도하는 프라이머 GlnPr1173(서열번호: 3) 및 GlnPr1174(서열번호: 4)로 증폭되었다. PCRs에 사용된 풀림 온도는 72°C이었다.
- [0225] 5' 및 3'GAPDH 단편(서열번호: 5 및 6)을 pCR-Blunt, 상업적으로 활용가능한 PCR-생성물 클로닝 벡터(pCR-Blunt, PCR Zero Blunt 클로닝 키트, Invitrogen)에서 클로닝되었다. 결합 생성물을 TOP10 컴피턴트 박테리아로 형질전환되었고 카나마이신 LB-아가 플레이트 상에서 분주하였다. 콜로니는 증폭되었고 플라스미드는 미니프랩에 의해 분리되었다. 조절 분해는 pCR-Blunt-5' GAPDH 및 pCR-Blunt-3' GAPDH 구조체를 수율하는 양성 클론을

식별하기 위하여 수행되었다.

II.2. 리포터 단백질 GFP 및 재조합 IgG1 단클론 항체(LC-IRES-HC-IRES-GFP)를 위한 DNA 단편 코딩의 제조

본 발명에서 사용된 리포터 구조체(REP)는 폴리스িস트로닉 유전자로 구성된다: IgG1 단클론 항체 경쇄 (LC)-IRES- IgG1 단클론 항체 중쇄(HC)-IRES-녹색형광 단백질(GFP). 뇌심(장)근(육)염 바이러스 (Encephalomyocarditis virus)로부터 유래된 내부 리보솜 엔트리 사이트(Internal Ribosomal Entry Sites(IRES))(Gurtu *et al.*, Biochem Biophys Res Commun.; 229(1): 295-298, 1996))는 3 펩티드 IgG1 단클론 항체 경쇄 (LC), IgG1 단클론 항체 중쇄 (HC) 및 GFP (도 1)의 번역을 허가하게 한다. 따라서, 형질전환된 세포는 IgG1 단클론 항체를 배출할 것이고 의존적 방법으로 세포 내 GFP를 축적한다. 그러나, 폴리크리토닉 mRNA는 진핵생물 세포에서 일반적이지 않고 자체의 번역은 상대적 낮은 적정량의 IgG1 및 GFP 발현을 유도하여, 매우 효율적이지 않다. REP 구조체를 포함하는 벡터는 제한효소 NheI 및 BstBI(BstBI는 65°C에서 사용됨)를 사용하여 절단된다. 발현 구조체를 포함하는 REP 단편은 자르고, 추가적으로 클로닝 단계에 사용된다.

II.3. 발현 벡터의 클로닝

벡터 pGLEX41, pcDNA3.1 (+) (Invitrogen, Carlsbad, CA)로 부터 유래된 발현 벡터는 안정 세포주 생성에 사용된다. GAPDH 서열이 있고 없이 벡터 A 및 B의 두 번째 생성을 생성시키는 변형된 초기 백본으로서 사용되었다. 모든 벡터를 위하여 동일한 프로모터-인트론 조합(mCMV 및 첫 번째 인트론 (IgDA)을 위한 기증자-수용체 단편 코딩)이 사용되었다(Gorman *et al.*, (1990) Proc Natl Acad Sci USA, 87: 5459-5463).

중간 벡터 pGLEX41-HM-MCS-amp^rA의 클로닝:

새로운 벡터 생성의 개발은 pGLEX41로부터 시작되었다. 상기 벡터를 앰피실린 저항 카세트를 방출하기 위하여 제한효소 NruI 및 BspHI를 사용하여 자른다. 백본 단편은 CIP되고 겔 전기이동에 의해 정제된다. 앰피실린 저항 유전자(블라(bla) 프로모터를 포함하여)의 코돈 최적화 버전(*E. coli*에서 발현을 위하여)을 위한 DNA 단편 코딩은 진아트(GeneArt)로부터 주문되었다. 삽입체는 제한효소 NruI 및 BspHI(백본으로서 사용된 동일한 효소)를 사용하여 진아트(GeneArt) 클로닝 벡터 #1013237 밖으로 잘렸고, 정제되었고 백본으로 클론되었다.

미니프렘은 제한 분해에 의해 분석되었다. 클론 pGLEX41-HM-MCS-amp^rA#2는 기대된 제한 프로파일을 가졌고 정확한 단편의 통합은 시퀀싱으로 확인되었다.

중간 벡터 pGLEX41-MCS-R6K-amp^rA의 클로닝

벡터 pGLEX41-HM-MCS-amp^rA#2의 복제의 pUC 기원을 교화하기 위하여 벡터는 PvuI 및 BspHI를 사용하여 절단되었다. 백본 단편은 CIP되고 정제되었다. 새로운 삽입체 단편은 발현 카세트의 일부로서 복제의 R6K 기원 및 변형된 SV40 poly(A) 서열을 포함한다. SV40 poly(A) 주위의 불필요한 박테리아 또는 바이러스 백본 서열은 삭제되어 왔다(하기의 표 2를 참조). 삽입체 단편은 진아트(GeneArt)로부터 주문되었고; 효소 PvuI 및 BspHI(백본을 위하여 사용된 바와 같이 동일함)를 사용하여 진아트 클로닝 벡터 밖으로 잘렸고, 정제되었고 백본 단편으로 클론되었다. 미니프렘을 준비하였고 서열 분석에 의해 확인되었다. 클론 pGLEX41-MCS-R6K-amp^rA#1은 정확한 서열을 가졌다.

표 2: 상이한 벡터에서 CPG 함량

표 2

	발현 벡터에서 CpG 함량		
	pGLEX41	코돈 최적화된 벡터: "A"	CpG 감소된 벡터 "B"
앰피실린(Ampicillin) 저항	49	43	19

퓨로마이신(Puromycin) 저항	93	36	1
제네티신(Geneticin) 저항	74	51	0
복제의 기원 (Origin of replication)	45	9	9
합계:	261	139	29

[0237] **중간 벡터 pGLEX41-MCS-R6K-엠포B의 클로닝**

[0238] 벡터 pGLEX41- MCS-R6K-amp^rA#1은 제한효소 BspHI를 사용하여 열렸고 엠포실린 저항을 방출하기 위하여 CIP되었다. 새로운 삽입체 단편은 *E. coli*에서 발현을 위하여 엠포실린 저항 코돈 최적화를 포함하나, 선택적 코돈 사용에 의해 삭제될 수 있는 모든 CpG 서열은 대체될 수 있다(상기 표 2를 참조). 상기 단편은 진아트에서 주문되었다. 삽입체 단편을 방출하기 위하여, 진아트 클로닝 벡터 #1016138는 BspHI를 사용하여 절단되었다. 젤 전기이동에 의해 삽입체 및 백본 단편 모두의 정제 후에, 이들은 결합되었고 PIR1 박테리아로 형질전환되었다. 미니프렘은 시퀀싱을 위해 직접 보내졌다. pGLEX41-R6K-MCS-엠포B#1은 정확한 서열을 가졌고 추가 클로닝 단계에 사용되었다.

[0239] **pGLEX41-derived 발현 벡터에서 리포터 구조체의 클로닝**

[0240] 발현 벡터 pGLEX41- MCS-R6K-amp^rA 및 pGLEX41-MCS-R6K-엠포B에서 리포터 구조체 REP를 클론하기 위하여, 벡터를 제한효소 NheI 및 ClaI을 사용하여 잘랐다. 발현 벡터 pGLEX41-HM-MCS는 제한효소 NheI 및 BstBI을 사용하여(65°C에서) 오픈하였다. 모든 벡터 백본은 분해 후에 CIP로 처리되었고 백본은 젤 전기이동에 의해 정제되었다. 백본은 리포터 구조체 REP를 위하여 코딩하는 NheI/BstBI 단편으로 결합되었다(BstBI은 ClaI으로 호환됨). 결합 생성물은 PIR1 또는 TOP 10 컴피턴트 박테리아로 형질전환되었고 엠포실린 LB-아가 플레이트 상에 플레이트되었다. 콜로니는 증폭되었고 플라스미드는 미니프렘에 의해 분리되었다. 양성 클론은 미니프렘의 제한 분해에 의해 식별될 수 있고 후속의 서열은 Fasteris SA에 의해 확인되었다.

[0241] **pGLEX41 유래된 발현 벡터에서 인접 GAPDH 서열의 첨가**

[0242] 상기 단락의 모든 제한 분해는 80 µl 최종 부피에서 수행되고 37°C에서 밤사이 배양되었다.

[0243] 5' GAPDH 서열 (서열번호: 7)은 제한효소 NruI을 사용하여 pCR-blunt-5' GAPDH로부터 절단되고 NruI을 사용하여 직선화된 발현 벡터 pGLEX41-R6K-amp^rA-[REP] 및 pGLEX41-R6K-엠포B-[REP]에서 결합되었고 재순환을 피하기 위하여 CIP 처리되었다. PIR1 콜로니의 증폭 후에(결합 생성물의 변형에 의해 얻어짐), 미니프렘을 제한 분해에 의해 분석하였다. 클론 pGLEX41-R6K-amp^rA-5'GAPDH-[REP] #2 및 pGLEX41-R6K-엠포B-5' GAPDH-[REP] #1은 제한 분석에서 예상된 사이즈의 밴드를 나타냈고 후속적으로 시퀀싱에 의해 확인되었고 추가 클로닝 단계에 사용되었다. 상기 새로운 벡터는 ScaI을 오픈되었고 CIP로 처리하였다. 3' GAPDH 단편(서열번호: 8)은 동일한 효소를 사용하여 pCR-Blunt-3'GAPDH로부터 절단되고 pGLEX41-R6K-amp^rA-GAPDH-[REP] 및 pGLEX41-R6K-엠포B-GAPDH-[REP] 발현 벡터를 생성하기 위하여 2개의 백본으로 결합되었다.

[0244] 클론 pGLEX41-R6K-amp^rA-GAPDH-[REP] #2 및 pGLEX41-R6K-엠포B-GAPDH-[REP] #8은 조절 분해는 제한 분석에서 예상되는 사이즈의 밴드를 나타냈다. 정확한 방향으로 3' GAPDH 단편의 삽입은 후속적으로 시퀀싱에 의해 확인되었다(Fasteris).

[0245] **II.4. 저항 벡터의 클로닝**

[0246] 저항 벡터의 클로닝을 위한 시작점은 벡터 pGLEX-MCS-R6K-amp^rA#1이다. 저항 유전자의 발현에 있어서 약한 프

로모터는 충분하고, mCMV 프로모터는 SV40 프로모터에 의해 대체될 수 있다. 저항 유전자를 위해 코딩하는 유전자는 진아트 SA(Regensburg, Germany)로부터 주문되었고 선별 코돈 사용에 의해(퓨로마이신: 퓨로B 및 네오마이신: neoB) 차이니스 햄스터(퓨로마이신: 퓨로A 및 네오마이신: neoA)에서 최적화되거나 CpG 함량에서 감소된다.

[0247] **pGLEX-R6K-ampA-puroA/puroB의 클로닝:**

[0248] 발현 카세트에서 퓨로마이신 저항을 클론하기 위하여, 벡터 pGLEX41-MCS-R6K-ampA#1는 제한효소 NruI 및 XbaI을 사용하여 오픈되었고 이어서 CIP 처리되었다. 삽입체 단편은 진아트로부터 주문되었고 진아트 클로닝 벡터 #1013239에서 삽입체로서 제공되었다. 이는 퓨로마이신 저항(CHO 세포의 코돈 사용을 위하여)을 위한 SV40 프로모터 및 코돈 최적화 유전자를 포함한다. 삽입체는 효소 NruI 및 XbaI(백본을 위해 사용된 바와 같이 동일함)을 사용하여 진아트 클로닝 벡터 밖으로 잘랐고, 정제되었고 백본 단편으로 클론되었다. 미니프랩을 준비하고 제한 분해로 분석하였다. 클론 pGLEX-MCS-R6K-ampA-퓨로A#1은 정확한 프로파일을 나타냈고 시퀀싱에 의해 확인할 수 있었다.

[0249] 상기 벡터는 퓨로마이신 저항 유전자를 위한 코딩 영역의 교환에 의해 벡터 pGLEX-MCS-R6K-ampA-퓨로B의 클로닝을 위해 사용되었고 반면에 SV40 프로모터를 남겼다. 새로운 삽입체 단편은 선택적 코돈 사용에 의해 삭제될 수 있는 모든 CpG 서열이 대체될 수 있는 퓨로마이신 유전자의 코돈-최적화된 버전을 포함한다. 상기 단편은 진아트에 의해 주문되고 삽입체 단편을 방출하기 위하여 클로닝 벡터 # 1016139에 운반되었고, 진아트 벡터는 제한효소 XbaI 및 NotI을 사용하여 절단되었다. 삽입체 단편은 젤 전기이동에 의해 정제되었고 XbaI 및 NotI을 사용하여 제한 분해에 의해 퓨로마이신 오픈 리딩 프레임의 방출 후, CIP 처리에 의해 pGLEX-MCS-R6K-ampA-퓨로A의 백본으로 클론되었다. 얻어진 벡터 pGLEX-MCS-R6K-ampA-퓨로B#1은 서열 분석에 의해 직접 확인되었다.

[0250] **벡터 pGLEX-R6K-ampA-NeoA 및 pGLEX-R6K-ampA-NeoB의 클로닝**

[0251] 발현 카세트에서 네오마이신 저항을 클론하기 위하여, 벡터 pGLEX-R6K-puroA#1를 제한효소 XbaI 및 NotI을 사용하여 오픈되었고, 이어서 CIP처리 되었다. 삽입체 단편은 진아트로부터 주문되었고 진아트 클로닝 벡터 #1013242(neoA) 및 #1026894(neoB)에서 삽입체로서 제공되었다. 이들은 각각 CHO 세포에서 코돈 사용에 대한 네오마이신 저항을 위한 코돈 최적화 유전자 및 CpG가 감소된 네오마이신 저항을 포함한다. 삽입체는 효소 XbaI 및 NotI(백본을 위해 사용된 바와 같이 동일함)을 사용하여 진아트 클로닝 벡터 밖으로 자르고, 정제되고, 백본 단편으로 클론되었다. 미니프랩을 제조하고 클론을 시퀀싱으로 확인하였다.

[0252] **벡터 pGLEX-R6K-ampB-NeoB 및 pGLEX41-R6K-ampB-puroB의 클로닝:**

[0253] 벡터 pGLEX41-R6K-puroB#1은 제한효소 BspHI을 사용하여 오픈되었고 후속적으로 CIP되었다. 삽입체 단편은 *E. coli*에서 발현을 위한 코돈 최적화된 엠피실린 저항 유전자를 포함하고, 반면에 선택적 코돈에 의해 삭제될 수 있는 모든 CpG 서열은 대체되어 왔다. 상기 단편은 진아트에서 주문되었고 클로닝 벡터 #1016138에서 도달되었다. 삽입체 단편을 방출하기 위하여 진아트 클로닝 벡터는 BspHI을 사용하여 절단되었다. 젤 전기이동에 의해 삽입체 및 백본 단편 모두의 정제 후에, 이들은 결합되었고 PIR1 박테리아로 형질전환되었다. 미니프랩은 시퀀싱을 위해 직접 보내졌고 확인할 수 있었다(pGLEX41-ampB-R6K-puroB#1).

[0254] 벡터 pGLEX-R6K-neoB-엠피B로 유도하는 클로닝은 백본 단편을 제조하기 위하여 제한효소 BspHI을 사용하여 pGLEX-R6K-neoB-ampA를 오픈하여 행하여졌다. 동일한 제한효소 조합을 사용하여 pGLEX-R6K-ampB-hygroB의 분해는 ampB를 위하여 삽입체 단편 코딩을 수율하였다. 상기 ampB 삽입체는 pGLEX-R6K-neoB-ampA 백본으로 클론되었다.

[0255] **II.5 저항 벡터로 인간 GAPDH 유전자의 서열 업스트림 및 다운스트림의 첨가**

[0256] 5' GAPDH 삽입체(3164 bps)를 얻기 위하여 벡터 pCR-blunt-5' GAPDH를 NruI로 절단하였다. 저항 유전자를 코딩하는 벡터는 NruI로 절단되었고, 백본 단편을 제조하기 위하여 후속적으로 CIP 처리되었다(Calf intestinal phosphatase, NEB, Ipswich, MA). 4개의 상이한 백본 단편(pGLEX-R6K-neoA-ampiA, pGLEX-R6K-neoB-ampiB, pGLEX-R6K-puroA-ampiA 및 pGLEX-puroB-ampiB)은 3164 bps 5' GAPDH 삽입체로 결합되었고 PIR1 컴피컨트 박테리아로 형질전환되었다. ApaI을 사용하여 미니프렘의 제한분해는 클론 pGLEX-R6K-neoB-ampiB-5'GAPDH #5, pGLEX-R6K-neoA-ampiA-5'GAPDH #6, pGLEX-R6K-puroA-ampiA-5'GAPDH #16 및 pGLEX-puroB-ampiB-5'GAPDH #5의 식별을 가능하게 하였다.

[0257] 상기 중간 벡터를 제한효소 ScaI로 자르고 결찰을 위한 백본을 제조하기 위하여 CIP 처리하였다. 두 번째 삽입체 단편을 운반하는 벡터, pCR-Blunt-3'GAPDH를 삽입체 단편 (3224 bps) GAPDH 다운스트림 인접 영역을 방출하기 위하여 ScaI을 사용하여 잘랐다. 4 개의 상이한 백본 분자는 정제된 3224 bps 삽입체 단편으로 결합되고 PIR1 컴피컨트 세포로 형질전환되었다. 미니프렘은 제한 분해에 의해 분석되었다. 예상되는 사이즈의 제한 단편을 나타내는 클론은 pGLEX-R6K-neoB-ampiB-GAPDH #8, pGLEX-R6K-neoA-ampiA-GAPDH #1, pGLEX-R6K-puroA-ampiA-GAPDH #1 및 pGLEX-puroB-ampiB-GAPDH #4이었다. 후속적으로 클론을 시퀀싱 분석에 의해 확인하였다 (Fasteris, geneva, Switzerland).

[0258] II.1.5. 형질전환을 위한 플라스미드 클론의 미디제조

[0259] 충분한 적정량의 플라스미드를 가지기 위하여, 미디프렘은 맥케레이나겔 키트(NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Switzerland)를 사용하여 제조되었다. 제한 분해에 의해 확인하고 시퀀싱한 후에, 플라스미드를 직선화하고 CHO-S 세포에서 형질전환을 위해 사용되었다. 표 3은 미디제조에서 얻어진 플라스미드 DNA 일괄의 농도를 요약하고 서열을 각각의 플라스미드의 식별 및 서열 정보를 확인하는 Fasteris SA로부터 파일한다. 모든 미디프렘은 형질전환에 사용되기 전에 시퀀싱에 의해 확인되었다.

[0260] **표 3: 클론된 플라스미드의 요약.** DNA 미디제조의 농도 및 직선화된 미디제조(해당 효소와 같이). GSC 숫자는 각각의 플라스미드에 대해 코드하고 관련있는 서열파일을 식별하는 것을 허가한다.

표 3

플라스미드	미디제조의 농도($\mu\text{g}/\text{ml}$)	직선화를 위한 효소	직선화된 플라스미드의 농도($\mu\text{g}/\text{ml}$)	글렌마크 (Glenmark) 플라스미드 코드
pGLEX41-R6K-ampiA-[REP]-GAPDH	1538	EcoRV	1019	GSC 2774
pGLEX41-R6K-ampiB-[REP]-GAPDH	1243	EcoRV	1233	GSC 2775
pGLEX-R6K-ampiA-neoA- GAPDH	890	AseI	766	GSC 2776
pGLEX-R6K-ampiB-neoB- GAPDH	594	AseI	979	GSC 2777
pGLEX-R6K-ampiA-puroA- GAPDH	917	AseI	859	GSC 2778
pGLEX-ampiB-퓨로B- GAPDH	869	AseI	1049	GSC 2779
pGLEX41-[REP]	2119	BspHI	868	GSC 2239
pGLEX41-R6K-ampiA-[REP]	865	BspHI	779	GSC 2240
pGLEX41-R6K-ampiB-[REP]	1751	BspHI	806	GSC 2249
pGLEX-R6K-ampiA-neoA	890	BspHI	764	GSC 2214
pGLEX-R6K-ampiB-neoB	767	BspHI	654	GSC 2244
pGLEX-R6K-ampiA-puroA	708	BspHI	659	GSC 2220

pGLEX-R6K-amp ^r B-puroB	574	BspHI	746	GSC 2213
------------------------------------	-----	-------	-----	----------

[0262] 실시예 2: 발현 벡터를 갖는 세포의 형질전환 :

[0263] 1. 재료 및 방법

[0264] CHO-S 세포 및 HEK293 세포

[0265] 포유동물 세포는 재조합 단백질의 정확한 폴딩, 집합 및 전사후 변형을 할 수 있기 때문에 단백질을 발현시키는 바람직한 숙주이다. CHO 세포주는 잘 특징되어있고 대부분의 인간 발병의 바이러스를 위한 숙주로서 제공하지 않기 때문에, 안정된 치료 단백질생성을 위한 상대적으로 안전한 숙주를 만드는데 사용된다. 차이니즈 햄스터 Ovary 세포 (CHO-S, Invitrogen, Carlsbad, CA, USA)는 4 mM L-glutamine (Applichem, Germany)으로 보충되는 PowerCHO-2 CD 배지 (Lonza, Verviers, Belgium)에서 현탁으로 배양되고 37°C, 5% CO₂ 및 80% 습도에서 진탕 배양기(2.5 cm의 원형 스트로크를 갖는200 rpm)에서 배양된다. HEK293 세포는 형질전환되기 쉽고 낮은 그림량 까지 재조합 단백질의 빠른 생성을 하게 하기 때문에 사용된다. 사용되는 세포는 HEK293-EBNA 세포 (ATCC, Manassas, VA)이고 Ex-cell 293 배지에서 현탁으로 일상적으로 배양된다(Sigma-Aldrich, St. Louis, MI).

[0266] CHO-S 및 HEK293 EBNA 세포의 계대배양은 신선한 배지에 0.5x10⁶ 살아있는 세포/ml의 분주 밀도를 사용하여 매일 3-4일 일상적으로 수행하였다. 세포를 가스교환할 수 있도록 침투할 수 있는 필터를 포함하는 50 ml 생물반응장치(bioreactor) 튜브에서 10 ml의 배지를 사용하여 배양하였다(Tubespin Bioreactor 50; TPP, Trasadingen, Switzerland). 세포 생존율 및 농도는 트립판 블루 세포 배제 방법을 사용하여 카운티스 자동화된 세포 카운터(Countess automated cell counter)로 결정되었다. 세포 농도는 CHO-S 세포에 대하여 PCV 튜브 (TPP, Trasadingen, Switzerland)를 사용하여 농축 세포 부피(PCV) 방법의 결정으로 확인되었다.

[0267] 농축 세포 부피 (PCV)

[0268] PCV 방법은 5000 rpm에서 1 분 동안 미니-PCV 튜브 (PCV 팩된 세포 부피 튜브; TPP, Trasadingen, Switzerland)에서 배양 액체의 특정 부피의 원심분리 상에서 기초되었다. 원심분리 동안에, 세포는 튜브의 베이스에서 눈금 측정기로 펠렛되었다. 농축 세포 부피의 퍼센트는 원심분리된 세포 배양액에 관계하여 펠렛의 부피를 측정하여 결정된다. 예를 들면, 1% PCV는 세포 펠렛의 10 µl는 1 ml의 배양액에 존재하는 것을 나타낸다.

[0269] 세포의 수를 세는 일상적인 세포를 위하여, 각각의 시료의 200 µl를 PCV 튜브에 파이펫하고 해당 펠렛의 부피 (µl에서)는 자(ruler)로 읽었다("easy read" 측정 기구; TPP, Trasadingen, Switzerland). 상기 부피는 1 ml에 대해 값을 갖도록 5를 곱한 후 살아있는 세포의 농도 측정을 얻도록 세포 특정 연관 요인을 사용하여 곱하였다(세포의 백만/ml에서)

[0270] "자동화" 세포 측정

[0271] 세포 농도 및 생존율은 트립판 블루의 동일한 양으로 시료를 혼합하여 카운티스® 자동화된 세포 카운터(Countess® Automated cell Counter)로 결정하였다. 용액은 기구에 의해 읽혀지기 전에 카운티스® 챔버 슬라이드로 파이펫되었다. 상기 기구는, 교정 후에, 뉴바우어 챔버의 자동화 읽기를 허가하였고 죽고 살아있는 세포의 세포 생존율 및 농도를 결정한다.

[0272]

[0273] 유동세포측정 분석

[0274] 유동세포측정은 각각의 세포의 다수 파라미터의 분석에 대한 기술이다. 상기 기술은 서로 표현형적으로 다른

세포, 예를 들면, 살아있는 세포로부터 죽은(세포의 크기 및 입상에 따른)세포의 정량적 및 질적 분석을 허가한다. 이는 또한 GFP와 같은 관심 단백질을 발현하는 세포의 정량을 허가한다. 세포는 300 μl 의 시료를 멸균 파이프팅에 의해 배양으로부터 수득되고 488 nm에서 공기-냉각된 아르곤 레이저 방출이 장착된 형광-연관된 세포 구분(FACS) 칼리브라 유동세포측정기(Becton, Dickinson 및 Company, Franklin Lakes, NJ, USA)로 분석하였다. 분석체는 셀퀘스트 소프트웨어로 만들어졌다. GFP 방출은 530/30-nm 밴드 패스 필터를 사용하여, FL-1로 검출되었다.

[0275] 첫 번째 문에서, 죽은 세포뿐만 아니라 세포 부스러기(debris)는 직선 규모 상의 SSC/FSC 점선에서 분석으로부터 배제되었다. 이후, 살아있는 세포의 GFP 형광은 대수(logarithmic) 규모상에서 히스토그램에서 디스플레이되었다. 형광 분포의 정중 값(median value)은 분석된 세포 집단의 GFP 발현 수준을 평가하는데 사용되었다.

[0276] **IgG 정량적 방법: OCTET QK**

[0277] 옥테트 QK 시스템(ForteBio, Menlo Park, CA, USA)은 항체, 단백질, 결합, DNA 및 다른 항체의 라벨-프리 정량을 수행하고 생분자 결합 상호작용의 운동특성(kinetic characterization)을 제공한다. 시료의 결합 비율(binding rate) 및 축적된 IgG1 농도 ($\mu\text{g/ml}$) 사이에 연관성은 표준곡선을 갖는 IgG 적정량의 정량화를 허가한다.

[0278] 세포 시료는 300 g에서 5 분간 원심분리한다. 이후 웰 당 항체 농도를 얻기 위하여 단백질 A 바이오센서(Protein A DIP 및 READTM Biosensor, Forte Bio, USA)를 사용하여 옥테트(Octet)로 분석되기 전에 상층액을 96 웰 플레이트에서 옥테트버퍼로 희석(IgG1 항체에 대하여 1/5)하였다.

[0279] **JetPEI를 사용하여 일시적 형질전환**

[0280] CHO-S 및 HEK293 EBNA 세포의 일시적 및 안정적 형질전환은 폴리에틸렌이민(PEI; JetPEI, Polyplus-형질전환, Illkirch, France)을 사용하여 수행되었다. PEI는 DNA와 같은 음이온 전하된 분자를 갖도록 복합체를 형성할 수 있는 양이온 중합체(cationic polymer)이다. 양이온 전하된 DNA-PEI 복합체는 음이온으로 전하된 세포표면에 결합하고 엔도시토시스(endocytosis)에 의해 내부화(internalized)된다. 이는 용해(lysis)에 의해 핵으로 방출되는 곳으로부터 라이소좀(lysosome) 구획에 도달한다. DNA-PEI 복합체를 갖는 높은 형질전환 효율은 라이소좀 분해로부터 DNA를 보호하기 위하여 PEI의 능력때문이다. 상기 세포는 제조사에 의해 제공된 매뉴얼에 따라 형질전환된다.

[0281] 모든 플라스미드는 안정 형질전환 (100 μl Tris-EDTA, pH 7.5에 재현탁된 100 μg 의 DNA)전에 직선화되었다. 일시적 형질전환을 위하여, 원형 플라스미드는 미디제조 DNA로부터 직접 사용된다. 상기 연구에서 일시적 형질전환은 50 ml 생물반응기 튜브에서 보존하고 항생제를 첨가하지 않았다.

[0282]

[0283] IgG1 및 GFP를 발현하는 안정적 CHO-S 클론은 하나의 발현 벡터 및 두 개의 저항 벡터를 공동-형질전환에 의하여 얻어질 수 있다(각각 푸로마이신 또는 네오마이신 저항에 대하여 코딩).

[0284] **안정적 풀 및 미니 풀(minipools)의 선별**

[0285] 형질전환 효율은 세포 내 GFP 발현을 분석함으로써 유동세포측정 (BD FACS Calibur cytometer, #1293)에 의해 형질전환 후 24시간에 결정되었다. 만약 GFP 양성 세포의 퍼센트가 20 %보다 높으면, 형질전환된 세포는 선택 배지로 희석되고 96 웰 플레이트로(분리된 안정적 미니 풀을 생성하기 위해 제한 희석을 위하여) 또는 T-플라스크(안정적 풀을 생성하기 위하여)에 분포되었다. 사용된 선택 배지는 제네티신 및 푸로마이신의 상이한 농도로 보충되는, PowerCHO-2, 4 mM 글루타민이었다.

- [0286] 형질전환 후 7일째에, 선별 엄중성(selection stringency)은 세포로 선별배지를 첨가하여 새롭게 하였다. 96 웰 플레이트에 있는 콜로니가 융합되자마자, 플레이트는 형광 리더를 사용하여 측정되었다.
- [0287] T-플라스크에 있는 풀은 항생제가 없는 PowerCHO-2, 4 mM L-글루타민을 사용하여 튜브스핀(tubespin) 규모로 확장되었다. 이의 생존율 및 농도는 카운티스 자동화된 세포 카운터로 평가되었다(Invitrogen, Carlsbad, CA, USA). 세포농도가 상기를 허락하자마자, 씨드 연쇄(seed train)는 50 ml 생물반응기 튜브(5% CO₂, 37°C 및 80% 습도에서 진탕기(200 rpm)에서 배양)에서 10 ml 배지의 0.5x10⁶ 세포/ml의 농도에서 세포 분주에 의해 각각의 풀에 대하여 시작되었다. 각각의 씨드 연쇄(seed train)는 성장 배지에서 0.5x10⁶ 세포/ml에서 세포를 분주하여 주 당 2회 계대하였다(세포농도는 PCV 분석에 의해 결정되었다). 씨드 연쇄는 모든 생성물의 실행(productions runs)의 접종원(inoculum)에 대하여 사용되었다(일괄).
- [0288] 다음 4-5 주 동안에 생성물의 실행은 중복(duplicates)으로 주 당 한번 분주 되었다. 풀의 안정성은 클론의 집단에 대하여 상기에서 언급된 바와 같이 FACS 및 IgG 발현에 의해 평가되었다.
- [0289] **생성물 실행(Production runs) (비연속 발효(batch fermentation))**
- [0290] 세포 풀의 일괄 실행은 접종을 위한 씨드 연쇄(seed train)를 사용하여 0.5x10⁶ 세포/ml의 농도에서 분주되었고 세포는 공급배지(Feed media)에서 7일 동안 배양되었다. 4 및 8일째에, 200 μ l의 세포를 300 g에서 5 분 동안 원심분리하였고 상층액을 옥테트(Octet)를 사용하여 측정된 IgG에 대하여 분석하였다. 추가적으로, 각각 일괄에 대한 GFP 발현을 FACS로 분석하였다.
- [0291] **2. 결과**
- [0292] **2.1 CHO 세포에서 일시적 발현:**
- [0293] 상기 연구와 비교하여 벡터는 자체의 백본에서 주요하게 다르다. 전체 발현 카세트(프로모터, 첫 번째 인트론, 발현 구조체, 폴리 (A))는 모든 벡터에 대하여 정확히 동일하다. 상기 벡터는 실시예 1에서 설명된 바와 같이 벡터 pGLEX41로부터 유래되었다. 하나의 벡터에서, 오피실린 저항 유전자는 *E. coli*에서 발현을 위해 코돈 최적화되었고 박테리아 백본은 최소로 감소되었다: pGLEX41-R6K-amp^rA-[REP](짧은 A에서). 두 번째 벡터에서, 오피실린 저항 유전자는 *E. coli*에서 발현을 위해 코돈 최적화되었으나, 모든 CpG 서열은 선택적 코돈 사용에 의해(가능할 때), 회피되었다: 상기 벡터는 pGLEX41-R6K-Amp^rB-[REP] (짧은 B에서)로 불린다. 세 번째 변형은 벡터 pGLEX41-R6K-amp^rA-[REP]-GAPDH(짧은 GAPDH_A에서) 및 pGLEX41-R6K-Amp^rB-[REP]-GAPDH(짧은 GAPDH_B에서)를 주는 벡터 A 및 B의 발현 카세트의 업스트림 및 다운스트림이 클론되는 GAPDH 인접 서열의 용도를 포함한다.
- [0294] 상이한 플라스미드 백본의 맥락에서 발현되는 리포터 단백질의 발현 수준을 비교하기 위하여 CHO-S 세포(Invitrogen)의 일시적 형질전환이 행하여졌다. 상기 형질전환(중복)을 10 ml의 최종 배지 부피를 사용하여 50 ml 생물 반응기 튜브(TPP, Trasadingen, Switzerland)에서 수행하였고 옥테트에 의해 형질전환 후 5일째에 분석하였다(도 2).
- [0295] 정확한 백본이 있는 모든 벡터(A 및 B)는 대조군 벡터 pGLEX41보다 약간 높은 발현 수준을 나타낸다. 벡터 A 및 B 사이에는 단지 최소 차이만 있다. 이는 예상되는데, 왜냐하면 백본에서 단지 차이가 일시적 발현 상에서 영향을 갖지 않아야만 하는 오피실린 저항이기 때문이다.
- [0296] 가장 현저한 관찰은 발현 상에서 GAPDH 서열의 긍정적 효과이다. 2-배 높은 발현 수준은 GAPDH 서열이 없는 하

나에 비교하여 GAPDH 인접 서열에 정착하는 플라스미드와 같이 얻어진다. 이는 A 및 B 구조체 모두에 대해 사실이다. pGLEX41 벡터와 비교하여, 3-배 높은 발현이 관찰되었다. 상기는 만약 플라스미드의 사이즈가 고려된다면 더욱 더 놀랍다. 벡터 A (7048 bps)는 벡터 GAPDH-A (13436 bps)와 비교하여 거의 절반 사이즈이다. 따라서, 일시적 형질전환 과정 동안에 운반된 DNA의 양은 모든 플라스미드에 대하여 동일하고, GAPDH-A의 몰라양의 단지 절반만이 핵으로 운반된다고 추정된다.

2.2 HEK293 세포에서 일시적 발현

상기한 플라스미드 백본의 맥락에서 발현된 리포터 단백질의 발현 수준과 비비교하기 위하여 HEK293 EBNA 세포의 일시적 형질전환을 수행하였다. 형질전환 (중복)을 10 ml의 최종 배지 부피를 사용하여 50 ml 생물반응기 튜브(TPP, Trasadingen, Switzerland)에서 수행하였고 옥테트에 의해 형질전환 후 10일째에 분석하였다(도 3).

도 3에 나타난 결과는 HEK293 EBNA 세포에서 GAPDH 인접 영역을 사용하여 얻어질 수 있는 발현에서 현저한 증가를 나타냈다. GAPDH-B 벡터는 발현에서 3배 증가를 나타내는 반면, GAPDH-A 벡터는 5 배의 발현에서 더욱 높은 증가를 나타낸다. 상기 벡터는 oriP 요소를 포함하지 않고 따라서 더욱 높은 적정량에 대한 잠재성을 가질 수 있다.

2.3 안정적 CHO 세포주에서 발현

안정적 형질전환된 세포의 구축

안정적 집단은 저항 유전자에 대하여 코딩하는 발현 벡터 및 벡터 코딩에 의하여, 이어서 항생제에 의해 매개되는 선별 압력에 의해 생성된다. 선별압력은 형질전환 후 14일째에 제거되었다. 상기 단계는 상이한 구조체 및 발현의 안정성의 리포터 단백질(IgG1 항체 및 GFP)의 발현 수준과 비교하기 위하여 생성물 실행에서 정기적 간격으로 배양되는 안정적 미니 풀 및 안정적 풀의 생성을 허가하였다

세포 풀로 수행된 생성물 실행 상에서 리포터 단백질 발현 연구

풀은 안정적 형질전환에 의해 생성되었다. 선별 절차 동안(형질전환 후 첫 번째 14일째)에 풀은 FACS에 의해 분석되었다. 배양의 생존율과 같이 GFP 양성세포 분획의 증가는 시간이 지남에 따라 관찰된다. 항생제에 의해 매개되는 선별압력은 14일 후에 풀로부터 제거되었다. 상기 접근법을 사용하여 "B" 플라스미드로 형질전환된 세포 풀은 얻을 수 없었다. 생성된 풀의 발현 수준은 세포가 50 ml 생물반응기 튜브에서 배양될 수 있자마자 분석되었다. 일괄은 중복으로 행하여졌다. 세포는 GFP 발현을 위해 FACS에 의해 분석되었고 상층액에서 IgG의 축적은 발현 8일째 후에 옥테트에 의해 분석되었다.

비례적 관계를 IgG 적정량 및 풀의 GFP 발현 사이에서 관찰할 수 있다. 따라서, 단지 IgG 데이터를 도 4에 나타냈다. GAPDH 서열을 포함하는 벡터를 갖는 형질전환된 모든 풀은 벡터 pGLEX41 또는 GAPDH 서열 없이 동일한 벡터(A 및 A-GAPDH 사이의 2.8 요소. B 풀이 생존하지 않았기 때문에 B 및 B-GAPDH 사이에서 결론을 나타낼 수 없음)와 비교하여 높은 발현을 나타낸다. A-GAPDH 및 B-GADPH로 수행된 형질전환은 pGLEX41 형질전환 (일괄-2에 대하여)보다 높은 발현의 IgG(각각 2.7 및 3.5 폴드 이상)를 유도하였다. 따라서, 풀에서, GAPDH 인접 서열은 단백질 생성에 대하여 호의적으로 보인다.

최종적으로, B-GADPH 벡터로 수행된 형질전환은 A-GAPDH로 수행된 형질전환보다 높은 발현의 IgG를 유도하였다 (1.25 요소). 따라서, 저항 유전자에서 CpG 감소는 역시 단백질의 안정적 생성을 위해 호의적으로 보인다.

of 1.25).

클론의 집단 상에서 발현 수준 연구

- [0310] 세포를 형질전환하고 클론 또는 올리고클론 집단을 얻기 위하여 선택 배지에서 96 웰 플레이트에 분포하였다. 7일 후에 선별압력을 세포로 선택 배지 첨가에 의해 새롭게하였다. GFP의 발현을 ELISA-플레이트 리더를 사용하여 형질전환 후에 14일째에 평가하였다. 상기 결과를 도 5에 나타냈다.
- [0311] 세포 내 폴에서 얻어진 결과를 확인함으로써, GAPDH 인접 서열을 포함하는 벡터로 형질전환된 세포는 GAPDH 업-및 다운스트림 서열(1.7부터 2 폴드까지 요소) 없는 동일한 백본 또는 대조군으로서 사용된 다른 벡터(pGLEX41: 2.5 폴드)보다 현저히 더 많은 GFP를 발현한다(도 5). 추가적으로, 단지 코돈 최적화된 해당 벡터 (A)보다 CpG 감소된 저항 서열을 포함하는 벡터(B)를 갖는 집단은 높은 발현을 유도하였다(A 및 B사이에서 1.5 폴드; B 및 B-GAPDH 사이에서 1.2 폴드).
- [0312] 발현연구로부터 몇 가지 결론을 그릴 수 있다, 첫째, GAPDH 업-및 다운스트림 서열은 벤치마크(pGLEX41)로서 사용된 표준 벡터보다 높은 발현을 허가하였다. 또한 세포가 GAPDH 서열이 없는 동일한 벡터 백본으로 형질전환 될 때 낮은 발현 수준이 얻어짐으로써 발현 상에서의 유익한 효과는 삽입된 GAPDH 인접 서열과 관련됨을 확인하였다. 추가적으로, 플라스미드의 발현 및 선별에서 CpG 수의 감소는 발현에 대해 역시 약간 호의적으로 보인다.
- [0313] **실시예 3: 새롭게 디자인된 벡터로 형질전환된 CHO-S GMP 세포의 일시적 발현 수준**
- [0314] GAPDH 프로모터의 5' 영역이 포볼 에스테르 응답 요소(phorbol ester response element) 뿐만 아니라 잠재적 인슐린을 정착한다는 것은 문헌에서 설명되어 있다(Alexander-Bridges *et al.*, (1992) *Advan Enzyme Regul*, 32: 149-159). 포볼 에스테르 응답 요소(-1040 -1010 bps)는 GAPDH 프로모터 (-488 - +20)로서 항상 언급되는 것의 업스트림에 위치해 있다. 안정적 H35 간암 세포주에서 수행된 삭제연구에서, 발명자는 베이스쌍(basepairs) -1200 내지 -488(전사시작 점에 비례하여)의 삭제의 현저한 효과를 나타낼 수 없었다. 따라서, 포볼 에스테르 응답 요소는 GAPDH 프로모터로부터 유도된 발현과 기능적으로 연결될 수 없다. 그럼에도 불구하고 GAPDH 인접 요소를 포함하는 플라스미드를 사용하여 관찰되는 일시적 및 안정적 발현의 증가에서 인슐린 및 PMA(phorbol-12-myristate-13-acetate, the most common phorbol ester)의 기여를 평가하기 위하여 일시적 형질전환 실험을 수행하였다.
- [0315] 인슐린이 없는 성장 배지를 얻기 위하여, PowerCHO2를 분말 배지로부터 제조하고 인슐린을 첨가하지 않았다. PMA를 시그마(St. Louis, MO)로부터 구입하였고, PowerCHO2(+/- 인슐린)에서 1.6 μ M의 최종농도(H35 간암 세포 주 상에서 알렉산더-브릿지(Alexander-Bridges)에 의하여 사용된 농도에 해당함)로 투여하였다.
- [0316] 형질전환을 상기에서 설명한 바와 같이 50 ml 생물반응기(bioreactor) 튜브(Tubespins, TPP, Trasadingen, Switzerland)에서 수행하였다. 옵티MEM(OptiMEM) (Life technologies, Carlsbad, CA)에 의해 제공된 인슐린의 존재를 피하기 위하여, 형질전환 배지를 4 mM Gln 및 25 mM 헤페스로 보충된 RPMI1640 (PAA, Pasching, Austria)로 변경하였다. 형질전환 후, 세포를 12 웰 플레이트에 분포하고 4개의 상이한 배지의 1 ml를 첨가하였다(PowerCHO2, 4mM Gln, +/-인슐린; PowerCHO2, 4mM Gln, 1.6 μ M PMA, +/- 인슐린). 다시, 2개의 IRES를 사용하여 IgG1 및 GFP를 발현하는 리포터 구조체를 사용하였다(실시예 2에서 설명됨). 상기 벡터는 형질전환 효율을 검증하는것을 허가하였다. 형질전환된 세포의 퍼센트 및 생존율을 모든 4 개의 상이한 배지 제조에서 유사함을 발견하였다.
- [0317] 도 6에 나타난 바와 같이, 인슐린 고갈 및/또는 PMA 첨가의 현저한 효과를 상기 실험 동안에 관찰할 수 없었다. 유사한 적정량을 발현에 사용된 모든 배지에서 얻었다. 상기는 GAPDH 유전자의 업스트림 인접 서열에 존재하는 잠재적 포볼 에스테르 및 인슐린 응답요소가 일시적 외래유전자(transgene) 발현에 영향을 주지 않는다는 것을 제시한다.

[0318] 실시예 4: 리포터 유전자 발현 상의 효과를 연구하기 위하여 프로모터의 업스트림 및 polyA 자리의 다운스트림 GAPDH 발현 카세트 인접하는 DNA의 단편화 분석

[0319] 인간 GAPDH 자리(locus)는 인간 게놈의 염색체 12 상에 위치해 있다. GAPDH는 효소가 글루코오스의 대사(metabolism)에서 핵심역할인 것처럼, 포유동물 기원의 모든 세포에서 계속적으로 활성화되도록 설명된다. 프로모터의 업스트림, GAPDH 유전자는 NCAPD2, 30000 bp 보다 이상으로 스트레치하는 유전자에 의해 옆에 있다. 폴리아데닐레이션 자리의 다운스트림, GAPDH 유전자는 IFF01에 의해 옆에 있다(상세설명을 위해 도 7을 참조).

[0320] GAPDH 및 프로모터 뿐만아니라, 또한 인접 영역은 상이한 종(species) 사이에서 잘 보존된다(표 4를 참조).

[0321] 표 4: 인간, 래트 및 마우스 GAPDH 인접 영역 사이에서 높은 상동(homologies)의 스트레치. 분석을 클론 매니저 9(manager 9)(ScieED, Cary, NC, USA)을 사용하여 행하였다. 숫자화는 각각 업스트림 또는 다운스트림 인접 요소(각각 서열 ID NO: 7 및 서열 ID NO: 8)의 첫 번째 베이스에 비례한다. 정렬에 사용된 서열은 서열 ID No 18의 마우스 베이스 532-3731 (업스트림) 및 8164-11364 (다운스트림) 및 서열 ID No 19의 래트 베이스 719-3918 (업스트림) 및 8495-11058 (다운스트림)에 대하여 사용된다.

표 4

[0322]

업스트림 영역				다운스트림			
상동(homology)의 서열 [래트]		상동(homology)의 서열 [마우스]		상동(homology)의 서열 [래트]		상동(homology)의 서열 [마우스]	
>80 %	>90 %	>80 %	>90 %	>80 %	>90 %	>80 %	>90 %
161-249	279-331	15-69	278-329	1608-1764	1706-1764	1614-1671	1904-2061
256-338	554-623	159-249	546-626	1894-2067	1912-2061	1888-2072	2927-3071
515-659		273-342				2918-3082	
2296-2349		515-647					
2381-2513		1143-1223					
2736-2818		1957-2009					
		2029-2080					
		2375-2485					
		2730-2821					

[0323] 설치류 및 인간 사이에서 DNA 상동(homology)의 비교는 38%의 최소 DNA 보존을 나타낸다. 프로모터 영역 또는 유전자에 대해 코딩하는 영역에 대한 DNA 외부의 보존된 스트레치의 존재는 DNA 서열을 유지하거나 단지 특정/소수 변경을 허가하는 세포 상에서의 선별 압력일 수 있음을 나타낸다. 본 발명의 특정 경우에, GAPDH 인접 영역은 GAPDH 유전자의 높은 발현 수준을 유지하기 때문에 세포에 대해 중요할 수 있다. 발현 감소를 유도하는 DNA 서열에서의 변경은 이에 대하여 선별될 수 있다.

[0324] 발현에서 증가가 관찰된 업스트림 및 다운스트림 GAPDH 요소의 기여를 평가하기 위하여, 구조체는 단지 업스트림 GAPDH 인접 영역(서열번호: 7), 업스트림 GAPDH 인접 영역 또는 다운스트림 GAPDH 인접 영역(서열번호: 8)의 단편을 포함하도록 만들어졌다. 리포터 IgG1 타입 항체는 RES 구조체 (경쇄-IRES-중쇄)에 의해 발현되고, 따라서 다수의 플라스미드의 공동-형질전환을 피한다. GAPDH 업스트림 단편의 단편화에 대한 상세설명을 도 8에 나타냈다. 하기의 업스트림 GAPDH 인접 영역의 단편이 사용되었다: 단편 1 (서열번호: 9), 단편 2 (서열번호: 10), 단편 3 (서열번호: 11), 단편 4 (서열번호: 12), 단편 8 (서열번호: 13), 단편 9 (서열번호: 14), 단편 11 (서열번호: 15), 단편 17 (서열번호: 16).

[0325] 사용된 업스트림 GAPDH 인접 영역(서열번호: 7)은 3개가 자체의 5'에 게놈 DNA에 연결되고 3 개가 자체의 3'말단에 게놈 DNA에 연결된 NruI 제한 자리의 2 곱하기 3(총 6)의 뉴클레오티드를 포함한다. 사용된 다운스트림

GAPDH 인접 영역(서열번호: 8)은 3개가 자체의 5'에 게놈 DNA에 연결되고 3 개가 자체의 3'말단에 게놈 DNA에 연결된 ScaI 제한 자리의 2 곱하기 3(총 6)의 뉴클레오티드를 포함한다. 각각의 제한 자리의 뉴클레오티드가 없는 업스트림 GAPDH 인접 영역 및 다운스트림 GAPDH 인접 영역은 서열번호: 20(제한효소 자리 없는 업스트림 GAPDH 인접 영역) 및 서열번호: 21(제한효소 자리 없는 다운스트림 GAPDH 인접 영역)에 나타난다. 사용된 업스트림 GAPDH 인접 영역의 단편은 게놈 DNA(단편 1은 자체의 5' 말단에 NruI 제한 자리(restriction site)의 3 뉴클레오티드를 포함하고; 단편 2는 자체의 3' 말단에 NruI 제한 자리의 3 뉴클레오티드를 포함하고; 단편 3은 자체의 5' 말단에 NruI 제한 자리의 3 뉴클레오티드를 포함하고; 단편 4는 자체의 3' 말단에 NruI 제한 자리의 3 뉴클레오티드를 포함하고; 단편 8은 자체의 3'말단에 NruI 제한 자리의 3 뉴클레오티드를 포함하고; 단편 9는 자체의 5'말단에 NruI 제한 자리의 3 뉴클레오티드 및 자체의 3' 말단에 NruI 제한 자리의 3 뉴클레오티드를 포함하고; 단편 11은 자체의 3'말단에 NruI 제한 자리의 3 뉴클레오티드를 포함한다)에 연결된 자체의 5' 및/또는 자체의 3' 말단에 있는 각각의 제한 자리의 3 뉴클레오티드를 각각 포함한다. 단편 17은 제한 자리의 뉴클레오티드를 포함하지 않는다. 각각의 제한 자리의 뉴클레오티드가 없는 업스트림 GAPDH 인접 영역의 단편은 서열번호: 22 (제한 자리 없는 단편 1), 서열번호: 23 (제한 자리 없는 단편 2) 서열번호: 24 (제한 자리 없는 단편 3), 서열번호: 25 (제한 자리 없는 단편 4), 서열번호: 26 (제한 자리 없는 단편 8), 서열번호: 27 (제한 자리 없는 단편 9), 서열번호: 28 (제한 자리 없는 단편 11)에 나타났다.

[0326] 발현 상의 업스트림 및 다운스트림 GAPDH 요소의 효과는 상층액에서 배출된 IgG1의 양을 정량화하기 위하여 옥테트(Fortebio, Menlo, CA, USA)를 사용하여 형질전환 후 10일 쯤에 평가되었다(도 9 참조). pGLEX41, 본래 벡터는 pGLEX41-amp^r 벡본에 사용된 개선된 새로운 벡터 디자인과 비교하여 낮은 발현 결과(80%)를 준다. 본래 pGLEX41 벡본과 비교하여 새로운 디자인은 복제의 상이한 기원(복제의 pUC 기원 대신에 R6K), *E. coli*에서 옴피실린 저항에 필요한 amp^r 유전자의 코돈 최적화 및 박테리아 기원의 불필요한 링커(또는 스페이서)서열의 제거를 포함한다. 벡터 모두는 대략 동일한 사이즈를 가진다.

[0327] 놀랍게도, 업스트림(서열번호: 7) 및 다운스트림 요소(서열번호: 8)을 포함하는 pGLEX41-amp^r (발현 결과를 나타내는 도 9에서 pGLEX41-업/다운으로 명명)는, 업스트림 및 다운스트림 서열없는 동일한 벡터와 비교하여 높은 발현(1.5 지수)을 준다. 만약 누군가 사이즈에서 차이(업/다운 단편은 대략 6000 bps에 의해 플라스미드의 사이즈를 증가시킴) 및 따라서 형질전환 동안에 운반된 플라스미드 복제에서 차이를 고려하면, 효과는 플라스미드 기본 당(per plasmid basis)에서 더욱 더 중요할 수 있다

[0328] 단지 업스트림 단편(업)을 포함하는 벡터는 본래 발현 구조체 pGLEX41-amp^r에 유사한 발현 수준을 나타낸다. 단지 다운스트림 단편(다운)을 포함하는 벡터는 본래 발현 구조체 pGLEX41-amp^r과 비교하여 발현에서 현저한 증가(1.2 지수)을 나타내고 있다. 발현에서 추가 증가는 만약 업- 및 다운스트림 단편 모두가 존재한다면, 관찰될 수 있다. 상기는 업스트림 단편의 단편화에 의해 확인된다. 단편 9 및 프로모터 근부(proximal) 단편 8은 pGLEX41-amp^r와 비교하여 발현에서 임의의 차이를 나타내지 않는다. 단편 1, 11 및 17은 발현에서 증가를 나타낸다. 가장 높은 증가는 단편 4에서 관찰되었다. 프로모터 근부(proximal) 단편 8이 임의의 효과를 나타내지 않음이 강조되어야만 한다. 따라서 발현에서 증가는 기존에 기재된 서열에 의해 설명될 수 없다(Alexander-Bridges *et al.*, (1992) *Advan 효소 Regul*, 32: 149-159), Graven *et al.*, (1999) *Biochimica et Biophysics Acta* 147: 203-218).

[0329] 흥미롭게도, 단편 2 및 3은 발현에서 현저한 감소를 유도한다. 상기는 예상되지 않았고, 특히 상기 단편이 반대 방향(도 9에서 안티센스(AS))으로 클론되는 사실의 관점에서 상기 효과를 초래하지 않는다. 단편 1, 8, 9, 11 및 17에 대하여 센스 또는 안티센스 방향(orientation)(데이터를 나타내지 않음)으로 통합된 단편에 대하여 발현에서 차이가 관찰되지 않았다. 비록 단편 2의 일부이지만, 단편 11은 상기 효과를 나타내지 않는다. 따라서 발현에 해를 끼치는 것처럼 보이는 서열 요소는 단편 11을 얻기 위하여 단편 2에서 삭제된 BstBI-BstBI 단편 상에 적어도 부분적으로 있어야만 한다.

[0330] 추가적으로, 부정적 요소가 BstBI-BstBI 단편 상에 위치한다는 가설은 단편 3 (BstBI-BstBI 단편을 포함하는)

및 단편 1 사이에 관찰된 발현에서 증가에 의해 지지된다.

- [0331] 부정적 효과(BstBI-BstBI)를 가지는 단편을 국한시키기에 쉬운 것으로 보이는 반면에, 상기 연구로부터 단편 2 및 3에서 관찰되는 부정적 효과가 어떻게 완전한 업스트림 단편에서 서열 요소 존재에 의해 보상되는 지는 덜 분명하다. 상기 부정적 효과는 단편 1 및 단편 4(그러나 단편 1에 대한 발현에서 증가는 단편 4보다 약함)에 의해 관찰된 적은 긍정적 효과에 의해 균형될 수 있다. 그럼에도 불구하고 상기 관찰된 단편 4에 대한 긍정적 효과(1.25 지수)가 부정적 효과 (0.4 지수)와 비교하여 덜 중요한 것처럼 보인다. 추가적으로 BstBI-BstBI 단편이 없는 전체 업스트림 영역인 단편 9은 전체 GAPDH 업스트림 인접 영역(nevertheless, 단편 9 포함한다s the EcoRV-BstBI 단편 which is part of 단편 2 및 3 및 might have a 부정적 효과 on 발현)과 비교하여 증가된 발현을 나타내지 않는다(그럼에도 불구하고, 단편 9은 단편 2 및 3의 일부인 EcoRV-BstBI 단편을 포함하고 발현 상에서 정적 효과를 가질 수 있다)
- [0332] 관찰된 효과 뒤의 기전에 대하여 단지 추측될 수 있다. 단편 2 및 3으로 관찰된 발현 상의 부정적 효과의 방향 의존성은 단지 하나의 방향의 발현을 유발할 수 있는 둘러싸는 프로모터가 없기 때문에, 비식별된 오픈리딩프레임의 발현(예를 들면, ncRNA의 발현)을 배제한다. 발현이 기저 수준 밑으로 감소된 사실은 긍정적 효과(예를 들면, 인핸서 활성)의 부재 뿐만아니라 부정적 효과에 의존하는 방향의 존재도 나타낸다.
- [0333] 요약하면, CHO 세포에서 일시적 발현의 놀라운 증가는 만약 인접 영역, 업스트림 및 다운스트림 영역 모두가 발현 플라스미드에서 존재하면, 관찰된다. 비록 단편 4는 발현상에서 현저한 긍정적 효과를 가지지만, 단일 단편은 관찰되는 발현의 전체 증가에 책임되는 것으로 확인될 수 없다. 발현 벡터 GLEX41-amp^rA (up/down) 발현 증가는 up- 및 다운스트림 인접 영역 모두의 효과를 요약하는 것으로 보인다.
- [0334] 실시예 5: 차이니즈 햄스터 GAPDH 유전자 및 차이니즈 햄스터 프로모터의 비번역된 게놈 DNA 서열 업스트림의 클로닝
- [0335] 1.1 발현 벡터로 차이니즈 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림의 클로닝
- [0336] 차이니즈 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림은 PCR에 의해 CHO-S (Life Technologies) 세포의 게놈 DNA로부터 증폭되었다. 게놈 DNA는 실시예 1에서 설명된 바와 같이 추출되었다. 구조체는 실시예 1에서 설명된 바와 같이 리포터 유전자 구조체 [REP]의 발현을 위하여 마우스 CMV 프로모터 또는 차이니즈 햄스터 GAPDH 프로모터를 사용하여 제조된다.
- [0337] 마우스 CMV 프로모터와 혼합하여 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림의 클로닝을 위하여, 프라이머 GlnPr1896 및 GlnPr1897는 실시예 1에서 설명된 PCR 프로토콜을 사용하고 서열번호 30으로 앰플리콘을 유도하여 3 kbs 단편 (bps 672 to 3671 of 서열번호 29)의 증폭에 사용되었다. 앰플리콘은 프라이머에 의해 도입된 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림 및 5' 및 3' 제한효소 자리를 포함한다.
- [0338] 차이니즈 햄스터 GAPDH 프로모터와 혼합하여 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림의 클로닝을 위하여, 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림 및 서열번호 31로 앰플리콘을 유도하는 GAPDH 프로모터 (bps 672 to 4179 of 서열번호 29)를 포함하는 게놈 DNA 서열을 포함하는 3508 bps 단편을 증폭하기 위하여 프라이머 GlnPr1902 및 GlnPr1905가 사용되었다. 두 번째 PCR에서, GlnPr1901 및 GlnPr1902은 서열번호 32로 유도하는, 단지 프로모터 영역(bps 3672 to 4179 of 서열번호 29)을 포함하는 508 bps 단편의 증폭에 사용되었다. 벡터 "A"에서 사용된 인트론(실시예 1에서 설명됨)은 프라이머 GlnPr1903 및 GlnPr1904를 사용하여 증폭되었다.
- [0339] 첫번째 융합 PCR은 템플레이트로서 서열번호: 32를 갖는 앰플리콘 및 인트론 서열을 갖는 앰플리콘을 사용하여

프라이머 GlnPr1904 및 GlnPr1901로 수행되었다. 상기 앰플리콘은 차이니즈 햄스터 GAPDH 프로모터, 인트론 및 프라이머에 의해 도입된 5' 및 3' 제한효소 자리를 포함한다. 모든 프라이머는 표5에 나타났다.

[0340] 두 번째 융합 PCR은 템플레이트로서 서열번호 31을 갖는 앰플리콘 및 인트론 서열을 갖는 앰플리콘을 사용하여 프라이머 GlnPr1905 및 GlnPr1904으로 수행되었다. 앰플리콘은 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림, 차이니즈 햄스터 GAPDH 프로모터, 인트론 및 프라이머에 의해 도입된 5' 및 3' 제한효소 자리를 포함한다.

[0341] 1% 아가로우즈 젤에서 정제 후에, "뉴클레오스핀 젤 및 PCR 클린업(NucleoSpin Gel and PCR Clean-up)" (Macherey Nagel, Oensingen, Switzerland)를 사용하여 관심의 밴드를 자르고 정제하였다. 정제된 단편은 제로블런트 PCR 클로닝 키트(Zero Blunt PCR cloning kit) (Invitrogen, Carlsbad, CA, USA)를 사용하여 플라스미드 pCR_Blunt로 클론되었다. 결합 생성물(Ligation products)은 컴피턴트(competent) *E.Coli* TOP10(One Shot®TOP 10 Competent *E.Coli*; Invitrogen, Carlsbad, CA, USA)으로 형질전환되었고 미니프랩의 제한 분석으로 분석되었다. 상기는 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림을 포함하는 플라스미드 pCR_blunt[CHO-업스트림GAPDH], 벡터"A"로부터 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림 및 GAPDH 프로모터 및 인트론을 포함하는 pCR_Blunt[CHO-upstreamGAPDH_GAPDH프로모터] 및 벡터"B"로부터 GAPDH 프로모터 및 인트론을 포함하는 pCR_Blunt[CHO-GAPDH프로모터]로 유도한다.

[0342] 배출된 유전자의 발현에 대한 자체 효과상에서 앰플리콘의 평가를 위하여, 벡터"A"(실시에 1에서 설명됨)가 사용되었다. 이전에 설명한 바와 같이, 상기 벡터에 사용된 발현 카세트는 배출된 IgG1 및 GFP를 위한 폴리시스트로닉(polycistronic) 유전자 코딩을 포함한다(실시에 1 참조). 따라서 형질전환된 세포는 IgG1 단클론 항체를 배출할 것이고 의존적 방식으로 세포 내 GFP를 축적할 것이다.

[0343] 차이니즈 햄스터 GAPDH 유전자의 게놈 DNA 서열 업스트림을 포함하는 3 kb 삽입체 단편을 배출하기 위하여, 플라스미드 pCR_Blunt[CHO-upstreamGAPDH]를 제한효소 NaeI를 사용하여 절단하였다. 상기 삽입체는 "A"의 백본 상에서 클론되고 제한효소 NruI를 사용하여 절단되고 CIP처리되었다(CIP; NEB, Ipswich, MA, USA). 백본 및 삽입체는 T4 DNA 라이게이즈를 사용하여 결합되었고(T4 DNA ligase, NEB, Ipswich, MA, USA) 후속적으로 컴피턴트 *E.coli* PIR1으로 형질전환되었다. 클론은 미니프랩 제조를 위하여 선택되고 후속적으로 제한분석을 하였다. 얻어진 플라스미드는 "A_GAPDH_UP"으로 명명하였고, 시퀀싱 분석으로 확인하였고 뉴클레오본드 엑스트라 미디키트(NucleoBond Xtra Midi kit)(Macherey Nagel, Oensingen, Switzerland)를 사용하여 미디프랩 규모로 생성되었다.

[0344] 차이니즈 햄스터 GAPDH 프로모터를 사용하여 발현 구조체의 클로닝을 위하여, 삽입체 단편은 제한효소 NheI 및 NruI를 사용하여 분해에 의해 플라스미드 pCR_Blunt[CHO-업스트림GAPDH_GAPDH 프로모터] 및 pCR_Blunt[CHO-GAPDH프로모터]로부터 방출되었다. 얻어진 단편은 벡터 "A"의 백본에서 클론되었고 동일한 효소를 사용하여 오픈되었고 CIP처리되었다. T4 DNA 라이게이즈로 결합 및 컴피턴트 *E.coli* PIR1로 전환 후에, 클론은 미니 제한 분석을 위해 선택되었다. 얻어진 플라스미드는 "A_GAPDH_UP_Prom" (차이니즈 햄스터 GAPDH 및 프로모터의 비번역된 게놈 DNA 서열 업스트림을 갖는 플라스미드)으로 명명되었고 "A_PR"(단기 프로모터를 갖는 플라스미드)는 시퀀싱 분석에 의해 확인되었고 키트 뉴클레오본드 엑스트라 미디키트(NucleoBond Xtra Midi)(Macherey Nagel, Oensingen, Switzerland)를 사용하여 미디프랩 규모로 생성되었다.

[0345] 2. 리포터 유전자 구조체의 발현 상에서 차이니즈 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림의 평가

[0346] CHO-S 세포는 10 ml의 배지 부피(실시에 2에서 설명됨)를 사용하여 튜브스핀생물반응기(tubespins bioreactors)에서 형질전환되었다. 형질전환된 세포는 37°C, 5 % CO2 및 80 % 습도에서 200 rpm 진탕으로 진

탕 배양기에서 배양되었다. 세포의 상층액은 단백질 A 바이오센서 (ForteBio, Menlo Park, CA, USA)가 있는 옥테트 QK 시스템을 사용하여 IgG1 발현에 대하여 분석되었다. 결과는 도 10에 나타냈다.

[0347]

마우스 CMV 프로모터 (A)와 비교하여 GAPDH 프로모터("A_PR")를 포함하는 플라스미드의 발현 수준은 50%에 의해 감소되었고, 차이니스 햄스터 GAPDH 프로모터는 바이러스 프로모터만큼 강하지 않다. 차이니스 햄스터 GAPDH 프로모터 ("A_GAPDH_UP_Prom")와 혼합하여 차이니스 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림을 포함하는 플라스미드는 단지 GAPDH프로모터("A_PR")을 갖는 구조체와 비교하여 발현에서 2 폴드 증가를 나타낸다. 차이니스 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림 및 마우스 CMV 프로모터 ("A_GAPDH_UP")을 포함하는 플라스미드는 단지 마우스 CMV 프로모터("A")를 포함하는 플라스미드에서 40% 이상의 가장 높은 발현 및 증가를 나타낸다("A"). 상기는 차이니스 햄스터 GAPDH 유전자의 비번역된 게놈 DNA 서열 업스트림은 리포터 단백질의 발현 상에서 인핸서 효과를 가진다.

[0348]

표 5: 실시예 5에서 클로닝에 사용된 프라이머

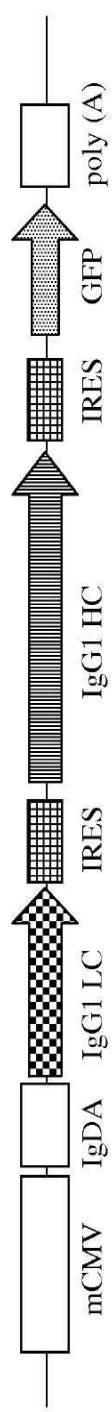
표 5

[0349]

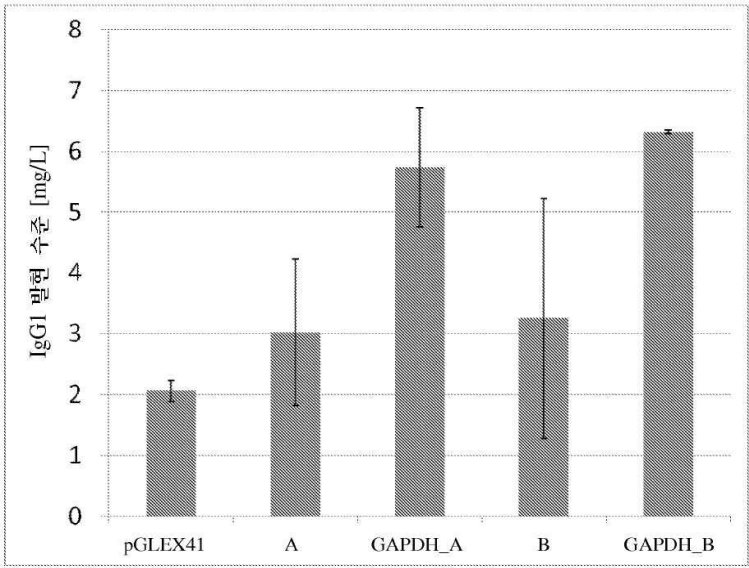
프라이머	서열식별번호 (SEQ ID No)	서열	방향 (Orientation)	제한자리 (Restr. site)
GlnPr 1896	서열식별번호 33	TACGGCCGGCTTCACTGTACAGTGGCACAT	정방향	NaeI
GlnPr 1897	서열식별번호 34	TCAGGCCGGCCGTGGTTCTTCGGTAGTGAC	역방향	NaeI
GlnPr 1901	서열식별번호 35	TACTCGCGAAGAAGATCCTCAACTTTCCACAGCC	정방향	NruI
GlnPr 1902	서열식별번호 36	GTTCATAAACGAGCTCTGCTATTTATAGGAAGTGGGTG	역방향	/
GlnPr 1903	서열식별번호 37	CACCCCAGTTCCTATAAATAGCAGAGCTCGTTTAGTGAAC	정방향	/
GlnPr 1904	서열식별번호 38	CGCTAGCACCGGTCGATCGA	역방향	NheI
GlnPr 1905	서열식별번호 39	TACTCGCGATTCACTGTACAGTGGCACATAC	정방향	NruI

도면

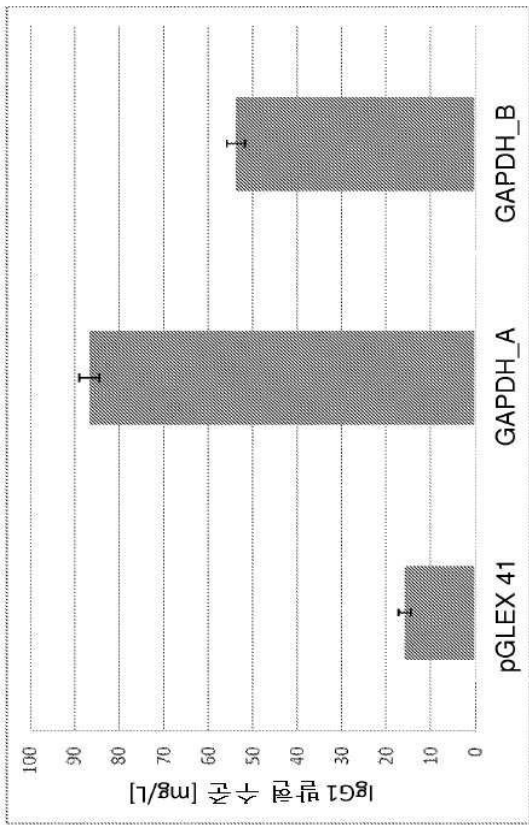
도면1



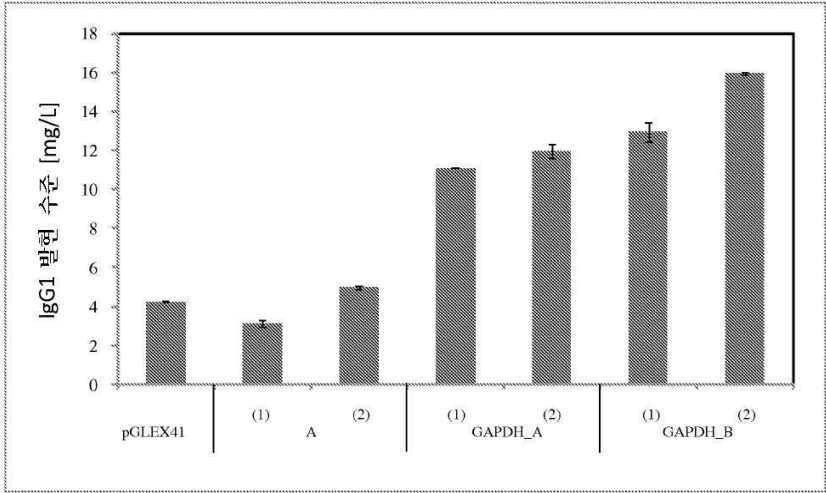
도면2



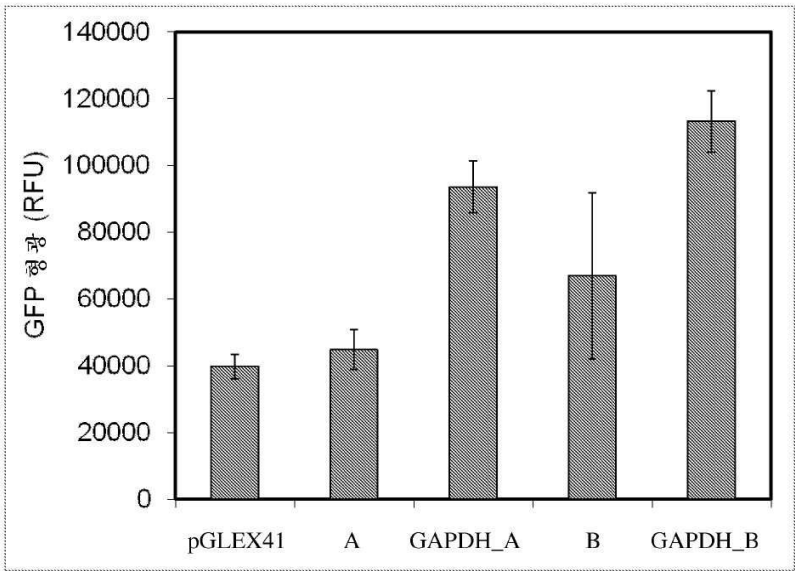
도면3



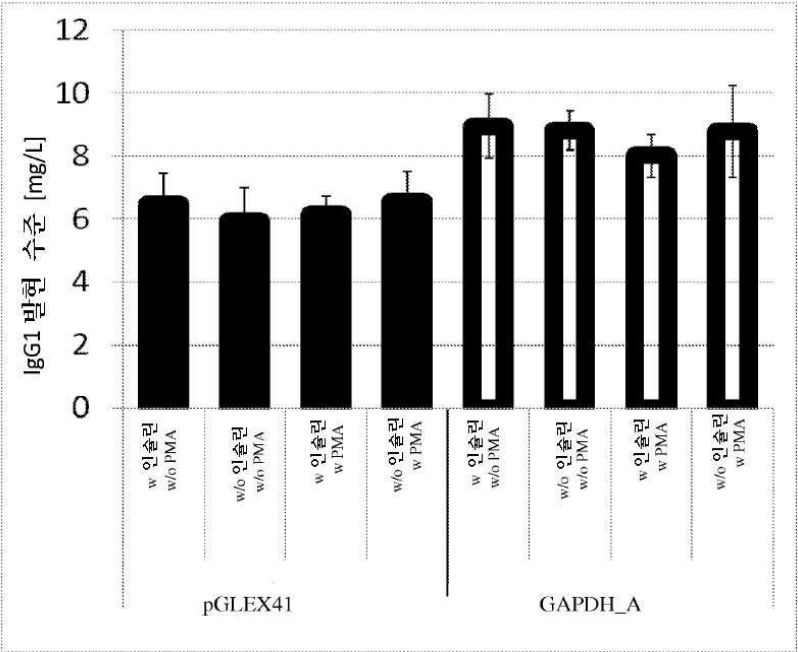
도면4



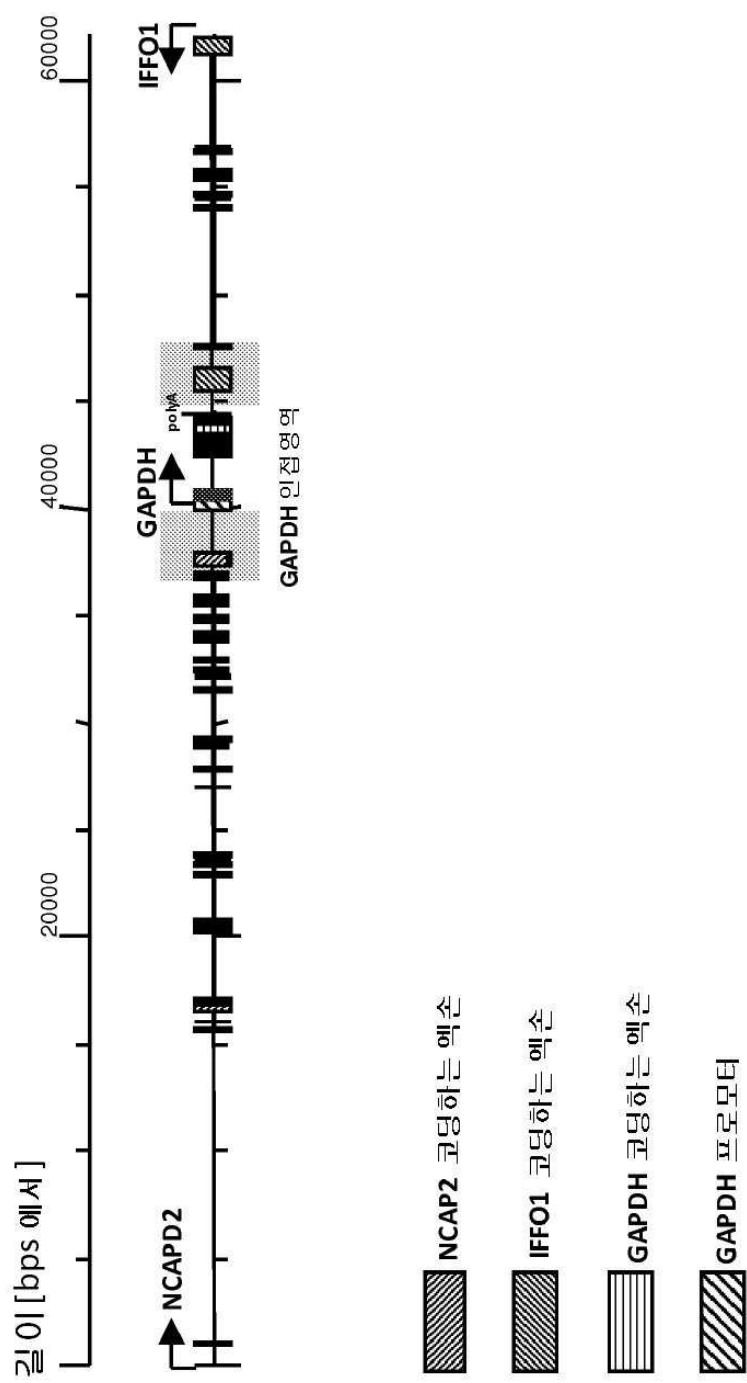
도면5



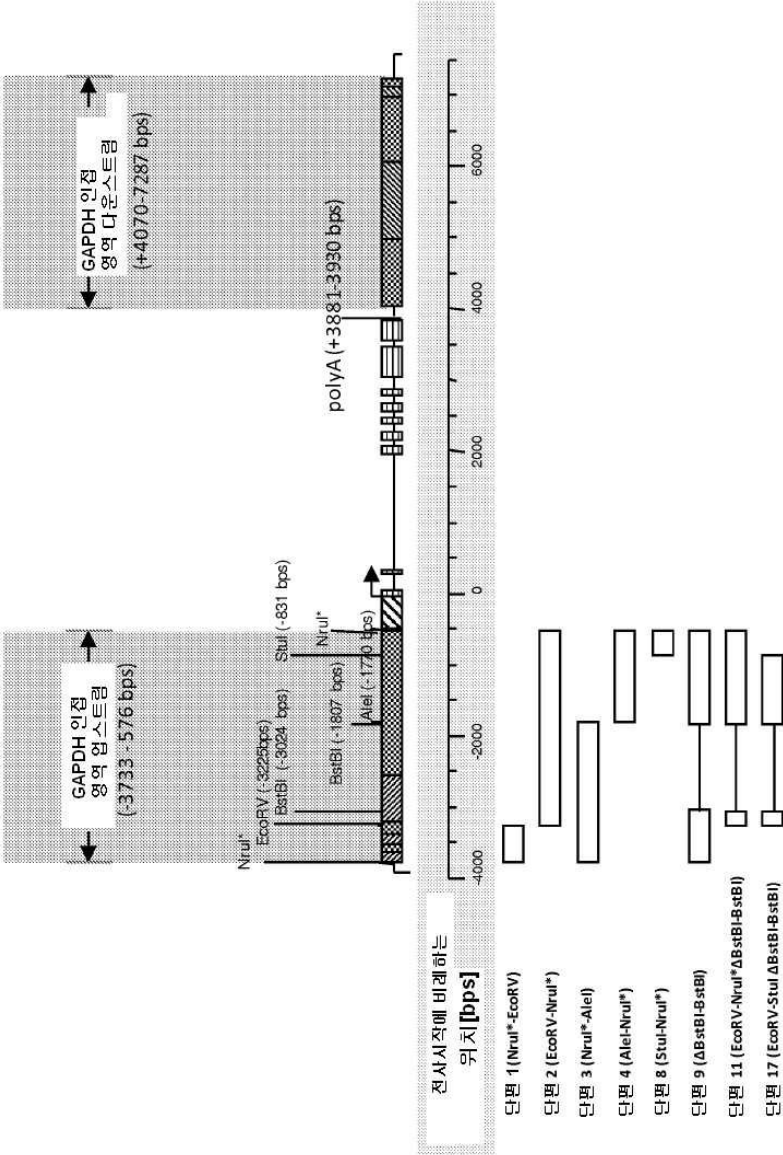
도면6



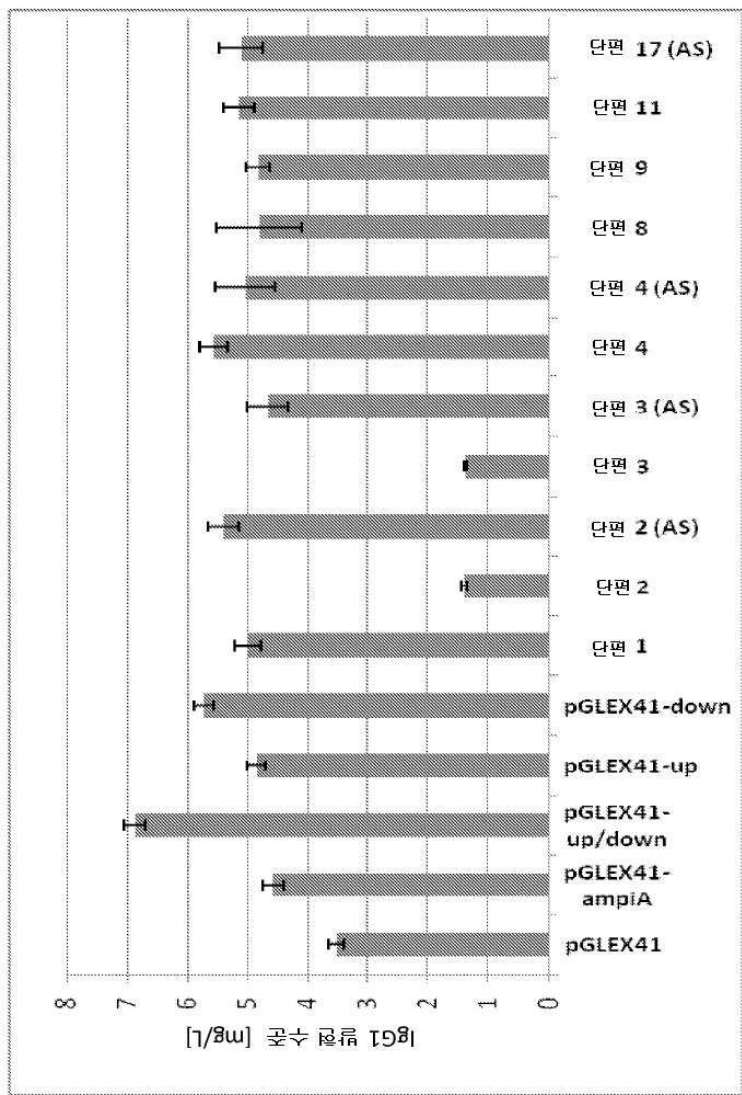
도면7



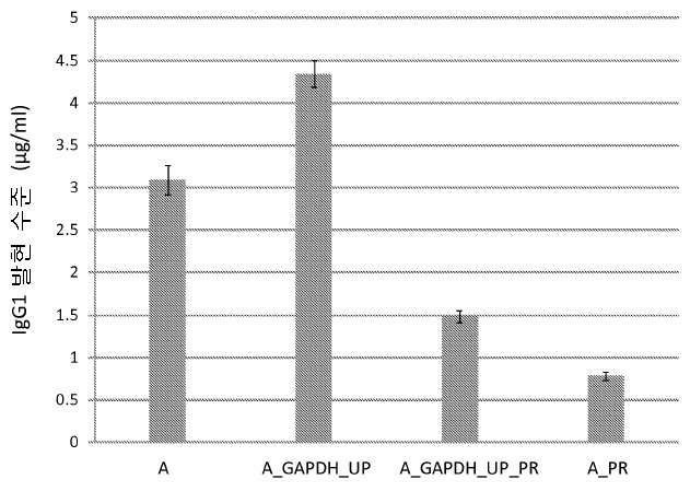
도면8



도면9



도면10



서열 목록

<110> Glenmark Pharmaceuticals SA

<120> Expression Cassette
 <130> 2014fpi-06-003
 <150> PCT/US US 61/567,675
 <151> 2011-12-07
 <160> 39
 <170> KopatentIn 2.0
 <210> 1
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer 1
 <400> 1
 attattcgcg atggctcctg gcatctctgg gaccgaggc 39
 <210> 2
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer 2
 <400> 2
 atcgctcgca agcttgagat tgtccaagca ggtagccag 39
 <210> 3
 <211> 36
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer 3
 <400> 3
 agcaagtact tctgagcctt cagtaatggc tgcctg 36
 <210> 4
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer 4
 <400> 4
 tggcagtact aagctggcac cactacttca gagaacaag 39
 <210> 5

<211> 3179

<212> DNA

<213> Artificial Sequence

<220><223> 5' GAPDH fragment

<400> 5

```

atcgtcgca agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa      60

ccatccactg gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct      120
gcagtgattt gtttcttctt cttttttaaa tcacgttttc ctgccttttc taggttctag      180
gtaccagcct ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg      240
tactaccctg cggcatccaa acaccagca gcgagcttcc aaaaagaaac ccaaagtgtg      300
cttctcaagt gatgagtcca gtgaggaagg tatgatgtc cgcctgttc cggcccgaga      360
aggcacacag ctagggtgca gagggctggt ttccatagga cctgtctcgg gggcctgagt      420
gtagatgtc tgccccactg ccgcagaagg gcctctcctg tacagcttgg attttatttc      480

ttctgtcggg tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg      540
tgactcagct ttttcttatt cctttaattc tttgcataga tcttcagca gagatgacag      600
aagacgagac acccaagaaa acaactccca ttctcagagc atcggtctgc aggcacagat      660
cctaggaagt ctgttctgt cctccctgtg cagggtatcc tgtagggtga cctggaattc      720
gaattctgtt tcccttgtaa aatatttgtc tgtctctttt ttttaaaaaa aaaaaaggcc      780
gggcactgtg gctcacgcct gtaatcccag cactttgcga taccaaggcg ggtggataac      840
ctgaggtagg gagttcgaga ccagcctgac caacatggag aaaccccatc tctactaaaa      900

ataaaaaatt agccgggctg attggcgtgc gcctgtaatc ccagctactc aagagctga      960
ggcaggagaa tcgcctgaac ccagaggcgg aggtttagt gagccgaat cacaccattg      1020
cactccagct tgggcaacaa tagcgaacct ccatctcaaa ttaaaaaaaaa aatgcctaca      1080
cgctctttaa aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct      1140
tcaactttgg aatatatgtt tgccaatctc ctgtttttct aatgaataaa tgtttttata      1200
tacttttaga cattttttcc taagcttgtc ttgttttcat ctttcacatt agcccagttt      1260
catgcagcag agagagggtt atcagtgcag agagagatga gtgagcccag agtcctaggg      1320

cctgtcccgg gatggcagat gagcttccctg ccccgctact gccaccttc ccctctcaac      1380
ctctggacce tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc      1440
tccagatcag tgtctctgaa cggggggcaa ttttgtctgc cagaggacat ctgacaacac      1500
ctggggcctg ttttgtgtc atagcctata ggggaagaat gctaccagca tttgtgggaa      1560

```

gaggccaggg atgtggctca acatcctgca gtgcacagga tggccctca acaaagaatc 1620
acacggccca caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg 1680
ctttctgccg tgtgcctcac cccaccccca ctcgtgttcc ctaatcccat ctccaaaggt 1740

tggcagcaga ccggcccagg ctcgtggaag ttcagatcat gatccctcc agctctgcag 1800
gagacaagac ctgtctccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct 1860
atgtcgcagg cgggctgccc cctgaagcct gcgcaccct ctcagctcc tcaagtcttc 1920
tctgtctgagt caccttcgaa ccggaggctg tgagctggct gtcgtgacca cactggtgcc 1980
tctgtctgca tgacaacagc acactacgtc agtagtctc cctgggcact gagctccctc 2040
tttgcgggga gaagacagta atgaaaaatg acaagcatga ggcagagggg aagatcacgc 2100
ttgggtgggt caggagcatg gaggtgctct taatgctctc aatgagaaag ggttaacggt 2160

cctggttgca ggaatagctg agtcagaggt ggggttctc ccactcccc accccaccc 2220
tttaccatt agggacctc ttgccttgct ctgtctactc tgctctgggt ggtcattgtg 2280
aaaagccgc accaaccatg ccagtggcag ccagacgagg acacagcctg gctctgggtc 2340
ccagcaggaa aggcaatccc agaaaggcag ggtcaggac tggagtctg tgggtgcttt 2400
ttaagcaaag attatcacca ggcaggctaa acttagcaac cggcttttag ctagaagggc 2460
agggggctgg tgtcaggta tgcctggcca gcaaagggc ccgggatccc cctcccatgc 2520
acctgtgat gggccaaggc caccacccc cacccttc cttacaagtg ttcagacccc 2580

tccatccca cactcacaaa cctggccctc tgccctccta ccagaagaat ggatccctg 2640
tgggaggggg caggggacct gtccaccg tgtgccaag acctctttc ccacttttc 2700
cctctcttg actcacctg cctcaatat ccccgccgc agccagtga aggaggtccc 2760
tggctcctgg ctgcctgca cgtcccagg cggggaggga cttccgcct cactcccgc 2820
tcttcgccc aggttgatg gaatgaaagg cacactgtct ctctccctag gcagcacagc 2880
ccacaggttt ccaggagtgc ctttgtgga gcctctggg ccccaccag ccatcctgc 2940
ctccgctgg ggccccagc cggagagagc cgttgtgca cacagggccg ggattgtctg 3000

ccctaattat caggtccagg ctacagggt gcaggacatc gtgaccttc gtgcagaaac 3060
ctccccctc cctcaagcc gcctccgag cctcttct ctccagccc ccagtgccca 3120
tgcccagtg ccagcccag gcctcgttc cagagatgcc aggagccatc gcgaataat 3179

<210> 6
<211> 3238
<212> DNA
<213> Artificial Sequence

<220><223> 3'-GAPDH fragment

<400> 6

tggcagtact aagctggcac cactacttca gagaacaagg ccttttctc tcctcgtcc	60
agtcctaggc tatctgctgt tggccaaaca tggaagaagc tattctgtgg gcagccccag	120
ggaggctgac aggtggagga agtcagggtc cgcactgggc tctgacgtg actggttagt	180
ggagctcagc ctggagctga gctgcagcgg gcaattccag cttggcctcc gcagctgtga	240
ggtcttgagc acgtgctcta ttgctttctg tgccctcgtg tcttatctga ggacatcgtg	300
gccagccccct aaggtcttca agcaggattc atctaggtaa accaagtacc taaaaccatg	360
cccaaggcgg taaggactat ataatgttta aaaatcggta aaaatgccca cctcgcatag	420
ttttgaggaa gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg	480
gaacttgggt aggggcaagg cgtgtagctg ggacctaggt ccagaccctt ggctctgcca	540
ctgaacggct cagttgcttt gggcagttac tcccgggcct cactttgcac gtgtgcttac	600
ctagtggaga caaaagtaca tacctcggta gagcgcgcac gcctgtaacc ccagcacttt	660
gggaggccaa ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat	720
ggtgaaactc cgtctctact aaaattacaa aaatcagcca ggcttcatgg cacatgccta	780
tagtcccagc tacaggcatg ctgaagcagg agaatcgtt gcaccccgga ggcagaggct	840
gcagtgagct gagaccacac cactgcactc cagcctagge aacagagtat gagactccat	900
ctcaaaaaaa aaaaaagtac ctacctcaga gttcaacta gtgaatatta ggaagtgctt	960
gagacagtga caccaaagtg cacaataaat actgccagt ttcatattta ttaaagaatc	1020
catttgaatg tcagctcaac acagcctcct ataccgagge attgtgaacc gcatctcccc	1080
agctttctca ggcttttcca agaatcaggg aactgtagc ctgttggctc cagtgtatga	1140
cagacacgga ggaagcacat ctttagctga tacttaacaa gagaccctga gcgcacatac	1200
acccgcgcac acatgcatgg agcttcacct tctctgtcat tctgcagtga ccaggagagc	1260
aagagctccc acctcccttc aaaacactgt gcccatcccg ggactaagg cctctttaa	1320
gcacggcacc tccacgaggg agggccacag ccacatacac tccacctggc aggtggacag	1380
ctgagcacg tggaccatag cagggacaag gtccccggc cagccccaac gccctctgcc	1440
gctgacaggg acagaagccc tctccagctg cgtgtgtctc agaggccatg cgtagcctcc	1500
agctgcattc tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga	1560
gctggctccg gacaacaggg atggaggaaa ggtcccatc tcacattcct gatacgtgga	1620
caaggtgagg ggccgcaatc gctctggcag cattttaaag atggggaagt agcagacacc	1680
cacgcgtgaa ggcaggagag cccaactgt ggtggaaatg gcccagaat ggtagggcc	1740

agcctagctc cagacacccc agagccctgg agaagccaag actgaggag aaagcctgag 1800

 ggaggagcgc cccagtcctc agggaccggc ctggtgcaga gctgcagctg atgttccct 1860
 ctgtgcagcc ccacctctg cctcgtgag ctccctgctg cgagggcctc ggggtcaagg 1920
 gggaggcagg tcctatctc atggagctgt cagatgagac atcgcgatcg gagtctcag 1980
 cctcgttgg cgcgccggc gggtcgctaa gcgggaccgc agtgaaagca ggagactttc 2040
 tagaaaaaa caccagttgt caaccttggg gcaggcagga atcctgaaga cggacggcac 2100
 tcctcctct gctgcctcac cctctggcag cccgtgagaa gtaccggaag cgaggcgagg 2160
 gccgcgggat ggcgaggag cggcaggac tgaactctt ccaaaccac cctgacagg 2220

 aaatgggccc cgcctgtgc ttgggaactc agaggctgag gtcaggcatg atggctcacg 2280
 cctgtaattc cagcacttg ggaggcagag gcgggtggat cacgacgtca ggagtcaag 2340
 gcaagcctgg ccaacatgtt gaaacccat ctctactaaa aatgcaaaaa ttagccaggt 2400
 tlggtggcgg gcgccttga ggctgaggca agataatcac ttgaacctgg gaggcggagg 2460
 ttgtagttag ccaaaaaaa aaaaaaaga aatagctgaa gtcacagtag gagagaagct 2520
 gctgagcctc cagcacctg actctagggc ctggcttta tgtctatctg cagtattttt 2580
 gtgattttta aaaattcact ttcttgttc ggtgtaactt acacagggtc aaatgcacaa 2640

 atcatggccc tgactttaga taaaaatctg cccccacaac cttctgttc ctgcccagtt 2700
 tttaactgc cttaaccag gggaaccacc agagctggtg gccttgggag gtttcagccc 2760
 tcccgtcatg aatggacata gctcatcaa ctgccaaggg agagagctgt gggcttgggc 2820
 cagccccacc aggtaaectc caaaggcag cccacagca agatgtgacc cagtcattgc 2880
 ctgagggtct ctggggctgt gtccaacct ttctccccgc tgtgtcccc tggaggccc 2940
 catgcccagg ggagggcct accggtctcc agactgggtg atgagccgc gcaggtctc 3000
 catctgcag tccaggccgc gttcatgt gcacatctc atgtactgt gcagggtccg 3060

 gttcatgtcg ttcttggccg tggccaactc cagctgtggt ggggcaggca gggccatcgg 3120
 ggttaacagg tggcgttcac agccctctg ttccccgc caggaggcca acacccaag 3180
 agcagtggtt gggccggggg cccaggcagc cattactgaa ggctcagaag tacttgct 3238

 <210> 7
 <211> 3164
 <212> DNA
 <213> Artificial Sequence
 <220><223> Upstream GAPDH flanking region
 <400> 7

cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaacatcca	60
ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga	120
tttgtttctt cttctttttt aaatcacgtt ttcctgcctt ttctaggttc taggtaccag	180
cctctggctt ctacagcctc agacaatgac ttgtcacac cagagccccg ccgtactacc	240
cgtcggcatc caaacaccca gcagcgagct tccaaaaaga aacccaaagt tgtcttctca	300
agtgatgagt ccagtgagga aggtatgatg ctcccgctg ttcccgccg agaaggcaca	360
cagctagggg gcagagggtt ggtttccata ggacctgctg cgggggcctg agttagatg	420
ctctgcccc ctgccgcaga agggcctctc ctgtacagct tggattttat ttcttctgtg	480
cgggtgtggga ttgtctcact tgttctctga tatctattt ttaccatct ttgtgactca	540
gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga	600
gacaccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga	660
agtctgttcc tgcctccct gtgcagggtt tctgttaggg tgacctgga ttcaattct	720
gtttcccttg taaaatat ttctgtctct tttttttaa aaaaaaaag gccgggcact	780
gtggctcacg cctgtaatcc cagcactttg cgataccaag gcgggtggat aacctgaggt	840
agggagtctc agaccagcct gaccaacatg gagaaacccc atcttacta aaaataaaaa	900
attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagagc tgaggcagga	960
gaatgcctg aaccagagg cggaggttgt agtgagccga aatcacacca ttgactcca	1020
gcttgggcaa caatagcga cctccatctc aaattaaaa aaaaatgcct acacgtctt	1080
taaaatgcaa ggttttctt taaattagcc taactgaact gcgttgagct gttcaactt	1140
tggatatat gtttccaat ctcttgttt tctaatgaat aaatgtttt atatacttt	1200
agacatttt tcttaagctt gtcttgttt catctttcac attagcccag ttatcatcag	1260
cagagagagg gttatcagtg cagagagaga tgagttagcc cagagtcta gggcctgtcc	1320
cgggatggca gatgagctc ctgccccgtc actgccacct ttccctctc aacctctgga	1380
ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtgccta ggctccagat	1440
cagtgcctt gaaccggggg caattttgtc tgccagagga catctgaca cacctggggc	1500
ctgttttgtt gtcatagcct ataggggaag aatgctacca gcatttgtg gaagaggcca	1560
gggatgtggc tcaacatcct gcagtgcaca ggatggcccc tcaacaaaga atcacacggc	1620
ccacaatgtc aatagcgtc cagttagaaa aacctgtct agaccaaggg ttgctttctg	1680
ccgtgtgcct caccacccc cactcgtgt tcctaatac catctccaa ggttggcagc	1740

agaccggccc aggtctgtgg aagttcagat catgatcccc tccagctctg caggagacaa 1800

gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860

aggcgggctg cccctgaag cctgcgcacc cctctccagc tectcaagtc ttctctgctg 1920

agtcaccttc gaaccggagg ctgtgagctg gctgtcgtga ccacactggt gcctctgctg 1980

tcatgacaac agcacactac gtcagtagtg ctccctgggc actgagctcc ctctttgcgg 2040

ggagaagaca gtaatgaaaa atgacaagca tgaggcagag gggaagatca cgcttgggtg 2100

gtgcaggagc atggaggctc tcttaatgct ctcaatgaga aagggttaac ggtcctggtt 2160

gcaggaatag ctgagtcaga ggtggggctt cctccactcc cccacccac ccctttcacc 2220

attagggacc ttcttgctt gctcttgctc ctctgctctg ggtggtcatt gtgaaaagcc 2280

cgcaccaacc atgccagttg cagccagacg aggacacagc ctggctctgg gtcccagcag 2340

gaaaggcaat cccagaaagg cagggtcagg gactggagtc ctgtgggtgc tttttaagca 2400

aagattatca ccaggcagcg taaacttagc aaccggcttt tagctagaag ggcagggggc 2460

tgggtgtcagg ttatgtctgg ccagcaaaga ggcccgggat cccctccca tgcacctgct 2520

gatgggcaa ggccacccca cccaccccc ttctttacaa gtgttcagca ccctccatc 2580

ccacactcac aaacctggcc ctctgccctc ctaccagaag aatggatccc ctgtgggagg 2640

gggcagggga cctgttccca ccgtgtgcc aagacctctt tteccacttt tteectcttc 2700

ttgactcacc ctgcccctca tatccccgg cgagccaggt gaaaggaggt ccctggctcc 2760

tggctgcct gcacgtccca gggcggggag ggacttcgc cctcacgtcc cgctcttcgc 2820

cccagctgg atggaatgaa aggcacactg tctctctccc taggcagcac agcccacagg 2880

tttccaggag tgcctttgtg ggaggcctct ggccccccac cagccatcct gtctccgcc 2940

tggggccca gcccggagag agccctggt gcacacaggg ccgggattgt ctgccctaat 3000

tatcaggctc aggttacagg gctgcaggac atcgtgacct tccgtgcaga aacctcccc 3060

tccccctcaa gccgcctccc gagcctctt cctctccagg cccccagtgc ccagtgccca 3120

gtgccagcc caggcctcgg tcccagagat gccaggagcc atcg 3164

<210> 8

<211> 3224

<212> DNA

<213> Artificial Sequence

<220><223> Downstream GAPDH flanking region

<400> 8

actaagctgg caccactact tcagagaaca aggccttttc ctctctcgc tccagtccca 60

ggctatctgc tgttggccaa acatggaaga agctattctg tgggcagccc cagggaggct	120
gacaggtgga ggaagtcagg gctcgcactg ggctctgacg ctgactggtt agtggagctc	180
agcctggagc tgagctgcag cgggcaattc cagcttggcc tccgcagctg tgaggtcttg	240
agcacgtgct ctattgcttt ctgtgccctc gtgtcttata tgaggacatc gtggccagcc	300
cctaaggtct tcaagcagga ttcatctagg taaaccaagt acctaaaacc atgccaagc	360
cggtaaggac tatataatgt ttaaaaatcg gtaaaaatgc ccacctcgca tagttttgag	420
gaagatgaac tgagatgtgt cagggtgact tatttccatc atcgtcctta ggggaacttg	480
ggtaggggca aggcgtgtag ctgggaccta ggtccagacc cctggctctg ccactgaacg	540
gctcagttgc ttggggcagt tactcccggg cctcactttg cacgtgtgct tacctagtgg	600
agacaaaagt acataacctg gtagagcgcg cagcctgta accccagcac ttggggaggc	660
caagtggggt gtatcacctg aggtcaggag tttagacca gcctggccaa catggtgaaa	720
ctccgtctct actaaaatta caaaaatcag ccaggcttca tggcacatgc ctatagtcce	780
agctacaggc atgctgaagc aggagaatcg ctgaccccc ggaggcagag gctgcagtga	840
gctgagacca caccactgca ctccagccta ggcaacagag tatgagactc catctcaaaa	900
aaaaaaaaag tacctacctc agagttcaaa ctagtgaata ttaggaagtg cttgagacag	960
tgacacaaaa gtgcacaata aatactcgcc agtttcatta ttattaaaga atccatttga	1020
atgtcagctc aacacagcct cctataccga ggcatgtga accgcatctc cccagcttct	1080
ccaggctttt ccaagaatca gggacactgt agcctgttgg tctcagtga tgacagacac	1140
ggaggaagca catcttttagc tgatacttaa acagagaccc tgagcgaca tacaccgcg	1200
cacacatgca tggagcttca ctttctctgt cattctgcag tgaccaggag agcaagagct	1260
cccacctccc ttcaaaacac tgtgcccac cggggcacta aggcctcttt aaagcacggc	1320
acctccagca gggaggggcca cagccacata cactccacct ggcaggtgga cagcgtgagc	1380
acgtggacca tagcagggac aagggtgccc ggccagcccc aacgccctct gccgctgaca	1440
gggacagaag cctctccag ctgcgtgtgc tgcagaggcc atgcgtagcc tccagctgca	1500
ttctattcca ctccagtgcc tgggccagtt agcaccagtg tggaagacag tgagctggct	1560
ccggacaaca gggatggagg aaaggtccca cattcacatt cctgatacgt ggacaagggtg	1620
aggggcccga atcgctctgg cagcatttta aagatgggga agtagcagac acccacgcgt	1680
gaaggcagga gagccccaac tgtgttgaa atggccccag aatggtaggg ccaagcctag	1740
ctccagacac cccagagccc tggagaagcc aagactgagg gagaaagcct gagggaggag	1800

cgccccagtc cccagggacc ggcctgggtgc agagctgcag ctgatgttcc cctctgtgca 1860

gccccaccct ctgcctcgct gagctccctg ctgcgagggc ctgggtgca agggggaggc 1920

aggtctctat ctcatggagc tgtcagatga gacatcgca tcggagtcct cagcctcgct 1980

tggcggcggc ggcgggtcgc taagcgggac cgagtgaaa gcaggagact ttctagaaaa 2040

aaacaccagt tgtcaacctt ggggcaggca ggaatcctga agacggacgg cactcctcct 2100

cctgtctgct caccctctgg cagccctgta gaagtaccgg aagcgagggc ggggccgagg 2160

gatggcgagg gagcggcagg gactgaactc tctccaaacc caccctgaca gggaaatggg 2220

ccccgcctgt gtcttgggaa ctgagaggct gaggtcaggc atgatggctc acgcctgtaa 2280

ttccagcact ttgggaggca gaggcgggtg gatcacgacg tcaggagtcc aaggcaagcc 2340

tggccaacat ggtgaaaccc catctctact aaaaatgcaa aaattagcca ggtgtggtgg 2400

cgggcgcctt ggaggtctag gcaagataat cacttgaacc tgggaggcgg aggtttgtagt 2460

gagccaaaaa aaaaaaaaaa agaaatagct gaagtcacag taggagagaa gctgctgagc 2520

ctccagcacc ctgactctag ggccttggct ttatgtctat ctgcagtatt ttgtgattt 2580

ttaaaaaatc actttcttgt tgcggtgtaa ctacacagg gtcaaatgca caaatcatgg 2640

ccctgacttt agataaaaat ctgccccac aaccctctg ttccttgcca gtttttaaac 2700

tgctctaac cagggaacc accagagctg gtggccttgg gaggtttcag ccctcccgctc 2760

atgaatggac atagctcatc caactgcaa gggagagagc tgtgggtctg ggccagcccc 2820

accagtaac tcccaaaggg cagcccccaca gcaagatgtg acccagtcac tgctgaggg 2880

tctctggggc tgtgttccaa cttttctccc cgtgtgtcc ccctggaagg ccccatgccc 2940

aggggaggcg cctaccggtc tccagactgg gtgatgagcc ggcggcaggt ctccatctgc 3000

acgtccagc cgcgcttcat gctgcacatc tccatgtact cgtgcagggt ccggttcatg 3060

tcgttcttgg ccgtggccaa ctccagctgt ggtggggcag gcagggccat cggggttaac 3120

aggtggcggt cacagcgct ctgttgcccc cgccaggagg ccaacacgcc aagagcagtg 3180

gctgggccgg gggcccaggc agccattact gaaggctcag aagt 3224

<210> 9

<211> 511

<212> DNA

<213> Artificial Sequence

<220><223> Fragment 1

<400> 9

cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60

ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga	120
tttgtttctt cttctttttt aaatcacgtt ttcctgcctt ttctaggttc taggtaccag	180
cctctggctt ctacagcctc agacaatgac ttgtcacac cagagccccg ccgtactacc	240
cgtcggcatc caaacaccca gcagcgagct tccaaaaaga aacccaaagt tgtcttctca	300
agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccgcccg agaaggcaca	360
cagctagggt gcagagggt ggtttccata ggacctgctg cgggggcctg agtgtagatg	420
ctctgcccc ctgccgcaga agggcctctc ctgtacagct tggattttat ttcttctgtg	480
cgggtgtggga ttgtctcact tgttctctga t	511
<210> 10	
<211> 2653	
<212> DNA	
<213> Artificial Sequence	
<220><223> Fragment 2	
<400> 10	
atctattttt tcaccatctt tgtgactcag ctttttctta ttcttttaat tctttgcata	60
gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga	120
gcatcggtc gcaggcacag atcctaggaa gtctgttctt gtctccctg tgcagggtat	180
cctgtagggt gacctggaat tcgaattctg ttcccttgt aaaatattg tctgtctctt	240
ttttttaaaa aaaaaaagg cggggcactg tggtcacgc ctgtaatccc agcactttgc	300
gataccaagg cgggtggata acctgaggta gggagtctga gaccagcctg accaacaatgg	360
agaaacccca tctctactaa aaataaaaaa ttagccgggc gtattggcgt gcgcctgtaa	420
tcccagctac tcaagaggct gaggcaggag aatcgctga acccagaggc ggaggttgta	480
gtgagccgaa atcacaccat tgcactccag ctgggcaac aatagcgaac ctccatctca	540
aattaaaaaa aaaaatgccta cacgtctttt aaaatgcaag gctttctctt aaattagcct	600
aactgaactg cgttgagctg ctccaacttt ggaatatatg ttgccaatc tccttgtttt	660
ctaataaata aatgttttta tatactttta gacatttttt cctaagcttg tctttgtttc	720
atctttcaca ttagccagct tcatgcagc agagagaggg ttatcagtgc agagagagat	780
gagttagccc agagtccatg ggcctgtccc gggatggcag atgagcttcc tgccccgtca	840
ctgccacctt tcccctctca acctctggac cctgcacagt gaccagacag cctctctggg	900
gagaattatg cagtgcctag gctccagatc agtgcttctg aaccgggggc aattttgtct	960
gccagaggac atctgacaac acctgggggc tgttttgttg tcatagccta taggggaaga	1020

atgctaccag catttgtggg aagaggccag ggatgtggct caacatcctg cagtgcacag 1080
gatggccct caacaaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa 1140
acctgtctta gaccaagggt tgccttctgc cgtgtgcctc accccacccc cactcgtgtt 1200
ccctaattcc atctccaaag gttggcagca gaccggccca ggctcgtgga agttcagatc 1260
atgatccct ccagctctgc aggagacaag acctgtctcc cagcattcct cattgttccc 1320

gggtctgcag agggcgtgag ctatctgca ggcgggctgc cccctgaagc ctgcgcaccc 1380
ctctccagct cctcaagtct tctctgctga gtcaccttcg aaccggaggc tgtgagctgg 1440
ctgtcgtgac cacactggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc 1500
tccctgggca ctgagctccc tctttgcggg gagaagacag taatgaaaaa tgacaagcat 1560
gaggcagagg ggaagatcac gcttgggttg tgcaggagca tggagggtct cttaatgctc 1620
tcaatgagaa aggggttaacg gtctggttg caggaatagc tgaagcagag gtggggcttc 1680
ctccactccc ccacccacc cctttacca ttagggacct tcttgcttg ctcttgctac 1740

tctgctctgg gtggtcattg tgaaaagccc gcaccaacca tgccagtggc agccagacga 1800
ggacacagcc tggctctggg tcccagcagg aaaggcaatc ccagaaagc agggtcaggg 1860
actggagtcc tgtgggtgct ttttaagca agattatcac caggcaggct aaacttagca 1920
accggctttt agctagaagg gcagggggct ggtgtcaggt tatgctgggc cagcaaagag 1980
gccccggatc cccctccat gcacctctg atgggccaag gccacccac cccacccct 2040
tccttacaag tgttcagcac cctcccatcc cacactcaca aacctggccc tctgcccctc 2100
taccagaaga atggatcccc tgtgggaggg ggcaggggac ctgttccac cgtgtgcca 2160

agacctcttt tccactttt tccctcttct tgactacccc tgccctcaat atccccggc 2220
gcagccagt aaaggagtc cctggtcct ggctcgctg cacgtcccag ggcggggagg 2280
gacttccgc ctcagctcc gctcttcgcc ccaggctgga tggaatgaaa ggcacactgt 2340
ctctctccct aggcagcaca gccacaggt ttccaggagt gcctttgtgg gaggcctctg 2400
ggccccacc agccatcctg tcttcgctt ggggccccag cccggagaga gccgctggtg 2460
cacacagggc cgggattgtc tgccctaatt atcaggcca ggctacaggc ctgcaggaca 2520
tcgtgacctt ccgtgcagaa acctccccct cccctcaag ccgctcccg agcctcttc 2580

ctctccaggc cccagtgcc cagtgccag tgeccagccc aggcctcggt cccagagatg 2640
ccaggagcca tcg 2653

<210> 11
<211> 1966
<212> DNA

<213> Artificial Sequence

<220><223> Fragment 3

<400> 11

cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca	60
ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga	120
tttgtttctt cttctttttt aaatcacgtt ttcctgcctt ttctaggttc taggtaccag	180
cctctggctt ctacagcctc agacaatgac ttgtgcacac cagagccccg ccgtactacc	240
cgteggcatc caaacaccca gcagcgagct tccaaaaaga aacccaaagt tgtcttctca	300
agtgatgagt ccagtgagga aggtatgatg ctcccgctg ttcccgccg agaaggcaca	360
cagctagggt gcagagggt ggtttccata ggacctgctg cgggggcctg agttagatg	420
ctctgcccc ctgccgaga agggcctctc ctgtacagct tggattttat ttcttctgtg	480
cgggtgtggga ttgtctcact tgttctctga tatctatctt ttcacatct ttgtgactca	540
gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga	600
gacaccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga	660
agtctgttcc tgcctccct gtgcagggt tctgtaggg tgacctggaa ttcgaattct	720
gtttcccttg taaaatatct gtctgtctct tttttttaa aaaaaaaaaag gccgggcact	780
gtggctcacg cctgtaatcc cagcactttg cgataccaag gcgggtggat aacctgaggt	840
agggagtctg agaccagcct gaccaacatg gagaaacccc atctctacta aaaataaaaa	900
attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagaggc tgaggcagga	960
gaatgcctg aaccagagg cggaggttgt agtgagccga aatcacacca ttgactcca	1020
gcttgggcaa caatagcgaa cctccatctc aaattaaaa aaaaatgcct acacgtctt	1080
taaaatgcaa ggttttctct taaattagcc taactgaact gcgttgagct gtttcaactt	1140
tggaatatat gtttgccaat ctcttgttt tctaataaat aaatgtttt atatacttt	1200
agacatttt tcttaagctt gtctttgttt catctttcac attagcccag ttatcatgag	1260
cagagagagg gttatcagtg cagagagaga tgagttagcc cagagtccta gggcctgtcc	1320
cgggatggca gatgagcttc ctgccccgtc actgccacct ttccctctc aacctctgga	1380
ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtgccta ggctccagat	1440
cagtgtctt gaaccggggg caattttgtc tgccagagga catctgacaa cacctggggc	1500
ctgttttgtt gtcatagcct ataggggaag aatgctacca gcatttgttg gaagaggcca	1560
gggatgtggc tcaacatcct gcagtgcaca ggatggcccc tcaacaaaga atcacacggc	1620
ccacaatgtc aatagcgtca cagttgagaa aacctgtct agaccaaggg ttgctttctg	1680

ccgtgtgcct caccaccacc ccactcgtgt tccctaatacc catctccaaa ggttggcagc 1740
agaccggccc aggtctgtgg aagttcagat catgatcccc tccagctctg caggagacaa 1800
gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860
aggcgggctg cccctgaag cctgcgcacc cctctccagc tcccaagtc ttctctgctg 1920

agtcaccttc gaaccggagg ctgtgagctg gctgtcgtga ccacac 1966

<210> 12
<211> 1198
<212> DNA
<213> Artificial Sequence
<220><223> Fragment 4
<400> 12

tggtgcctct gctgtcatga caacagcaca ctacgtcagt agtgcctcct gggcactgag 60
ctccctcttt gcggggagaa gacagtaatg aaaaatgaca agcatgaggc agaggggaag 120
atcacgcttg ggtggtgcag gagcatggag gtgctcttaa tgctctcaat gagaaagggt 180
taacggtcct ggttgcagga atagctgagt cagagggtgg gcttctcca ctccccacc 240
ccacccttt caccattagg gacctcttg ccttgctctt gctactctgc tctgggtggt 300

catttgaaa agcccgacc aaccatgcca gtggcagcca gacgaggaca cagcctggct 360
ctgggtccca gcaggaaagg caatcccaga aaggcagggt cagggactgg agtcctgtgg 420
gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg cttttagcta 480
gaagggcagg gggctgggtg caggttatgc tgggccagca aagaggcccg ggatccccct 540
cccatgcacc tgctgatggg ccaaggccac cccacccac ccccttcctt acaagtgtc 600
agcacctcc catcccacac tcacaaacct ggcctctgc cctctacca gaagaatgga 660
tccctgtgg gagggggcag gggacctgtt cccaccgtgt gcccaagacc tcttttcca 720

cttttccct cttcttgact caccctgcc tcaatatccc ccggcgcagc cagtgaagg 780
gagtccttgg ctcttgctc gcctgcacgt cccaggcggg ggagggactt ccgcctcac 840
gtcccgtct tgcgccagg ctggatggaa tgaaaggcac actgtctctc tccctaggca 900
gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca 960
tcctgtctc cgctggggc cccagcccgg agagagccgc tggtgcacac agggccggga 1020
ttgtctgcc taattatcag gtccaggcta cagggtgca ggacatcgtg acctccgtg 1080
cagaaacct cccctcccc tcaagccgc tcccagacct ccttctctc caggcccca 1140

gtgccagtg cccagtgcc agcccaggcc tcggtcccag agatgccagg agccatcg 1198

<210> 13
 <211> 259
 <212> DNA
 <213> Artificial Sequence
 <220><223> Fragment 8
 <400> 13
 cctctggggc cccaccagcc atcctgtcct ccgcctgggg ccccagcccg gagagagccg 60
 ctggtgcaca cagggccggg attgtctgcc ctaattatca ggtccaggct acagggtgc 120
 aggacatcgt gaccttcgt gcagaaacct cccctcccc ctcaagccgc ctcccagacc 180
 tccttctctt ccaggccccc agtgcccagt gcccagtgcc cagcccaggc ctcggtccca 240
 gagatgccag gagccatcg 259

<210> 14
 <211> 1947
 <212> DNA
 <213> Artificial Sequence
 <220><223> Fragment 9
 <400> 14
 cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60
 ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120
 ttgttttctt cttctttttt aaatcacgtt ttctgcctt ttctaggttc taggtaccag 180
 cctctggctt ctacagcctc agacaatgac ttgtcacac cagagccccc cgtactacc 240
 cgtcgcatc caaacaccca gcagcgagct tcaaaaaaga aacccaaagt tgtcttctca 300
 agtgaatgag ccagtgagga aggtatgatg ctcccgcctg ttcccgcccg agaaggcaca 360

cagctagggt gcagagggt ggtttccata ggacctgtg cgggggcctg agttagatg 420
 ctctgcccc ctgccgaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480
 cgggtgtggga ttgtctcact tgttctctga tatctatctt ttaccatct ttgtgactca 540
 gctttttctt attcctttta ttctttgcat agatctttca gcagagatga cagaagacga 600
 gacaccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660
 agtctgttcc tgcctccct gtgcagggt tctgtaggg tgacctgaa ttcgaaccgg 720
 aggctgtgag ctggctgtcg tgaccacact ggtgcctctg ctgtcatgac aacagcacac 780

tacgtcagta gtgtccctg ggcaactgagc tccctctttg cggggagaag acagtaatga 840
 aaaatgacaa gcatgaggca gaggggaaga tcacgttgg gtggtgcagg agcatggagg 900

tgctcttaat gctctcaatg agaaagggtt aacggtcctg gttgcaggaa tagctgagtc 960
agaggtgggg ctctctccac tccccaccc caccctttc accattaggg accttcttgc 1020
cttgctcttg ctactctgct ctgggtggtc attgtgaaaa gcccgacca accatgccag 1080
tggcagccag acgaggacac agcctggctc tgggtcccag caggaaaggc aatcccagaa 1140
aggcagggtc agggactgga gtcctgtggg tgctttttaa gcaaagatta tcaccaggca 1200

ggctaaactt agcaaccggc ttttagctag aagggcaggg ggctgggtgc aggttatgct 1260
gggccagcaa agaggcccg gatccccc ccatgcacct gctgatgggc caaggccacc 1320
ccacccacc ccttctcta caagtgttca gcacctccc atccacact cacaacctg 1380
gccctctgcc ctctaccag aagaatggat cccctgtggg agggggcagg ggacctgttc 1440
ccacctgtg cccaagacct ctttccac ttttccctc ttcttgactc acctgcct 1500
caatatcccc cggcgagcc agtgaaagg agtccctggc tctggctcg cctgcacgtc 1560
ccaggcgagg gagggacttc cgccctcacg tcccgtctt cgccccaggc tggatggaat 1620

gaaaggcaca ctgtctctt ccttaggcag cacagccac aggtttccag gagtgccttt 1680
gtgggaggcc tctgggccc caccagccat cctgtctcc gcctggggcc ccagcccgga 1740
gagagccgtt ggtgcacaca gggccgggat tgtctgcct aattatcagg tccaggctac 1800
agggtgcag gacatcgtga cttccgtgc agaaacctc cctccccct caagccgct 1860
cccagcctc ctctctcc agggccccag tgcccagtgc ccagtgccca gccaggcct 1920
cgggtcccaga gatgccagga gccatcg 1947

<210> 15

<211> 1436

<212> DNA

<213> Artificial Sequence

<220><223> Fragment 11

<400> 15

atctatttt tcaccatctt tgtgactcag ctttttctta ttctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
gcatcggtc gcaggcacag atcctaggaa gtctgttct gtctccctg tgcagggtat 180
cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggtgtcgt gaccacactg 240
gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gcactgagct 300
ccctctttgc ggggagaaga cagtaatgaa aatgacaag catgaggcag aggggaagat 360

cacgcttggg tgggtcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420

acggtcctgg ttgcaggaat agctgagtca gaggtggggc ttcttccact cccccacccc 480
 acccttttca ccattagga cttctttgcc ttgtcttgc tactctgtc tgggtgggtca 540
 ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
 ggggtcccagc aggaaaggca atcccagaaa ggcaggggtca gggactggag tcctgtgggt 660
 gctttttaag caaagattat caccaggcag gctaaactta gcaaccggct tttagctaga 720
 agggcagggg gctgggtgtca ggttatgtg ggccagcaaa gagggccggg atccccctcc 780

catgcacctg ctgatgggcc aaggccaccc cccccaccc ctttcttac aagtgttcag 840
 caccctccca tccacactc acaaacctgg ccctctgccc tctaccaga agaattggatc 900
 ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc ttttccact 960
 ttttccctct tcttgactca ccctgccctc aatatcccc ggccgagcca gtgaaaggga 1020
 gtccctggct cctggctgc ctgcacgtcc cagggcgggg agggacttcc gccctcacgt 1080
 cccgtcttc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
 acagcccaca ggtttccagg agtgcctttg tgggaggcct ctgggcccc accagccatc 1200

ctgtctccg cctggggccc cagcccgag agagccgtg gtgcacacag ggccgggatt 1260
 gtctgccta attatcaggt ccaggctaca gggctgcagg acatcgtgac cttccgtgca 1320
 gaaacctccc cttccccctc aagccgctc ccgagcctc ttctctcca gggccccagt 1380
 gcccagtgcc cagtgccag cccaggcctc ggtcccagag atgccaggag ccatcg 1436

<210> 16

<211> 1177

<212> DNA

<213> Artificial Sequence

<220><223> Fragment 17

<400> 16

atctatttt tcaccatctt tgtgactcag ctttttctta ttctttaat tctttgcata 60

gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
 gcatcggctc gcaggcacag atcctaggaa gtctgttctt gtcctccctg tgcagggtat 180
 cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
 gtgcctctgc tgcatgaca acagcacact acgtcagtag tgctccctgg gactgagct 300
 ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
 cacgcttggg tgggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
 acggtcctgg ttgcaggaat agctgagtca gaggtggggc ttcttccact cccccacccc 480

acccttttca ccattaggga ctttcttgcc ttgctcttgc tactctgctc tgggtgggtca 540
 ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
 ggggtcccagc aggaaaggca atcccagaaa ggcagggtca gggactggag tcctgtgggt 660
 gctttttaag caaagattat caccaggcag gctaaactta gcaaccggct tttagctaga 720
 agggcagggg gctgggtgtca gggttatgctg ggccagcaaa gagggccggg atccccctcc 780
 catgcacctg ctgatgggcc aaggccaccc caccceaccc ctttccttac aagtgttcag 840
 caccctccca tcccacactc acaaacctgg ccctctgccc tcctaccaga agaatggatc 900

ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc ttttccact 960
 ttttccctct tcttgactca ccctgccctc aatatcccc ggccgagcca gtgaaaggga 1020
 gtccctggct cctggctcgc ctgcacgtcc cagggcgggg agggacttcc gcctcacgt 1080
 cccgctcttc gcccaggtt ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
 acagcccaca ggtttccagg agtgcctttg tgggagg 1177

<210> 17
 <211> 18106
 <212> DNA
 <213> Homo sapiens
 <400> 17

gtgagcagca caggacactt caatgcctgt tgggttctgg gctggctaag acatctgccg 60

gccctgggca gcatacggct cttgcagtca ctttccgctc ctcttatcc ccagctgggt 120
 tgcaacaaaa ttgccagagt gacctaaagc cagatctttg tctccagttc ttttttatt 180
 actcaaaaa cacaacaaa gcagcatctc atccaattct tgtttgtttg tttttaatag 240
 tttttatatt tcagagcagt tttaggttca aagcaaaatt gagcagaaag tacaggaggt 300
 tcccttctac cccttgcccc tacacatcac agccttcccc acctcaaca tcctgcacca 360
 ggggtggcaca ttgtttacag ctgaacctac acttacacat catctcctaa agtcatggtt 420
 taccttggag ttactgtcac gtaatgacat gtaccacca ttgcagtatc atacagaaga 480

gtttcactgc cttacaaac ccctgcactc cacctattta tccctctctc ccacaaccc 540
 ctgatctttt tactgttgcc atcactttgt cttttccaga atgtatcatt ggaatgatec 600
 gggtatggagc ctctcactt tggcttctta gtaatgtgcg ttttaaggcct ccatgtcttc 660
 catggccttg tttcttttta atcagaagta actgttttca ggcctgctct gaatctcctt 720
 ttctccctcc aggtataat agatgaattt gagcagaagc ttggggcctg tcataccaga 780
 ggtttggatg gaatcaagga gcttgagatt ggccaagcag gtagccagag agcgccatca 840

gccaagaaac catccactgg tacgtaaggc agcctgtgcg ggcgagacca gactgggccc	900
tcccctcctg cagtgatttg tttcttcttc ttttttaaata cacgttttcc tgccttttct	960
aggttctagg taccagcctc tggcttctac agcctcagac aatgactttg tcacaccaga	1020
gccccgccgt actaccgctc ggcatccaaa caccagcag cgagcttcca aaaagaaacc	1080
caaagtgtc tttcaagtg atgagtcag tgaggaaggt atgatgtcc cgcctgttcc	1140
cggccgagaa ggcacacagc tagggtgcag agggctggtt tccataggac ctgctgcggg	1200
ggcctgagtg tagatgtct gcccactgc cgcagaaggg cctctcctgt acagcttga	1260
ttttatttct tctgtgcggt gtgggattgt ctacttgtt ctctgatac tatttttca	1320
ccatctttgt gactcagctt tttcttattc ctttaattct ttgcatagat ctttcagcag	1380
agatgacaga agacgagaca cccaagaaaa caactcccat tctcagagca tcggctcgca	1440
ggcacagatc ctaggaagtc tgttctgtc ctccctgtgc agggatcct gtagggtgac	1500
ctggaattcg aattctgtt ccttgtaaa atatttgtct gtctctttt tttaaaaaa	1560
aaaaaggccg ggcactgtgg ctacgcctg taatccagc actttgcgac accaaggcgg	1620
gtggataacc tgaggtaggg agttcgagac cagcctgacc aacatggaga aaccccatct	1680
ctactaaaaa taaaaaatta gccggcgta ttggcgtgcg cctglaatcc cagctactca	1740
agagctgag gcaggagaat cgctgaacc cagaggcgga ggtttagtg agccgaaatc	1800
acaccattgc actccagctt gggcaacaat agcgaacctc catctcaaat taaaaaaaa	1860
atgctacac gctctttaa atgcaaggct tttctttaa ttagcctaac tgaactgcgt	1920
tgagctgctt caactttga atatatgtt gccaatctc ttgttttcta atgaataaat	1980
gtttttatat acttttagac attttttct aagcttgtct ttgtttcatc ttacacatta	2040
gcccagtttc atgcagcaga gagagggtta tcagtgcaga gagagatgag tgagcccaga	2100
gtcctagggc ctgtcccgga atggcagatg agcttctgc cccgtcactg ccacctttcc	2160
cctctcaacc tctggaccct gcacagtgc cagacagcct ctctggggag aattatgcag	2220
tgcctaggct ccagatcagt gcttctgaac cgggggcaat tttgtctgcc agaggacatc	2280
tgacaacacc tggggcctgt tttgtgtca tagcctatag gggaagaatg ctaccagcat	2340
ttgtgggaag aggccaggga tgtggtcaa catctgcag tgcacaggat ggcctcaa	2400
caaagaatca cacggccac aatgtcaata gcgtcacagt tgagaaaacc tgcctagac	2460
caagggttgc tttctgccgt gtgcctcacc ccacccac tcgtgttccc taatccatc	2520
tccaaaggtt ggcagcagac cggcccaggc tcgtggaagt tcagatcatg atccctcca	2580

gctctgcagg agacaagacc tgtctcccag cattctcat tgttcccggg tctgcagagg	2640
gcgtgagcta tgctgcaggc gggctgcccc ctgaagcctg cgcacccctc tccagtcct	2700
caagtcttct ctgctgagtc accttcgaac cggaggtctgt gagctggctg tctgaccac	2760
actggctgct ctgctgtcat gacaacagca cactacgtca gtagtgctcc ctgggcactg	2820
agctccctct ttgcggggag aagacagtaa tgaaaaatga caagcatgag gcagaggga	2880
agatcacgct tgggtgggtc aggagcatgg aggtgctctt aatgctctca atgagaaagg	2940
gttaacggtc ctggttgagc gaatagctga gtcagagggtg gggcttcctc cactcccca	3000
ccccacccct ttcaccatta gggaccttct tgccttctc ttgctactct gctctgggtg	3060
gtcattgtga aaagcccga ccaaccatgc cagtggcagc cagacgagga cacagcctgg	3120
ctctgggtcc cagcaggaaa ggcaatcca gaaaggcagg gtcagggact ggagtctgt	3180
gggtgctttt taagcaaaga ttatcaccag gcaggctaaa cttagcaacc ggcttttagc	3240
tagaaggga gggggctggt gtcaggttat gctgggcagc caaagaggcc cgggatcccc	3300
ctcccatgca cctgctgatg ggccaaggcc accccacccc accccttcc ttacaagtgt	3360
tcagcacct cccatccac actcacaac ctggccctct gccctctac cagaagaatg	3420
gatccctgt gggagggggc aggggacctg tccccacct gtgccaaga cctctttcc	3480
cactttttcc ctctcttga ctcacctgc cctcaatgc ccccgcgca gccagtgaag	3540
gggagtcct ggctcctggc tcgctgcac gtcccagggc ggggaggac ttccgcctc	3600
acgtccctgt cttcgccca ggctggatgg aatgaaaggc acactgtctc tctccctagg	3660
cagcacagcc cacaggttc caggagtgc tttgtgggag gcctctgggc cccaccagc	3720
catcctgtcc tccgcctggg gcccagccc ggagagagcc gctggtgcac acagggccgg	3780
gattgtctgc cctaattgc aggtccaggc tacagggtg caggacatcg tgacctccg	3840
tgagaaacc tccccctcc cctcaagccg cctcccgagc ctccttctc tccaggcccc	3900
cagtgccag tgcccagtc ccagcccagg cctcggtccc agagatgcca ggagccagga	3960
gatggggagg gggaagtggg ggctgggaag gaaccacggg cccccgccg aggcccatgg	4020
gccccctcta ggcttttgc tgagcagtc ggtgtcacta ccgcagagcc tcgaggagaa	4080
gttcccaac ttccccct ctcagcctt gaaagaaaga aaggggaggg ggcaggccgc	4140
gtgcagccg gagcgtgtc gggctccgc tccaattccc catctcagtc gttccaaag	4200
tcctctgtt tcatccaagc gtgtaagggt ccccgctctt gactccctag tgtcctgtg	4260
cccacagtc agtcctggga accagcaccg atcacctccc atcgggcca tctcagtc	4320
ttcccccta cgtcggggcc cacacgtctg gtgcgtgccc agttgaacca ggcggtg	4380
gaaaaaaaa agcggggaga aagtagggcc cggctactag cggttttacg ggcgcacgta	4440

gctcaggcct caagaccttg ggctgggact ggctgagcct ggccgggaggc ggggtccgag	4500
tcaccgcctg ccgccgcgcc cccggtttct ataaattgag cccgcagcct cccgttcgc	4560
tctctgctcc tctgtttcga cagtcagccg catcttcttt tgcgtcgcca ggtgaagacg	4620
ggcggagaga aacccgggag gctagggacg gcctgaaggc ggccggggcg ggccgaggcc	4680
ggatgtgttc gcgccgtgc ggggtgggccc cgggcggcct ccgcattgca ggggcgggcg	4740
gaggacgtga tgcggcgccg gctgggcatg gaggcctggt gggggagggg aggggaggcg	4800
tgtgtgtcgg ccggggccac taggcgtca ctgttctctc cctccgcga gccgagccac	4860
atcgctcaga caccatgggg aaggtgaagg tcggagtcaa cgggtgagtt cgcgggtggc	4920
tggggggccc tgggctgcga ccgccccga accgcgtcta cgagccttg cggctccggg	4980
tctttgcagt cgtatggggg cagggtagct gtccccgca aggagagctc aaggtcagcg	5040
ctcgacctg gcggagcccc gcaccaggc tgtggcgccc tgtgcagctc cgcccttgcg	5100
gcgccatctg cccggagcct ctttccctta gtccccagaa acaggaggtc cctactcccg	5160
cccagatcc cgaccgggac ccctaggtgg gggacgcttt ctttctttc gcgctctgcg	5220
gggtcacgtg tcgcagagga gccctcccc caccgcctcc ggccaccgag ccccgggat	5280
gctagtgcgc agcgggtgca tccctgtccg gatgctgcgc ctgcggtaga gcggccgcca	5340
tgttgcaacc gggaaggaaa tgaatgggca gccgttagga aagcctgccg gtgactaacc	5400
ctgcgtcct gccctgatgg gtggagtcgc gtgtggcggg gaagtcaggt ggagcgaggc	5460
tagctggccc gattttctct ccgggtgatg cttttcttag attattctct ggtaaatcaa	5520
agaagtgggt ttatggaggt cctcttgtgt cccctcccc cagagggtgtg gtggctgtgg	5580
catggtgcca agccgggaga agctgagtca tgggtagtgt gaaaaggaca ttccaccgc	5640
aaaatggccc ctctggtggt ggcccccttc tgcagcgccg gtcacctca cggccccgc	5700
cttccccctgc cagcctagcg ttgacccgac cccaaaggcc aggtgtgaaa tgtaccggg	5760
aggattgggt gtctgggcgc ctcggggaac ctgcccttct cccattccg tcttcggaa	5820
accagatctc ccaccgacc ctggtctgag gttaaatata gctgtgacc tttctgtagc	5880
tgggggcctg ggctggggct ctctcccatc ctttctcccc acacacatgc acttacctgt	5940
gtcccactc ctgatttctg gaaaagagct aggaaggaca ggcaacttgg caaatcaaag	6000
ccctgggact agggggttaa aatacagctt cccctcttc caccgcccc agtctctgtc	6060
ccctttgtag gagggactta gagaaggggt gggcttgcgc tgtccagtta atttctgacc	6120
tttactctg ccctttgagt ttgatgatgc tgagtgtaca agcgttttct ccctaaagg	6180
tgcagctgag ctaggcagca gcaagcattc ctggggtggc atagtgggt ggtgaatacc	6240
atgtacaaag ctgtgcccc gactgtgggt ggagtgccc cacatggccg cttctcttgg	6300

aagggtctcg tatgactggg ggtgttgggc agccctggag cttcagttg cagccatgcc	6360
ttaagccagg ccagcctggc agggaaagctc aaggagata aaattcaacc tcttgggccc	6420
tcctgggggt aaggagatgc tgcattcgcc ctcttaatgg ggaggtggcc tagggctgct	6480
cacatattct ggaggagcct cccctcctca tgccttcttg cctcttgtct cttagatttg	6540
gtcgtatttg gcgcctggtc accagggctg cttttaactc tggtaaagt gatattgttg	6600
ccatcaatga ccccttcatt gacctcaact acatggtgag tgctacatgg tgagccccaa	6660
agctggtgtg ggaggagcca cctggctgat gggcagcccc ttcataacct cacgtattcc	6720
cccaggttta catgttccaa tatgattcca cccatggcaa attccatggc accgtcaagg	6780
ctgagaacgg gaagcttgtc atcaatggaa atcccatcac catcttcag gagtgagtgg	6840
aagacagaat ggaagaaatg tgctttgggg aggcaactag gatggtgttg ctcccttggg	6900
tatatggtaa ctttgttgc ctcaatatgg tctgtcccc atctcccc caccgccata	6960
ggcgagatcc ctccaaaac aagtggggcg atgctggcgc tgaglacgtc gtggagtcca	7020
ctggcgtctt caccaccatg gagaaggctg gggtgagtgc aggaggggccc gcgggagggg	7080
aagctgactc agccctgcaa aggcaggacc cgggttcata actgtctgt tctctgtgt	7140
aggctcattt gcagggggga gccaaaaggg tcatcatctc tgccccctct gctgatgccc	7200
ccatgttcgt catgggtgtg aaccatgaga agtatgaca cagcctcaag atcatcaggt	7260
gaggaaggca gggcccgctg agaagcggcc agcctggcac cctatggaca cgctccctg	7320
acttgcccc cgctccctct ttctttgcag caatgcctcc tgcaccacca actgcttagc	7380
accctggcc aaggteatcc atgacaactt tggatatgtg gaaggactca tggatatgaga	7440
gctggggaat gggaactgagg ctcccacatt tctcatcaa gactggctcc tccctgccgg	7500
ggctgcgtgc aaccctgggg ttgggggttc tgggactgg ctttccata atttccttc	7560
aaggtgggga gggaggtaga ggggtgatgt ggggagtacg ctgcagggcc tcaactcctt	7620
tgcagaccac agtccatgcc atcactgcca ccagaagac tgtggatggc cctccggga	7680
aactgtggcg tgatggccgc ggggtctctc agaacatcat cctgcctct actggcgtg	7740
ccaaggctgt gggcaaggtc atccctgagc tgaacgggaa gctcactggc atggccttc	7800
gtgtccccac tgccaacgtg tcagtgggtg acctgacctg ccgtctagaa aaacctgcca	7860
aatatgatga catcaagaag gtggtgaagc aggcgtcgga gggccccctc aagggcattc	7920
tgggctacac tgagcaccag gtggtctcct ctgacttcaa cagcgacacc cactcctcca	7980
cctttgacgc tggggtggc attgcctca acgaccactt tgtcaagctc atttcctggt	8040

atgtggctgg gccagagac tggctcttaa aaagtgcagg gtctggcgcc ctctggtggc	8100
tggctcagaa aaagggccct gacaactctt ttcattctct aggtatgaca acgaatttgg	8160
ctacagcaac aggggtggtgg acctcatggc ccacatggcc tccaaggagt aagaccctg	8220
gaccaccagc ccagcaaga gcacaagagg aagagagaga ccctcactgc tggggagtcc	8280
ctgccacact cagtcaccca ccacactgaa tctccctcc tcacagttgc catgtagacc	8340
ccttgaagag gggaggggccc tagggagccg caccttgtca tgtaccatca ataaagtacc	8400
ctgtgtctca ccagttactt gtcctgtctt attctagggt ctggggcaga ggggagggaa	8460
gctgggcttg tgtcaagggt agacattctt gctggggagg gacctggtat gttctcctca	8520
gactgagggt agggcctcca aacagccttg cttgcttcca gaaccatttg cttcccgctc	8580
agacgtcttg agtgctacag gaagctggca ccactacttc agagaacaag gccttttctt	8640
ctctcgctc cagtcctagg ctatctgtg ttggccaaac atggaagaag ctattctgtg	8700
ggcagcccca gggaggctga caggtggagg aagtcagggc tcgactggg ctctgacgt	8760
gactggttag tggagctcag cctggagctg agctgcagcg ggcaattcca gcttggcctc	8820
cgcagctgtg aggtcttgag cagtgctctt attgctttct gtgccctcgt gctttatctg	8880
aggacatcgt gccagcccc taaggtcttc aagcaggatt catctaggta aaccaagtac	8940
ctaaaacat gcccaaggcg gtaaggacta tataatgttt aaaaatcggt aaaaatgccc	9000
acctcgcata gttttgagga agatgaactg agatgtgtca gggtgactta tttccatcat	9060
cgctcctagg ggaacttggg taggggcaag gcgtgtagct gggacctagg tccagacccc	9120
tggctctgcc actgaacggc tcagttgctt tgggcagtta ctcccgggcc tcactttgca	9180
cgtgtgctta cctagtggag acaaaagtac atacctcggg agagcgcgca cgcctgtaac	9240
cccagcactt tgggaggcca aggtgggtgt atcacctgag gtcaggagtt tgagaccagc	9300
ctggccaaca tgggtgaaact ccgtctctac taaaattaca aaaatcagcc aggcctcatg	9360
gcacatgcct atagtcccag ctacaggcat gctgaagcag gagaatcgct tgcaccccg	9420
aggcagaggc tgcagtgagc tgagaccaca ccactgcact ccagcctagg caacagagta	9480
tgagactcca tctcaaaaaa aaaaaaagta cctacctcag agttcaaact agtgaatatt	9540
aggaagtgtc tgagacagtg acaccaaaag gcacaataaa tactcgccag tttcattatt	9600
attaaagaat ccatttgaat gtcagctcaa cacagcctcc tataccgagg cattgtgaac	9660
cgcattctcc cagcttctcc aggcctttcc aagaatcagg gacactgtag cctgttggtc	9720
tcagtgtatg acagacacgg aggaagcaca tctttagctg atacttaaac agagaccctg	9780
agcgcacata caccgcgca cacatgcatg gagcttcacc ttctctgtca ttctgcagt	9840
accaggagag caagagctcc cactccctt caaaacactg tgcccatccc gggcactaag	9900

gcctcttttaa agcacggcac ctccacgagg gagggccaca gccacataca ctccacctgg	9960
caggtggaca gcgtgagcac gtggaccata gcagggacaa ggtgccccgg ccagcccaa	10020
cgcctctgc cgctgacagg gacagaagcc ctctccagct gcgtgtgctg cagaggccat	10080
gcgtagcctc cagctgcatt ctattccact ccagtgcctg ggccagttag caccagtgtg	10140
gaagacagtg agctggctcc ggacaacagg gatggaggaa aggtcccaca ttcacattcc	10200
tgatactgg acaaggtgag gggccgcaat cgctctggca gcattttaaa gatggggaag	10260
tagcagacac ccacgcgtga aggcaggaga gcccactg tggaggaaat ggcccagaa	10320
tggtagggcc aagcctagct ccagacaccc cagagccctg gagaagccaa gactgaggga	10380
gaaagcctga gggaggagcg cccagtcctc cagggaccgg cctggtgcag agctgcagct	10440
gatgttccc tctgtgcagc cccaccctct gcctcgctga gctccctgct gcgaggcct	10500
cgggtgcaag ggggaggcag gtctctatct catggagctg tcagatgaga catcgcgatc	10560
ggagtcctca gcctcgcttg gcggcggcgg cgggtcgcta agcgggaccg cagtgaagc	10620
aggagacttt ctagaaaaa acaccagttg tcaaccttg ggacggcagg aatcctgaag	10680
acggacggca ctctctcc tgctgcctca cctctggca gcccgtgaga agtaccgaa	10740
gcgagggcgg ggcgcggga tggcgaggga gcggcaggga ctgaactctc tccaaacca	10800
ccctgacagg gaaatggcc ccgctgtgt cttgggaact cagaggctga ggtcaggcat	10860
gatggctcac gcctgtaatt ccagcacttt gggaggcaga ggcggtgga tcacgacgtc	10920
aggagttcaa ggcaagcctg gccaatggtg tgaacacca tctctactaa aaatgcaaaa	10980
attagccagg tgtggtggcg ggcgccttg aggctgaggc aagataatca ctgaaacctg	11040
ggaggcggag gtgtagtga gccaaaaaa aaaaaaaag aaatagctga agtcacagta	11100
ggagagaagc tgctgagcct ccagcacct gactctagg ccttgcttt atgtctatct	11160
gcagtatctt tgtatcttt aaaaattcac tttctgttg cggtgtaact tacacagggt	11220
caaatgcaca aatcatggcc ctgactttag ataaaaatct gccccacaa ccttctgtt	11280
ccttgccagt ttttaactg cctctaacca gggaaccac cagagctggt ggccttgga	11340
ggtttcagcc ctccgctcat gaatggacat agctcatcca actgccaagg gagagagctg	11400
tgggtctggg ccagccccac caggtaactc ccaaaggga gccccacagc aagatgtgac	11460
ccagtcattg cctgagggtc tctggggctg tgttccaacc tttctcccg ctgtgtcccc	11520
ctggaaggcc ccatgcccag gggaggcgcc taccggtctc cagactgggt gatgagccgg	11580
cggcaggctt ccatctgcac gtccaggccg cgcttcatgc tgcacatctc catgtactcg	11640
tgcaggtgcc ggttcatgtc gttcttggcc gtggccaact ccagctgtgg tggggcaggc	11700
agggccatcg gggtaaacag gtggcggtca cagcgctct gttgccccg ccaggaggcc	11760

aacacgcaa gagcagtggc tgggcccggg gccaggcag ccattactga aggctcagat	11820
tttaaaataa accagaccaa ggcagaaggc agcgaataca ccattgtttc ctctgctttt	11880
cctggaaatc ttgacctatg gaggaatgca aaggctactgc tgcaggctgt gttcttttaa	11940
gactgataag agtggtaggg aaagtcagt tgggcccggg ggagcagatc cacaggctcc	12000
cagggccaca caggttccta tggggcccgg ctacccctc atcccctgta ctgcgtggt	12060
cctgggctgt cccctactg cactgtgagc tcccgtagg gtttagcgaca ggttttctc	12120
atcctggcag acccagacct cgaagggaat taactctctc tgtgacctg gccatacaag	12180
cctggctaac tgcgaaagg ggccaatcc tcttaccatt ctctccttat cactttattt	12240
tctcacctaa aagccatttt cagacactaa aaggaagact cccatattgg aggtgccatt	12300
gattctgaga cctgtcctga ttccagaaat gtcacacgtg gaaaacacgg ggctcagaat	12360
caaagaaatc cggctctcct ccccgggccc attctgcagc ccacactctg aggcaggcgg	12420
agaggcaggg cccaggcagc acaactcag tgccaggac agtcgactgc ctctctttgt	12480
tttgggtaca aagggcaggg gccgcaagg gaagcaaat gaaacttct ggagatgacg	12540
ttcgtgactg cctgagtga gagaacgggg aggataccag ctgccataat gtggattgga	12600
cagagagggt gggaacttag agaaagaagg aaacaggtta cagggccaga cagcagagat	12660
ggtcaggaca gtgacatttg gggtagatg gcggctggct gggtagaag ccctagtggg	12720
gacagcttc ctggcaggct tggatccaca gctggggccc cacagtagtg ggaggagcag	12780
gacacggtga ctgttcaacc cagggggaaa atacagtgc accagccctt gttcagatca	12840
gtgcctaca cccaatttc tccaccttg aagaaccgga gtgcaggag ctgggttctg	12900
gaaccatcac tcaggagagc tgggaagaag gaaatgaacc aaatatatt tactgaaata	12960
caaggctagt actgccaca gtagtctctg catatgtcta tgtccccact taattcaca	13020
atttacgca ggaacaaggt caggagggtga acgtgcatgt caggagatgt ccttgacaag	13080
ccacctgaa aactgcaggc tcgtgtgtg aacagacttg ggggtgggac gggaggagac	13140
agagagaggt gaggagcgtt ggggaagaac tgctcccg ccaccactct gggccttggc	13200
tttaccttg tccatcttag tgacctgag agatcaaagg tcacaaagat aaaggcacct	13260
tgggtactcc tgaggctagc aagtctcct aacggttccc ctaatgcaga agtcccaaa	13320
ccccaggcca cggaccagta cctgtctgtg gcctgttagg aaccaggcca aatagcagga	13380
ggtgagcagt gggcaagtga gtgaagcttc atctgtattt acagctgctc cccatcactg	13440
gcattacagc ctgagctcca cctcctgtca gatccgcggc ggcatagat tctcacagga	13500

tcctgaaccc taccttacgt gaactgtgca tgcgagggat ctaggctgtg tgctccttat	13560
gagaatctaa tgcttgatga tctgtcacta tctcccatca cccccagatg ggaccgtcta	13620
gttgcggaaa acaagttcat ggctcccact gattctacat tatggtgaga tgtataatta	13680
tttcattata tattataatg taataataat agaaataaag tgcacaggcc ggggtgcggtg	13740
gctcatgcct gtaatccag cactttggga ggccgagggtg gtggatcacc tgaggtcggg	13800
agttcgagac cagcttgacc aacatggaga aaccccatct ctactaaaaa tacaaaatta	13860
gccaggcgtg gtggcacatg cccgtaatcc cagctacttg ggaggctgag gcaggagaat	13920
tgcttgagcc cgggaggcgg aggcctgcagt gagccaagat tgcgccattg cactccagcc	13980
tgggcaacaa gagcaaaact ctgtctctaa aaataataat aataatTTTT aaaaatagta	14040
ataaagtaca caataaatgt aatgtttgaa tcactcctaaa atcatttccc ccgggtcagt	14100
ggaaaaactg tcttcacaa accagtcctt ggtacaaaaa aggttgggga ctgctgccta	14160
aagtcattgc actgggagag aggaggagtc agatgatttg aaaatacttt ttttttttt	14220
tttctgaga tggaatcttg ctctgtcacc caggctgttg tgcagtggca caatctcagc	14280
tcactgcagc ctccacctcc cagattccag caattctcct gcctcagcct cccaagtagc	14340
tgggattaca ggtgcccacc acaatgcctg gctacttttt gtatttttag tagagacggg	14400
gtttcacatg ttggtggcca ggctggcttc acactcctga cctcaagtga tctgcccgcg	14460
ttggcctccc aaagtgtctg gattacaggc gagagccacc acgcccagcc aaaaatagct	14520
taaaaggcat gattccattt aagatattat ttctgttctt tcagatcatg ttttactg	14580
ggcacttgag gtccaggga gactaggag agatggcagc ccgagatctc ttctcacgaa	14640
gatggccgca caaaggcccc tagttggttg aaggcacagt gaaaacacaa accttaaatc	14700
caccatccaa aagagggaga caagaatcag gtcaggcagg aggaggaaac agtgagccgc	14760
ggtagacggg aggggtggga gtactggttt ccttggcat gtcttttagct agcagcacia	14820
atccaggaaa cctcatctac ctggccaaga aaaccagggt tgatgcagga gcaggacagt	14880
ttgatgtgtg ttgggagaat gagtatatag gaaaataatt tctcgacttc cttctcaaac	14940
acctccaatg tgacagcaat aagaataaag tcaaggccgg gcgcggtggc tcacgcctgt	15000
aatcccagca ctttgaagg ccgaggcggg tggatcacga ggtcaggaga tcgagaccat	15060
cctggctaac atggtgaac cccgtctcta ctaaaaata caaaaaatta gccgggcgta	15120
gtggcaggcg cctgtagtcc cagctactct ggaggctgag gcaggagaat ggcgtgaacc	15180
tgggaggtgg agcttgcatg gagctgagat cgcgccactg cactccagcc tgggcgacac	15240
agcgagacta cgtctcaaaa aaataaataa ataataaat aaatgaaaag aataaagtca	15300
ctctctcctt atccactggt ttactttcc acggcttcag gtaccaagg tgaacctga	15360

tctgaaaata ttaaatgaaa tattccagaa ataaacaatt cctaagtttt aaattgcacg	15420
ctgttctgag tggcctgata aaatctcact ctctctgccg ggcgcggtgg ctcacgccta	15480
taatcccagc actttgggag gccgaggtgg gcggatcatg aggtcaggag atcgagatca	15540
tcctggctaa cacggtgaaa ccccatctct actaaaaaat acaaaagaaa ttagccgggc	15600
gtgatggcgg gcgcctgtag tcccagctac tcaggaggct gaggcaggag aatggcgtga	15660
accagaggagg tggagcttgc agtgagccga gttcaagcca ctgcactcca gcctgggtga	15720
cacagcgaga ctccgtctcg gaaaaaaaa aaaaaacctc actctatcca gaccaggaca	15780
tgaacctcc ctctgtccca aggatccaca ctgtctatac tgctgacca ttagttactt	15840
aggagccatc tcagtatga gattgactgt cgtggatgg cagtgcccg gttcaaggaa	15900
cccttatctt tacttcataa tggcccaaaa gcacaagagt agtgatgctg gcaatttga	15960
tatgtgaaag agaagccaaa atatgctccc tttaatgaa aaggtaaaag ttttcaactt	16020
aataagaaaa agaatccgat gctgaggttg gtaagctcta cagtaagaat gaatcttcta	16080
tccgtgaaat tgtaaagaag gaaaaagaaa tttgtgccag ttttctgtc acacctgaa	16140
ctgcaaaagg taccgccaca gcgcacgcta agtgcttagt taagatagaa aaggcatiga	16200
atgtgtgggt gggagacatg aacagaaaca cgtccagtt gacagcaact ggaccaagc	16260
tatcaggagg tcagtctat ctttggtttc aggcatccac tgtgggtctt ggaacacatg	16320
ccctgaagat aaggggggac tacgtatgaa tctccatgt gggcctgaag agaggctgag	16380
acgcagagaa ggctccatgc ctctccaaa gtcagagcgt cgggaaagga ccacctacc	16440
aggccaggag agctcacggg ggtcccgctc atccctctc ccacagcaac agcaacctcc	16500
attctgtgtt tccacaaatc cttcttctt tgtttgtatt tttgagacgg agtcttgctc	16560
tgtcgcccg gctggagtgc agtgggtgca tctcggtca ctgcaagctc cgcctcccg	16620
gttcacgcca ttctctgcc tcagcctccc gagtagctgg gactacaggt gcctgccacc	16680
acggccagct aattttttt ttttgtattt ttcgtagaga cggggtttca ccgtgttagc	16740
caggatggtc tcgattctct gacctcgtga tcttcccgc tcggcctccc aaagtcttg	16800
gattacaggc gtgagccacc gcgccggcc tccacaaatc cttcttgag ctctaaagc	16860
ctagagtttg agctaaagca ccagctccac ttgcctctag gtacctctag aaggtacttc	16920
aaggcttggg aaaaagcatg gagatttctc tgttatctca aaggctgagt ggaagaaaca	16980
ggaaagtccc agtccatctc tgagtccat ggcaatgata gtgttaataa taataaaaca	17040
gtccccacta tgctacactc gctatgagcc agacactgcc ctaagctcta tatatgcatt	17100
atttcatctg atcttctaaa caacctata ggtgaggtat tattaaattc ctcatcttac	17160
aaataaaaag gcccggcaca gtggctcatg cctgtaatcc caacacttg tggggccgag	17220

ttaggaggac tgcttgaatc cagaagtttg agggcagcct ggaaaacata gtgacaccct	17280
gttctctgta aaaaaaatca aaaaattagc cgggcgtggt agcacacgcc tgtggtccca	17340
gctacttggg aggctgaggt aggaggattg cttggatcca gaagatcacg gctgcagtgg	17400
gcctggagtg acagagcaag accctgtctc aaaaaagaa agaaagaaag aaagaaagaa	17460
aaagaacag gtgtagggtt ggacacatgg ctcatgtctc taattctagt actttgggag	17520
gccgaggcag gaggatcact taagcccaag agtttaagac ctgcctggga aacacaggga	17580
gacaccatct ctacaaaaa atacaataa aaattagctg ggtatggtgg cacacacctg	17640
tgggccagc tactcagaag gctggggtgg gaggatcact ccagcctagg caacagagtg	17700
aaacagagtg agacctgtc tcaaaaaaga aaggacagaa agacaaggga gagaggagg	17760
gagggataga gggggagggg aagagagaga gagagagaaa ggaaggaagg aaggaaggaa	17820
agaaaggaag ggaaggaag ggaagggagg ggctcagagg gaacaagtaa cttattacaa	17880
gttcagggtg gtgttaagta acagagctga gatttgaacc gatatctgtc tgtttccaga	17940
gcctgaggtc ttcagagccc accaggagga ctatagctga gagatcactc ctctcccaa	18000
gaccttggtc tctagggttc tgcagcaggg atgaggccac tgcgcctgca gcccactca	18060
aaacctctg ggacaccacg cccctggctt ccgctcctgc ccttac	18106
<210> 18	
<211> 16312	
<212> DNA	
<213> Mus sp.	
<400> 18	
gtgagtggtc ctgttaggtt ccggtgcagt aaacttcac tgccctcccta agtgctctga	60
gactcctgtc tgaactttct gccttgctag cctcactaca gctatggaac cagagtgact	120
aagaccagat cttaattcct ggtccctttt ttaaaggcca ttcatttcct tttttgttaa	180
catgcattgg tattttgcct gtgtatatgt ttgcatgaag gtgccaagtc cccttggaat	240
tggagttaga ctgttgtggg ccctgagaac tgaacacagg tcttttgtaa gagcagccag	300
tgctcctaac cactatgcca ccttcccagc cccagtgaag gccttgagat agggctctgc	360
tatgtagacc cagctgcctc agactaacag agattccctt gcctctgtct cctggtagta	420
gaggtagtgc gcacatgttt gcactggact gcttcttttag agtagtttca tgtttaaagc	480
aaaagtcagt aagtagagct cccatatgct ctgtcccaca cccgcatggc ctccaccca	540
tgaccttctt gtcttggtat acgtgggtct ggatgaagga acgtgattga cagatcagca	600
tccaaagtcc actcactgga cacagtggca catacttacc acttgtgtca tactgggcat	660

tttcaggact ttgaggtcac cttaccccc acccctgtgg tggcatcagt tagttggaat	720
cacagtttca gattggtttc tttagttttc tctcttttca tggctatttc tttttaagca	780
gggagataat gtttccagcc ctgtctttta ctttctgtcc tccaggctat aatagatgaa	840
tttgagcaga agcttcgtgc ctgtcacacc agaggcatgg atggaataga ggagtttgaa	900
actggccagg gaggcagcca gcgagccctg tctgccaaga aacctccgc tggatgaga	960
gggatgcaag cccacatggg gctcctatga gtgggtccca ccttttttt gttcctttgt	1020
ggttttttt tggttttgaa atgggtccct tgcctttctt ttctagtctc aaggctacag	1080
cctctgactt ctgtagactc agacaatgac ttgttcacac ctaagccgcg acgtacaaa	1140
cctggtcgtc caaaaactca gcagcgcaag aagtcccaga ggaaagccaa agttgtcttc	1200
ttgagtgcgc agtccagtga ggacggtag atacaccctc ccaacagatc agctgcctgt	1260
gtggggtctt ccagtggggc ttgcctagct gtggattcat tcttcctggc catgtcttcc	1320
tcactatgca gctcggctga ctttcacagt gtgccgtgtg cgtgtgtgtt tgtgtgtgtg	1380
tttgtgtgtg tgtgtgcaca cgtgtgcgcg tgcgcacgcg cgcatgtgtt tactttgaga	1440
tccattttgc cccatcttta tgactcagct ttgtttccta taattctttg aatagaactt	1500
tcagcagaga tgacagaaga agagacacc aagagaacca ccccatccg cagagcatct	1560
gggcgaagac acaggtccta ggagctgcct ttgcctactc ccgtccctgc cgtgcagggt	1620
atcttactgg atgatcttgg attcagtgcc cccccccct tgtagaagtg cttgttgtct	1680
gtggttttta catggaaat gcctttgcat tgaggctctt ctcttagagc cctgctgcca	1740
ctttcccttt ctccgagcct gtgccaattg ctgttttcta atgaataaat gtttttatag	1800
acttttaagc atcctgtcct ggctgggtctg tcccatcctt agccctgagc tgtgtctgga	1860
tcttcatggc cctgtgtgtg cctcatcctt gcacaactga aactctgctg tgcagaatta	1920
tgcggtttct aggttcacgt gaacagtctt cagcagctct ggggtactttg caccacagaa	1980
aacattcgtt aatacctaga gacttatttc tgtttgcctt gtggggcagg gggtgcagt	2040
gaaagcetta agatgaaggc caccttcatg cacattgtgc ccataaac acgatgtcag	2100
gagtacacaa cttagaaaat gctgtcttag acctagggt tgtttcctct gtgcgcccc	2160
ccccatccc cgaagactgg caaatgactc agcctggcta atggcagaga ttatgaacc	2220
cactcagttc tataggtagc aagggtccatc gcccagagc cagtcttca ctacactaag	2280
gggtctgcag gaggtccagc tgtaagccat gctgtgtgcc caccatcccc aagcccttaa	2340
tctgtctgagt cacttggagc aggaggcaca actgtgttt ctgtgcat gacaacagca	2400

cagtgtctgct gaagtgtctcc ctacctcttg gtgaggggac agtaatgata gatgacaggc	2460
gtgaggcaaa agggaaggtg cacctctgaa aactaggtt agggctgggc tatggaaggg	2520
ctctttgggt aggaaaaggg ttaacgctgt gcatagagcc tcggtaggaa ggcagaaaag	2580
ggactctgaa tcigccatgc ctctctccc caccactgcg ctggggccac tcccaccctt	2640
ccccacctgt tcatctggcc tgcctcccgc tcatgggaag cctgggcacg ggccacacag	2700
ctgtcagtct aacacagcct tggctctggg gccccaaagga ggtagggcaa tctggaaggg	2760
caaggagcca agactagatt tggggtgcag ccacgctgc tccctgcctt ttaagcaaag	2820
gttatcacca ggccagctaa acttagcaat aggcctctga gctagatgag cagggggctg	2880
gtttcatgtt gcactggcct agcaaagagg cgccagagtc cccctgcctt gcacctgcta	2940
cagtgtctcc aagcccttcc acccttcca gcctccctac ccttcggagt gaagaatccc	3000
ggtctcgaaa agacaaaata aaaccaaga atgccttttc tcccttcac ttgtggcaag	3060
aggctagggg ctccctgtc ctggctcagg ggtataatat ttcctctctt gtgttctccc	3120
ctcactgac tcacctgtg tccacgaggg cagccaaaaa aggagagtcc ctggctccag	3180
ggccagttag cagctccctg gatggggagg gacttccgc ctcacgtccc aactctccac	3240
cctggggcta cagtgggtga aaggggcagt gtctcctagc ctgggcgggg cagctctcag	3300
gttccgagga gggatactat aggaggcctc tgagcctcct ccaattcaac ccttaagagg	3360
gatgtctccc ttaccccggt gtcccagctt aggttcatca ggtaaactca ggagagtgtt	3420
tgtaagtctc aattatcagg ttccaggct gcagggcac ctgacctatg gcgtagcaat	3480
ctcttttca agcgtcttcc tggccacccc ctgcccctca cagcgcggtc cacaagccag	3540
gcctctggcc aaagacagaa gccaggaggg gggggggggc agataggga atggggctcc	3600
tgagggtccc cgcgggagga aaagtgtccc caccctggca ttttcttcca ctctttttt	3660
tttcgggggg gggggttct gtgtcactac cgaagaaca cgaggagaag atcctcaact	3720
tttccgcagc cttttcaata atggggagag gttcgatgat gcagtgccag ggagaccac	3780
acttctccat ttccctgtt ctccatttt actcggaag cagcattcag gtctctgggt	3840
cctggatgtc ctgggtgcac actccaagga ctctctgtcc ttaagttcat agtctgtatt	3900
ccctgagtcc tatcctggga accatcacc ggtcacctcc tgagcggggc aatctcagct	3960
cccccccc tatcagttcg gagccacac gcttgggtgcg tgcacatttc aaaaatgagg	4020
cgggtccaaa gagaggagg aggggaaatg agagaggccc agctactcgc ggctttacgg	4080
gtgcacgtag ctcaggcctc tgcgcccttg agctaggact ggataagcag ggcgggaggc	4140

ggggcgcgcg tcatcagctc cccccacca tccgggttcc tataaatacg gactgcagcc	4200
ctccctgggt ctctctgctc ctccctgttc cagagacggc cgcatcttct tgtgcagtgc	4260
caggtgaaaa tcgcggagtg ggccgcagga ggccggggac agtcggaaac tgggaagggg	4320
agtgggcact gtacgggtct agggatgctg gtgcgaagtg tgcaagccgg acccaggctc	4380
cgcatcgcag ggggggatg atggaggacg tgatggggcg cacggcggga atggaggcgg	4440
ggtgggggag gggactgcct ggtgtccttc gggccacgct aatctcattt tcttctcctg	4500
cagcctcgtc ccgtagacaa aatggtgaag gtcggtgtga acgggtgagt tccagggcgg	4560
ggccctgctc cgttgcttac gcaggtcttg ctgaccggg ggctctgcag tactgtgggg	4620
aggtggatga ggtggccgaa gcgccaagg agacctcaag gtcagcgtc ggacctggcg	4680
atggctcgca ctgcccgc ccagctgctg cacctctggt aactccgct ttgcggggat	4740
gagcggcccc gactcttaag tattaggaac aacccacgc gcccgttcag acccatccc	4800
taatccccag tccggggctt ctttctttac ttctgcgccc tgaggagtca cgtgccagga	4860
gggaagcccc tccccatct ccccttctt cggggtcggg ggcatggcgt tgtgaggtgc	4920
atacctttgc gcatcatctc ccaggcttgg gcttcttta gggtaactgg ccgccccat	4980
gttgcaaacg ggaaggaaat gaatgaaccg ccgttatgaa atcttgctta ggccttcctt	5040
cttcttagct tgtgactaac ctcatctctc tcggctgggt ggagtgtcct ttatctgta	5100
ggccagggtg tgcaaggctt ccgtgctctc gagagagttc tacctcaca tctgtctcac	5160
cttattagcc ttaaaagccc ttgagcctta ttgtcctcgg gcataatgcg tattctagat	5220
tattctctga aaatcaaagc ggacttacag aggtccgctt gacctccaa cccagaggt	5280
agttatggcg tagtgcagag ccgtgggatg gggagctgag tcatgggtgt tctgaaaaga	5340
aattttccac cacaaaatgg ctctgtagt agcagccct tccatccct gcacttcca	5400
tcacagctc gactgacct aggcctata ggccaggatg taaaggtcat taagaggatt	5460
gggtgtccct gcgcctcaga atcctgcctt tctcccgtt ccatctcca gaaaccagat	5520
ctctccactc cgcctgac tgaggttaaa tttagccgtg tgacctttct g gatctgggg	5580
tctgagcggg ctctccacc tgctccctt acacacatct gttgtccgg ctctcatitt	5640
tgcccagaaa gaacagggtt ttgcgaac agccctggga ttagggttgg aaacccccca	5700
catgttttct cagtctttcc ccttagttcg agggacttgg aggacacagg tgggcccgc	5760
ctgtgtgtct cagctgacc tttagccttg cctttgagc ttgtgatga atgagttcac	5820
aggtctgccc tgtccagggg gtgtagcctg aagtcagcc atgttggaac aaacttcca	5880
gggcatgagt gatgggggtg atgtgccaag ctgtgacca gactggggca aactgccact	5940
tcttaagaga cttagaatga cttggaggag gtttgctggg caagcaatca cctcttgac	6000

aggaaagaaa cctccacttt ataaccgtgc tataaaagcc ctgccaggcc tcggctgctc	6060
aaagaatata aaattagatc tctttggact ttctaggggtg ggaacagtct atatttgggt	6120
tgtacatcca agcattcaac tagctttatt aaagggaatt ctgaacaaaa catgaacttc	6180
ctgatgcata catttctaata gtactgtgtc tgcataaagg cttgtacctt tgtttggta	6240
cgatgcatagc tgatggctgc aggtttctcca cacctatggt gcaacagtat tccactctga	6300
agaacatgag atagcctggg gctcactaca gacctatgag gagttctgat ctcagctccc	6360
ctgtttcttg tctttcagat ttggccgtat tgggcgctg gtcaccagggt ctgccatttg	6420
cagtggcaaa gtggagattg ttgccatcaa cgaccttctc attgacctca actacatggt	6480
ctacatgttc cagtatgact ccactcacgg caaattcaac ggacacagtca aggccgagaa	6540
tgggaagctt gtcatcaacg ggaagcccat caccatcttc caggagcgag accccactaa	6600
catcaaatgg ggtgaggccg gtgctgagta tgcgtggag tctactggtg tcttcaccac	6660
catggagaag gccggggtaa gtggccgaa agctgaaggt gacgggcacc cttgatatgg	6720
tgaacctga aaaccaagaa ctgagttctga aatcaacttc tttcccttaa acaggccac	6780
ttgaagggtg gagccaaaaa ggtcatcatc tccgccctt ctgccgatgc ccccatgttt	6840
gtgatgggtg tgaaccagca gaaatgatc aactcactca agattgtcag caatgcatcc	6900
tgcaccacca actgcttagc ccccttgcc aaggtcatcc atgacaactt tggcattgtg	6960
gaagggtcga tggatgtag gcagtgaggga gacagctcat gcatttctta tcttaccctg	7020
ccatgagtgg acccttcttt gtaggtgtcc cttttggta gaggggtgcc gtgcaggacc	7080
tactcattg ccccggtgtt ttctagacca cagtccatgc catcactgcc acccagaaga	7140
ctgtggatgg cccctctgga aagctgtggc gtgatggccg tggggctgcc cagaacatca	7200
tccctgcac cactggtgtc gccaaaggctg tgggcaaggt catcccagag ctgaacggga	7260
agctcactgg catggccttc cgtgttctta ccccaatgt gtccgtctg gatctgacgt	7320
gccgcctgga gaaacctgta tgtatgggga gagctgggct tgtctctggt gtgacagtga	7380
cttgggacaa ggatagtcat tttgggggtt gtctctatca ggccaagtat gatgacatca	7440
agaaggtggt gaagcaggca tctgagggcc cactgaagggt catcttgggc tacactgagg	7500
accaggttgt ctctgcgac ttcaacagca actccactc ttccacctc gatgccgggg	7560
ctggcattgc tctcaatgac aactttgtca agctcatttc ctggtatgtg gggatgggaa	7620
acctgacttt taagagcaac tgggggtttg gtgccctctg gtggctagct cagaaaagaa	7680
acccaaacta acagtgtgcc caatttgttc taggtatgac aatgaatacg gctacagcaa	7740
cagggtggtg gacctatgg cctacatggc ctccaaggag taagaaaccc tggaccacc	7800
acccagcaa ggacactgag caagagaggc cctatccaa ctcggccccc aacactgagc	7860

atctccctca caatttccat cccagacccc cataataaca ggaggggcct agggagccct	7920
ccctactctc ttgaatacca tcaataaagt tcgctgcacc cacctctcct tggttttaat	7980
gcggtgttgg gggagggagg gtgtgtatgg ggaggtgaag caggctcaat caacttttgt	8040
gcatgtatct ttattggctc tgacatccag ctgggtgccg gaagtcagg gctacattct	8100
atccaagtc cagcccagg cgcctgtcct gaacagctga tatgtctggc aatgcggagg	8160
gtggggcttg acaatgggac aggaaatgca gtctgagact cacttggttg acagcactga	8220
cttccagagc tcaactgtct tgagaaagat cgggtgtcctt ttatctgcaa acaaaggtta	8280
attgtacca gccctctgc agacagacgt aaaattagat aaaattggaa gggcgcccag	8340
tacagtaaaa cctagtggta tccatgaaga ggggctgtag cactggctga tgctttggtt	8400
cgtgcgtcac acagtgactc agtgcctggg aagtattcca caccgtcatt gcttttaaag	8460
gttccattcc aatagttcaa catagcgtct cctgggaggt cgtgactgac tccctaggtc	8520
agggacatct taaccttttg gttttgtcag tgacagccat ggcggcatat cctcagctgc	8580
tcactggaga aacaggcccc gagcaccagt aacctgcagg ccacacccat gtgcgggcag	8640
cagcttcacc tctcagccg cgacctgggc ccagcacagg acttcccaaa ggcgagcctg	8700
cacggccatg ttactgcaa gaggcaagtg aacagctggg agcatgtgga cagagcacc	8760
cacacctcc aaagccttac tagggagcaa agccctatcc caaggtacat gatgaatctg	8820
cttcacacga gtgcctaggc aagtiagcac cagtacaaaa tccaagatac ccaaggaata	8880
caggcaggtc ccgatcac attccttaca aagtggaaaa gggaggtca caagtgtcct	8940
ggcagcattc cgaagatgga gaagtagcag ataccacat gtgaggaagg caggaaggag	9000
tgtcccaatg tcatggcgct tgtacaggga ggggtgggct gagcctagct ccagacaccc	9060
cagaacctgg ggaggacac cccagggccca ggcttcatgc agtggcacag aagaccaca	9120
cccctgcaga cccacacccc tgccccagtg agctctgggg tgccagacac tggtttctat	9180
ctcatggagc tatcagatga gacgtctga tcggagtct cagtctcact cggcgggtggc	9240
ggcgggtcgc taagcgggac cgcagtgaac gcaggagact ttctagaaaa aaacaccagc	9300
tgtcaccttg gaccaagcag gaaggctgca gacttctaag cttgctttct catggccttc	9360
tgaggtacag ggggtgccag gcaagggtg gggagggaag aagacattc cttttcaac	9420
ttgatgggga gcagtgcctc tcttatcttg gggggctctg agaggctcag agatcagaaa	9480
agtgcgaagt cacagcacgg agatgcccc tgagccacgc tggggcctta gcttctgtc	9540
cagcagtaat ctggtatggg ggtgggggct gctggacccc aaaactcctg ggtgcaagca	9600

atcctcttgc ctccaaagtg gctggggtgg agccacagat ctcttgtaag catgcacact	9660
atactcagct tgatgtaatt tgggagattg ggttgggttt tttgttgtct ttgagacagg	9720
gcctgtacgc attcctgtct ggccctggaac tctatacaga ggtgaggctg gcctgttact	9780
tggagtacct gcctgttgtct tctatgcatt tgaatgacag atatgcgcta gcacacctga	9840
cttactgtag tttcttggat tttgtttgtt gttgctatctt ggtttgtttg ggacaaattt	9900
ttgctgctgt cgtccaggac agcctaaaac tcatgacaat ccttctgcct ctgccttcca	9960
agactgaaac tgcagacatg taccactctg cctggtttgc agtaaatttt aaaaaaatt	10020
tttaatgtgt gtgtgtctgt gtatgtgtat ctgtctgtgt gtctgtgtgt gtatgtgtgt	10080
gtctgtgtgt gtatctgtct gtgtgtgttt atctgtgtgc ctgtatctgt ctgtgtgtat	10140
gtctgtgtgt atctgtgtgt ctgtgtctgt ctgtgtgtgt atctgtgtgt gtgtgtttgt	10200
atctgtgtgt gtgtctgtgt gtctgactgt ctgcctgtct gtctctgtgt gtgtgtgtgt	10260
gtgtacacag ggccccacac aggctaggca agtctctac cactaagcta tactccgaaa	10320
ccccagattt actcttacac aaggctcaaac acataaacc gtgatttgga taagatcctg	10380
tccccgtatc tttggctctt gcctgggtta ggttctaaga cactcagagt tgggacactg	10440
gtcagtttca tctgatggcc aaggcaggga gaggaagctt cattctgggc cacacctaca	10500
ggcgactccc aaggacagct ggacataaag atgtgacctg gtcctgccc acagaccatc	10560
ggggcctatt ccagccttcc tccctgggag agagaccag cccgctcggg gcaggccctt	10620
accggtcccc ggactgtgtg atgagccggc ggcagggtctc catctgcacg tccaggcccc	10680
gcttcatgct gcacatctcc atgtactcat gcagggtgctg gttcatgtcg ttcttggtg	10740
tggccagctc cagctgcagt aagggtgggt caggaggagc cacggttagc aaaccaggc	10800
aaggccctgg ggccaagca gctcaggaga caagtaagct acaactgcct gtttcccca	10860
ggatatggag gagggtgtgg gcaactggcg atctgacttc aggcctcac tgttgctagg	10920
caagggtca ggacagtgt ccccgctcgg cctgtcctg ccaatccact cctactgcca	10980
cccaagaatc aatcagctga tgtctgagtt ctctgtaaga acagcaagag cacacaacac	11040
agtaaggatg gtgcggtgct gtgcggtgct gggagcacgc atccctccag cgcttactcg	11100
gctcaggcag cagggtcccg aaggcaaacc taaatataaa acccggtctc aaaaagggt	11160
aaggaggcca caccaggcgg ctcatggat aaggctctg cgcatctggc aatgtgaatt	11220
tgttctctaa cctccacaca tgcaccgtgg catgtggaga gagagagaca cacacacact	11280
aataacaata ataataaaat ttagtatctt ggggctggt gagatggctc agtgggtaag	11340
agcaccgcac tgccttcca aaggctccaga gttcaaatcc cagcaaccac atggtggctc	11400
acaaccatcc gtaacagat ctgacgcct cttctggtgt gtctgaagac agctacagt	11460

tacttacata taataaataa ataaatctta aaaaaaaaaa aaaaaattta gtattttgtt	11520
tgttttgttt tcgaacagg gtctctccat gtagccctgg ctgtcctgga actcactcta	11580
tagaccaggc tggcctggaa ctacagaaatg cgctgcctc tgcctcccga gtgttgggat	11640
taaaggtgtg caccaccact cccggcaagg cagcagttct taacctgtga gtctcaatcc	11700
ccttgggggt tgactatcag acactctgca tgtctgtatt tatgttataa ttcgtaacag	11760
tagcaaaatt agttatgaag cagcaatgaa atagtttatg cctggggcca ccacagcgtg	11820
aagcactgtg ttaagggtca cagcgtctgg aaggctgagc accactgctt taaggacaga	11880
cagggccact catctgtcca tctttgccac tgagcctccc ctcaactgaa gctaattcac	11940
tgtgaaatat tccaaagcct cgctgcaatg gagggaaacct ggctgcttct gtttatagca	12000
tttcattcca catcatcttc tcccttataat cctgctcgcc cacataagag agagaccttc	12060
ctcctgagtg attacagggt aaatcccaa caaaagaaat gggggggggg agtctctgta	12120
tcaagaaaag tctcccttcc agagctgact ctgcagccag gtttgaaggc agcaaaaaa	12180
gtaggctgag gcggaaaagt atctaggcca gagacaaact gcggcctctc tttgtttgtg	12240
atacaaaaggc caggggccag ataggagaa tcagactgcg gtggggagat ttgcttggt	12300
gcctgagggc aaagacaagg aggccgagtgc gccttgaagc atgtggatca gctgtgggaa	12360
gtcaccgaa acaaatgtga actggaccgg gtgggagtgg gcccaggag cagacaaagg	12420
agacagccag gacactggca ttttgggtca tctgtgaac agctgggtgg taagccacag	12480
aggagatggc ttgaaaggca tgcttgttcc cacagcagac tgcaccattt gtgagaaaa	12540
catgacactg aggatcagag ctacagtgc accccccaca gacaaatgca cacagcattc	12600
caggtcagag atcagctgcc acccatgtac atcaacacgc acacacacac aactattac	12660
aggtcagaga gcagctgtca cccatacata ccaacacaca cacacactat tacaggtcag	12720
agagcagctg tcacceatac ataccaacac acacacacac acacacacac acacactatt	12780
acaggtcaga gagcagctgt caccattca taccaacaca cacacacaca cacacacact	12840
attacaggtc gagagcagct gtcaccata aatacaagca taccctcac tccctccct	12900
acccacctc cattatagtc tctcaaagca taaggcagta ctaccagcca tgtctctaag	12960
tgtcatatgt caccctgatt tacacactga aggggaagat gcacattcta gaagatgcct	13020
tcaacaaatc acccccaaat ttcagactca gtgtgggaga ggcaggggta gaggtagata	13080
ggagcgtga aggggcgggg ccatggaaga aaactccacc caccactctg catcaccatt	13140
caataatggg ggctggtcta tctctatgcc aataaccctg agatcaaggg tcatggagat	13200
aaaggtccta cagatgtcc agggagaccc aggagctact ctgtaccaa cgctaccaag	13260
tgagcatcct ttggctttct cctcaactca ttctcccagg agagaacaga agtcatacaa	13320

cttaggaaca gcttcaaca tggtcccaat ttttttggga gggcaggggg agagtagagg	13380
tttgaagcag ggcttctctg tgcagccctg gccatcctgg aagtagttct gttccccagg	13440
ctgtcctcga actcacagag atcctcctgc ctgttttccc actaccagga ctaaagggtg	13500
caccaccaca ctccatttta taggtactgg tatcaaacct agggcttcac gcatgaaggt	13560
aggcaggcat tctaccatga gagcacagtc ccagcccagg gtgtcttctt ctgtcctacc	13620
acatcatgtt tacacttagt tatttgtggc ctaagggatg aaccacagaa gggcagcagc	13680
tcctgatcac tgtgctaaga ctctacctg gctgacggta gggaaaaaag gaccaagcct	13740
taaatctacc gtctgtaaga ggtggacaca acccagggag ggggacagga aggaggaagg	13800
agcagtaaga ggagcggaca ggaggagatg ggagatcatt tgtccccagg caaagcttaa	13860
ctggtggcag aacactggct ggggtgggaat tctgggttaa aaacaaaag cactggtttt	13920
cctcagagta gcacggccat ctgacatgtg ctggccactg cagaagagtt acttccaaag	13980
ttccctttca aaatgcaaac agagtgacct ccatgggcac gggaatgcac cacagggtc	14040
gcagggagcc caatggcatt agtctcctgc ctcatccctg catcagtgtc taagacatag	14100
agccctgttc actgtccaat ggcgaccccc ttctgcagt cctgcctgc agtccagcag	14160
gacaatctaa accccagccc caccaccatt cataccgat gcccacaggc tcaaaaaaca	14220
aaagagggag acacctccag gagctccaag gctgagccaa ggagacagga gtccctctgg	14280
gtgtacggc acttagggtg ttcccagtaa taaatgcga cactaccga tgtcaacctg	14340
ccagatcctg tcctaagagc cagcgcgcc tcactctatc tgatcatctt aataaccctg	14400
cctgcctacc agcctgtcgg gagagattat gaacatctct attggacaaa taaagagaca	14460
ggcccacct tagcagggtg gaaaccatct ctaaccctaa caggctgagg caggaggatt	14520
gcctcaagtt tgactggcta cacagaaatc tggctcgggt gggggcttag gatggctcag	14580
tagttaagag catgcctgt tcttcagag gaccaagtt tattttccag cgccacatc	14640
aggttcacaa acacctgtga gcgccagctc ctggaaacct gatgccacc tctgacctcc	14700
aaggagagag ggggagggag ggaggaagg aaggagggag ggaggaggag agggagggag	14760
ggaggaggag agagagagag aaagagagag agagagaaag agagaggaga gatggggcg	14820
tggatctgcg aggaaggcag gcaggcaaga tgaatgtgtg tgagagccta caacaaaaga	14880
ataaaaagga atgaaggagc tagctgagga acacctcatc agtgggaggc ttactgcaca	14940
gcatgaagac ctgagtttga tcttagaacc catctagaaa ataagctagg gctgaagaga	15000
tggctcagca gttaagagca ctgtgtttgc tcttgcaatt ccagcaacc acatgtagc	15060

tcacaacctg ggacctgatg ccctcttctg gcatgcagggt gtacatacag atagaacct 15120
 cataatattaa ataaataaat aaataaataa atcttaaaaa aataaaataa actagccata 15180
 gtagaacaca cttgcaatcc caaagctaag aatgtacaga cagacagaca gacagacaga 15240
 cagacaaaag atgcctaggg ctccctggcc agctcggcct actcagcaag tttcagacca 15300
 gtgaaagacc ctgtctcaat taaaaaaaa aaacaaaaaa cccttaacca cactcaaaga 15360
 acactacceca ggittgtctt ctggcctcca cacttgtatg tgtacacatg tgcgcgtgca 15420
 cgcacacaca cacacacaca cacacacaca cacacacaca cacagaggaa aggaggaaga 15480

cccaaggaag gataaaaagc tggtcagtca aagctggcct aaagctcctg ccgtaatcct 15540
 atcgcacagg aggcgtgggc agaagcacca ggagcttgag gccaacctag actacaagac 15600
 aagatcttga aaagaaacag agaggcaggg gcaggaagga cactcacaag tgcgagagag 15660
 cttgcagttg tgatgcacac ttttaatccc aggagaggca cagtaacttc tccagagtga 15720
 gticcaggac agccagggct acagagaagc cctgtctgtc agggggaaaa gaaaaaaaaa 15780
 agaaaggggg ggacagaaaa gaaaaagata gtctgggtggg aaatgattca ccgggtaaag 15840
 gcatgagcca ccaagtccaa agacctaaac ttgattgctg tcaccccat ggtaaaagag 15900

aacttctcc cacacgttgt tctctgacct ctacacatgt gcatggactg cgcagtctct 15960
 ctgacactcc atccatccat ccatccatcc atccatccat ccatccagt tgtgtgtgtg 16020
 tgtgtatgtg tgtgtgtgtg tgtgtgtaaa ataaataaaa aggaaggagg gagggacagc 16080
 aaaagatttg ttccatgtgg tgacagagcc gggatttgaa cccacatcta tgtgcctcca 16140
 gatcctgaag tctcaaggac ctgctacagc taccacaact gagaggatca ctactgacaa 16200
 cagatggaag ctggggcctt ggggttcagc gaggaggagg ccagcaacac acctacagcg 16260
 tggctcctga gccctggacc tacagcgtgg ctcttagggc cggcaccctt ac 16312

<210> 19
 <211> 11058
 <212> DNA
 <213> Rattus sp.
 <400> 19

gtgagtggac ccggtacact ggccctgttg ggttccggtg taggaaacgg aatccgctct 60
 ccctaagtgc tcagagatcc ctgtctggac ttgctagcct cactacagcc gtgcaaccga 120
 agtgacttaa ggccagacct tttttttttt ttttttctga gctggggacc gaaccaggg 180
 ccttgcgctt gctaggaagc tgctctacca ctgagctaaa tcccaaccc caggccagac 240
 cttttttaa ggccattcgt ttctctttt atttatttt tagtgtgcac tggatatttg 300

cctatgttta tgictgtgaa ggagcccagt gccctggagt tggagccaca gacagtgtg	360
ggtgctgaga actaaacaca ggtcttaacc actatgccgc ccccccccc cccagcctca	420
gtaaaggcca ttttctactt gcgatagggt cttgctatgt agaccagct ggcctcaaac	480
taacagagat tcccctgcct ctgtctccag tgtgctggta gtaaaggtgt acccccatgt	540
tcggcactag actgtttctt tagagtggtc tcatgtttaa agtaaacctg agttcccatc	600
tgctgtccc acacctcat ggcctctacc ccatcacctt atcctcggtc cgtgaccta	660
caatctacat tgactgggat gaaggaatga gattgacaga tcagcatcca aagtcactc	720
actgaatcca gtggcacatt cttgccacta gtgtcatact gggcattttc aggactgtga	780
ggtcacctca ctccccacc tgtggtggta tcagtcagtt ggaatcacag tttcagagtt	840
tgtttcttta acttgctgta tctttcatgg ctgcttttta aatagggaga taatgcttcc	900
gaccctgtct ttaccttttt ctccaggtg taatagatga atttgagcag aagcttcgtg	960
cctgtcacac cagaggcatg gatggaatag aggagcttga aactggccaa ggaggcagcc	1020
agcgggccct gtctgccaag aaacctccg ccggtgtgag aggcaagccc acgtgggcc	1080
ccacaggcag tcccctcttt tctttttttt cttttgtggt ttgtttttgg ttttgaaatc	1140
atgtccctct gcatattcca gtctcaaggc aacagcctct gacttctgta gactcagaca	1200
atgactttgt cacacctaa gctcgacgta ccaaaccctg tcgtccaca actcagcaac	1260
gcaagtccca gaggaagcc aaagtgtct tctcaagtga cgagtcagat gaagatggta	1320
agatacacc tccaacagg acaactgctt gtgtggggtc tccagccggg ctgacctggc	1380
tgtggattca ttctccctgg cagtgtcttc cccactgtgc agctgctggg ctgactttca	1440
ctctgtgtgt gtgtttgtgg tgtgtgtcca gatccgtttt gccccatctt tatgactcag	1500
cgtttgttcc tataattctt tgaatagaac ttccagcgga gatgaccgaa gaagagacac	1560
ccaagaaaac cacccecatc cgcagagcat ctgcgcaaag acacaggtcc taggagctgc	1620
ctttgcctat tccatccct gctgtgcagg gtatcttact gaatgatcga gaattcaacg	1680
ccccccccc ctttggaata tacttgttgc ctcttgttt ttaacatgta aatggtcttt	1740
tcgagcaagg cttttctctt agagccctg actgctactt tccaagatcc aatcaccatg	1800
ttttctagt gataaatgtt tttatatact ttttaagcatt ctggccggct ggtctgttcc	1860
attattaggc agcagatagc agtccagaga gagaacccca ggcatcctga gccatctctg	1920
gacctcatg gacctgtgg tgcttcatct ttgcacaact gctgagattt gctgtagaga	1980
gttctgtggt ttctagggtc acataagcac ttctcagcct gggggtactt tgcacccag	2040

aacacgtttg gcaatatcta gagaccaggg gtgggagaga tggctcagcg attaggagca	2100
cttactgccc ttccagaggt cctgagttca attcccagca accacatggt ggctcacagc	2160
catctgtaat gagatctggt gccctcttcc ggtgtgtcta aagacagcta tataataaat	2220
aaataaatct ttaaaaaaaaa aatctagaga cctatttctt ctgcccacct tgtggggcag	2280
gggctgctgt aggcacccag tcagtcagag cctaaggagg aagactcctt catccatagc	2340
actccaaca cagcatgtcg ggactgtgca actagagggt ctgctctaga cctgcggttt	2400
gtttcctgtg tgccacaccc catcccagaa gactggcaaa tgagtcagcc tggctggtgg	2460
cagagatcat gccattagct ctgttgggtga caaggccac ctccaagag tcagtcctcc	2520
actatgctaa ggggtctctc gctcagctgc agggccaggc tgtgctatgt gctgcccacc	2580
ttccccaagc cctaatctg ctgagtcact tggagcagga ggaggactg ttgtttccgc	2640
tgctatgaca acagcacagt gcagctgagc ctctccctac ctcttgggtat ggaaagagta	2700
atgacaggcg tgaggcagaa gaagatgcac ctcaaaaaca ctaggttagg ggtggactgt	2760
ggaagggtc ttgagacctc tgggtaggaa aagggttaac tcggtggata gacctcagt	2820
aggaagacag aaaagggaca gctgagctctg agatgcctgt ctcccagca ctgcactggg	2880
gccactccta cctttccac cctgtttttg gcctgtctcc tgctcatgaa aagcctaggc	2940
atggaccaca cagctgtcag tctaaacaca gctttggctc tggggcccca aggaggtagg	3000
gcaatccag aagggaaggg agccaggact agatttgggg tgcagcccag cgtgtccct	3060
gccttttaag caaaggttat caccaggcca gctaaactta gcaataggct cttcagctag	3120
aagtgcaggg ggtgtgtctc aagtgtcact gggctagcaa agaagcccca ggtccccc	3180
ggcccgacc tgctgatggc ccgcacctcc caatcaaccc cttccacct ttctagctc	3240
cctacctgc aaagcaaaac agtccctttc tcctccact gtgcaaagag gctaggggtt	3300
gcctgttctg gctcgggata ctccctttct gctttcttcc ctttctgac tcacctgct	3360
ttcacgaggg cagccaagga aagagagtcc ttggctcctg ggccagtcag caggtcgct	3420
ggatggggag ggacttccgc cctcacgtcc cagctctcca cctggggct gcggtgggtg	3480
aaagggacag tgtctccaat cctgggcggt acaactcagg ttccggggag ggaaggcctc	3540
tgagcctcct ccaacccaac cctcaacagg gatgcttacc ccggggcccc ggtataggtt	3600
cagcaggtaa actcaggga gtgttgtctc aatcaggttt ccagactgca ggggatcccc	3660
gacctatggc gtaggaatct ccttttcaag cctctagtca cttccttget ttcacatgc	3720
ggtacacaag ccaggactct ggcctaagag atagaagcca ggcggtgggg agagaggaaa	3780
atggggctgc gaagggtccc caccggagga aaagtgtccc ccccgctcct cccccagct	3840
tttcttctgc tccttttttc ggggcggggg gggcgtgctg tgtcactacc gaagaacaac	3900

gagaagatcc tcaacttctc ctaagccttt tcactaatag ggagaagttc gatggggcag	3960
ccttgggcag acccacactt ctgtccatt tccttggttc ctgcagctct cagattctcc	4020
cattttattc gggaagcagc tttctggttt ctgggtcctg gatgtccttg gtgcacactc	4080
caaggactcc tcgtccttaa tccatagtct gtattccctg agtcctatcc tgggaaccct	4140
catccggtca cttcctcggc gggacaatct cagctccctt cccctctca ggtcggagcc	4200
cacacgcttg gtgcgtgcac atttcaaaaa cgaggcgggt ccaaaaagag ggaggggggg	4260
aatgagagag gcccagctac tcgcggcttt acgggtgcac gtagctcagg cctctgcgcc	4320
cttagacttg gactggatga gccgagcggg aggcggggcg cgcgtcatca gtcccccca	4380
ccatccagtt cctataaata cggactgcag cctccctgg tgctctctgc tcctccctgt	4440
tctagagaca gccgcatctt cttgtgcagt gccaggtgaa aaagaaaaga aaaaaagga	4500
ctgggccgca ggaggccgga gacgaatgga aattaggaat ggggggaagg acgtgtacg	4560
ggtttagggg cgctggtgcg aggtccgga gccgagcca ggctccgcat tgcagaggat	4620
ggtagaggac gtgatgggc atgcggcggg aatggaggcg ggtggggga ggggactggc	4680
cacgctaate tgactttctt ctcccgcagc ctgctctcat agacaagatg gtgaaggteg	4740
glgtgaacgg gtgagttcca gggcggggcc ccgctccgtt gcgcatgccg gtctggctaa	4800
attcggggct gtgctgtagc gtgcggggag gtggataggg tggccgaagt acccaaggag	4860
acctcaaggt cagcgtcgg acctggggaa ggctcgact taacgggacc ccagctgttg	4920
ctcccctgt aactccgct ttgctgggga gcggcccgga gtctaaagta gcaggagcac	4980
ccccgacact ggcgcgacac cgctcaggat ccgaccgta atccccagc tggggttct	5040
ttctttactt tcgcgccagc aggagtcacg tgccagaagg ggagccctc cccatcgtc	5100
cccttctca gggtcggttg cacgggggtg tgagggtcat acctttgcgc atcttctaag	5160
ggcttcactt atggtaactg gccgccgcca tgttgcaaac gggaaggaaa tgaatgaacc	5220
accgttaaga aatcttctt cggccttctt cattcttagc ttgtgactaa ctttctcatt	5280
cctctcagct gggtaggtg tcctttattc ttaggccag gtgatggatg gcagggttc	5340
cgtgctctca ttagagctct actatacaat ccgtctctcg tcagcctta taggcctca	5400
gccttatttc cctcgggcat aacttatatt ctagattgtt ctctggaagt cgaagtgc	5460
ttacagaggt ctacttggcc tccagcccc ggaggtgcgg tagcaatggc gtagtgccga	5520
gcccgggggt ggggtgggga gctgagtcac gatggtcctg aaaagaaatt ttctgccaca	5580
aaatggctcc ggtggtagca gccccctcc tccaggcct ccacttccat cccagctcag	5640
cactgacca aaccctatag gccaggatgt aaaggtcact aagaggattg ggtgtctctg	5700
agcctcagga tctaccctt tcccccaacc catctccag aaaccagatc tccccattcc	5760

gccctgatct ggggttaaatt ttagctgtct gacctttctg tatctggggt ctgagctggg 5820

ctctctgccc tgttcatccc tccacacatc tgttgctcct gctccgattt tttttttttt 5880

tttttttttg ccgggaaaga cagggtgttt gcaaatgagt cctgggatta gggctggaaa 5940

atcactgggt ttcttgatct ttccccaga aagggacttg gaggaagcag gtgggctggc 6000

cctgtcctgc tcaactctgac ctttagcctt gccctttgag ctgctgatga atgagttctc 6060

tcctgtccgg gagtgtagcc tgaagtccag ccatgctgga accggcttcc caggcctgtg 6120

tgggtgggggt gatgaatacc atgtgccaag ctgtaccca gaccggggca aaccgccact 6180

tcttaagaga cttaaaatga ctttgggggtg ccgggcaagg ctgtgggcct aatcacctct 6240

tggacaggaa aggaacctcc actttatagc tgctataaaa gtctctccag ccctgggtgg 6300

ctcaaggaat ataaaattag atctcttttg actttctagg gaggaacat tctatatattg 6360

ggttgtacat ccaagcattc aaccggcttt attaaaggaa attctggaca aaacatgaac 6420

ttcctgctgc ttacgtttct gatgctgtac tgaatcagcg taagaacttg cacctttttg 6480

tgatgcgtgt gtagcgggct gctgtaggat ctccacgcc atggtgcagc gatgctttac 6540

tttctgaaga acatgcgttg gccagggct gactacaaac ccaggagggg ttcactgac 6600

ccaactaact cgctatttc ttgcctcaga ttggccgta tcggacgcct ggttaccagg 6660

gtgccttct cttgtgacaa agtggacatt gttgccatca acgacctt cattgacctc 6720

aactacatgg tctacatgtt ccagtatgac tctaccacg gcaagttcaa cggcacagtc 6780

aaggctgaga atgggaagct ggtcatcaac gggaaacca tcaccatctt ccaggagtac 6840

glatggaaac atgcacaggg tacttcgagg agatggtaag gggcggcatt aactctgtcc 6900

tgttccccgc ctgtaggcga gatcccgct acatcaaatg gggatgatgt ggtgctgagt 6960

atgtcgtgga gtctactggc gtcttcacca ccatggagaa ggctgggta agtggccagg 7020

aagctggagg taaggggaac ctttgatatg ggtgcaacct ccaaactgaa gagctgagtc 7080

tgaaatcaac ttctctgca caggctcacc tgaagggtgg ggccaaaagg gtcacatct 7140

ccgcccttc cgtgatgcc cccatgtttg tgatgggtgt gaaccacgag aaatatgaca 7200

actccctcaa gattgtcagc aatgcatcct gcaccaccaa ctgcttagcc ccctggcca 7260

aggtcatcca tgacaacttt ggcatcgtgg aagggtcat ggtatgtagg caatggagac 7320

agctcatgca tgagtggacc tttctttgaa gatgtccctt tgggtaggag ggctgccctg 7380

caagacctca cccattgcct ctatgctttc tagaccacag tccatgcat cactgccact 7440

cagaagactg tggatggccc ctctggaaag ctgtggcgtg atggccgtgg ggcagcccag 7500

aacatcatcc ctgcatccac tggctgctgcc aaggctgtgg gcaaggtcat cccagagctg	7560
aacgggaagc tactggcat ggccttccgt gtccctaccc ccaatgtatc cgttgtggat	7620
ctgacatgcc gcctggagaa acctgtacgt agggggagag ctgggtttgt tttgtctatg	7680
gtatgggtgt gacttggggc aagagcagtc atctttggtt ttgcccttac aaggccaagt	7740
atgatgacat caagaagggt gtgaagcagg cggccgaggg cccactaaag ggcacccctg	7800
gtacactga ggaccaggtt gtctcctgtg acttcaacag caactcccat tcttccacct	7860
ttgatgctgg ggctggcatt gctctcaatg acaactttgt gaagctcatt tcctggtatg	7920
tggggaccgc aagcctggct cttgagagta actgaaggtt tggcgccctc tggcggccag	7980
cttagaaaag aagcccaaac taaccgttgt cccaatctgt tctaggtatg acaatgaata	8040
tggctacagc aacagggtgg tggacctcat ggcctacatg gcctccaagg agtaagaaac	8100
cctggaccac ccagcccagc aaggatactg agagcaagag agaggccctc agttgctgag	8160
gagtccecat cccaactcag ccccaaacac tgagcatctc cctcacaatt ccatccaga	8220
ccccataaca acaggagggg cctggggagc cctcccttct ctcgaatacc atcaataaag	8280
ttcgtgcac cctcttcttg tgtgtgaatg cagtgttgtg ggggttatgg gcaggtgaag	8340
gaggctccat tcggttttgt gcacatatct ttattggctc tgccatccag ctgggtgcta	8400
gaggtccagg gctacattct aaccgagtgc cagcccaggc cgctgtcca gacacctgat	8460
gtgtctggca gtgtggaggg tggggcttga cactggaaca gaaggtgcag tctttgaaca	8520
gcactgactt tcagagcttg ctgctcttga gaaagatctg tgtcctttat ctgcaaaaa	8580
aggttaattg taccagctc tcaaagtctc cagatgttaa attaggtaaa gttggaagg	8640
tgccagtac agtaggacta attgtgtcca tgaagagga aggtccctgc tgagctgttc	8700
acctgtcttc aggtgtgggc acttgaaggt gctttggttc tggggcagtg actagccagc	8760
gtcactgtt ggttgggaat gaaatgaagg tgcattacac ggtgactcga ctcaacgcct	8820
tcattgcttt caaaggtccc attcaaatgt cagctctgca tagcctctcc tgggagatcc	8880
tagctcccta ggaccaggga cattgtaacc ttttggtttt gtcagtaca gccatggagg	8940
cattccctca gctggctact gaggagacac cctgagcacc cctaaccttg tgcaggccac	9000
acccaagtgt ggcagcagct tcattctcag ggccacctgg cccaggcact ggacttccca	9060
aaggcacagc catgttctact gcagtgggca ggtgggcagc tagctgtgag catgtgcagt	9120
gtccacgccc tcaaggcctt cctagggagc aaggctctat ccaaaggtag atgcggaatc	9180
tgtttcacat gagtgcctag gccagttagc accagtacaa aaccaagga aaacaggcag	9240
gtcccataac cacattcctt acaaatggaa aaggagggt ccaagtgtgc tggcagcatt	9300
ccaaagatgg agaagtagca gatacccaca tgtgaggaag gcaggagtgc cccaatgtgt	9360

tgggagtgtg acagggggtg gggccgagcc tagctccaga caccacagaa ccctggagag 9420
 gacaagccag ggccaggctt catgcggtgg cacagaagat ggtcctctgc agaccacac 9480
 ccctgcccc gagagctctg gggcaccagg ccctggttcc tatctcatgg agctgtcaga 9540
 tgagacatct cgatcggagt cctcagtctc gcttggcggc ggcggcgggt cgctaagcgg 9600

 gaccgcagtg aaagcaggag actttctaga aagaaatacc agttgtcacc ttggggcaaa 9660
 gcaggaaagc tectgagact tccaaacttg ctttctctgg cttatcttt tgaggtagag 9720
 ggggatgaca ggcaaggacg ggggagggaa ggaaaacttg tttcaactt gatgggagga 9780
 acagtctgc tctttgggac tcagagatca gaaaagtctg aagtcacagc atggagatgt 9840
 tccctgagcc aactaaggc cctagcttct tatccagcaa taattcgggtg tgcggggggc 9900
 tgtggggact gtaggaggtg agactggggg agtatggggg tttagggggc tgtttcactc 9960
 tgttgtcact gctgaccca aattcctggg tgcaagcaat cctcttcct tcagttgatt 10020

 ttctttttt ttttaattta ttttattat atgagtacac tgtaactgtt ttcagacata 10080
 ccagaagagg gcatcagatc ccattacaga tggttgtgag ccaccatgtg gttgctggga 10140
 attgaactca ggacctctgg aagagcagtc agtgccttta accgctgagc catctctcca 10200
 gccccctctt ccttcaaagt ggctgggggtg gccacaaacc cagttagcgt gctcaccata 10260
 caatgcaatt tggggggctg ggttgggttg ggtttttgt tctctttgag atgagtcagt 10320
 ttgcattcct gectggcctt gaactcctca cagaggcgag gctggtctag tactttggc 10380
 aattcacctg cccgttgctt ctgtgcatct gaatgatagg tatgcgctca taagcctgac 10440

 ttaactgcca ttccctggat ttttgttgt tgttgtgtg gtttggtttg tgtgggacaa 10500
 ggttttactg ctgtccagaa cagcctaaaa ctcatgacaa tccttctgcc tccgccttcc 10560
 aagactgaga ctgcagacat gtaccgctag gcctggttta gagatttcta aaaaaatctt 10620
 agtgtgtgtg tgtgtgtctg tgtgtgtgtg tgtgtgtgtc tgtgtgtgtg tgtgtgtgtg 10680
 tctgtgtgtg tctgtgtgtg tctgtgtgtg tgctagacag gctagacaag gactctacca 10740
 ctgagctata ttccaaaacc ccagatttac tcttaacaa ggtcaaacag ctaaactgtg 10800
 atttgattt ggataatcct gtccctctat tcttttctc cttgccaggt ttgggctctc 10860

 agaccctcag agttgggcac tagtcggttt catctgacgg ccaaggcagg gagaggaagc 10920
 ttcatcttag gccacacct cagggactcc caaggacagc tggacataaa gatgtgacct 10980
 ggtcactccc acaggccact ggggcctgtc cagccttctt cctggggaga cagccacgcc 11040
 actcagggcg gcccttac 11058

<210> 20

<211> 3158

<212> DNA

<213> Homo sapiens

<400> 20

agcttagat tgcacaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg	60
gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt	120
gtttcttctt cttttttaaa tcacgttttc ctgccttttc taggttctag gtaccagcct	180
ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccctg	240
cggcatccaa acaccacga gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt	300
gatgagtcca gtgaggaagg tatgatgtc cgcctgttc ccggccgaga aggcacacag	360
ctagggtgca gagggctggt ttccatagga cctgtgcgg gggcctgagt gtagatgtc	420
tgccccactg ccgcagaagg gcctctctg tacagcttgg atttatttc ttctgtcgg	480
tgtgggattg tctcacttgt tctctgatat ctatttttc accatcttg tgactcagct	540
ttttcttatt cttttaattc ttgcataga tctttcagca gagatgacag aagacgagac	600
acccaagaaa acaactccca ttctcagagc atcggtctgc aggcacagat cctaggaagt	660
ctgttctgt cctccctgtg cagggtatcc ttaggggtga cctggaattc gaattctgtt	720
tcccttgtaa aatatttgtc tgtctctttt ttttaaaaa aaaaaaggcc gggcactgtg	780
gtcaccgct gtaatccag cactttgcga taccaaggcg ggtggataac ctgaggtagg	840
gagttcaga ccagcctgac caacatggag aaacccatc tctactaaaa ataaaaaatt	900
agccgggctg attggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa	960
tgcctgaac ccagaggcgg aggttgtagt gagccgaaat cacaccattg cactccagct	1020
tgggcaacaa tagcgaacct ccatctcaaa ttaaaaaaaa aatgcctaca cgctctttaa	1080
aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct tcaactttgg	1140
aatatatgtt tgccaatctc cttgttttct aatgaataaa tgtttttata tacttttaga	1200
cattttttcc taagcttgtc ttgttttcat ctttcacatt agcccagttt catgcagcag	1260
agagagggtt atcagtgcag agagagatga gtgagcccag agtcctaggg cctgtcccgg	1320
gatggcagat gagcttctg ccccgctact gccacctttc ccctctcaac ctctggaccc	1380
tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc tccagatcag	1440
tgcttctgaa ccgggggcaa tttgtctgc cagaggacat ctgacaacac ctggggcctg	1500
ttttgtgtc atagcctata ggggaagaat gctaccagca tttgtgggaa gaggccaggg	1560
atgtggctca acatctgca gtgcacagga tggccctca acaaagaatc acacggccca	1620
caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg ctttctgccg	1680

tggtgctcac cccaccccca ctctgtttcc ctaatcccat ctccaaaggt tggcagcaga	1740
ccggcccagg ctcttggaag ttcagatcat gatccctcc agctctgcag gagacaagac	1800
ctgtctccca gcatttctca ttgttcccgg gtctgcagag ggctgagct atgtgcagg	1860
cgggtgtccc ccigaagcct gcgcaccct ctccagctcc tcaagtcttc tctgtgagt	1920
caccttcgaa ccggaggctg tgagctggct gtctgacca cactggtgcc tctgtgtca	1980
tgacaacagc aactacgtc agtagtctc cctgggcact gagtccctc ttgctggga	2040
gaagacagta atgaaaaatg acaagcatga ggcagagggg aagatcacgc ttgggtggtg	2100
caggagcatg gaggtgtctt taatgtctc aatgagaaag ggtaaacgt cctggttga	2160
ggaatagctg agtcagaggt ggggttctt ccactcccc accccacccc ttaccatt	2220
agggaccttc ttgcttctt cttgtactc tgctctgggt ggtcattgtg aaaagcccgc	2280
accaacctg ccagtggcag ccagacgagg acacagcctg gctctgggtc ccagcaggaa	2340
aggcaatccc agaaaggcag ggtcaggac tggagtctg tgggtgtctt ttaagcaaag	2400
attatcacca ggcaggctaa acttagcaac cgcttttag ctagaagggc aggggctgg	2460
tgtaggtta tgctgggcca gcaaagaggc ccgggatccc cctcccatgc acctgtgat	2520
gggccaaggc cccccccc cacccttc ctacaagtg ttcagcacc tccatccca	2580
cactcacaaa cctggcctc tgcctccta ccagaagaat ggatccctg tgggagggg	2640
caggggacct gtcccaccg tgtgccaag acctctttc ccacttttc cctctcttg	2700
actaccctg cctcaatat ccccggcgc agccagtga agggagtccc tggtcctgg	2760
ctgcctgca cgtcccagg cggggaggga ctccgcct cactccgc tcttgcctc	2820
aggctggatg gaatgaaagg cacactgtct ctctccctag gcagcacagc ccacaggtt	2880
ccaggagtgc ctttgtgga ggcctctgg cccaccag ccacctgtc ctccgctgg	2940
ggccccagc cgagagagc cgctggtgca cacaggccg ggattgtct ccctaattat	3000
caggtccagg ctacagggt gcagacatc gtgacctc gtgcagaaac ctccccctc	3060
ccctcaagc gcctccgag cctcttct ctccagccc ccagtgccca gtgccagtg	3120
cccagcccag gcctcggtc cagagatgcc aggagcca	3158
<210> 21	
<211> 3218	
<212> DNA	
<213> Homo sapiens	
<400> 21	
aagctggcac cactactca gagaacaagg cttttctc tctcgtcc agtcctaggc	60

tatctgctgt tggccaaaca tggaagaagc tattctgtgg gcagccccag ggaggctgac	120
aggtggagga agtcagggt cgcactgggc tctgacgtg actggttagt ggagctcagc	180
ctggagctga gctgcagcgg gcaattccag ctggcctcc gcagctgtga ggtcttgagc	240
acgtgctcta ttgctttctg tgccctcgtg tcttatctga ggacatcgtg gccagcccct	300
aaggtcttca agcaggattc atctaggtaa accaagtacc taaaacctg cccaaggcgg	360
taaggactat ataatgttta aaaatcggta aaaatgccca cctcgcatag ttttgaggaa	420
gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg gaacttgggt	480
aggggcaagg cgtgtagctg ggacctaggt ccagaccct ggctctgcc ctgaacggct	540
cagttgcttt gggcagttac tcccggcct cactttgcac gtgtgcttac ctagtggaga	600
caaaagtaca tacctcggtg gagcgcgcac gcctgtaacc ccagcacttt gggaggccaa	660
ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat ggtgaaactc	720
cgtctctact aaaattacaa aaatcagcca ggcttcatgg cacatgccta tagtcccagc	780
tacaggcatg ctgaagcagg agaatcgctt gcaccccgga ggagaggct gcagtgagct	840
gagaccacac cactgcactc cagcctaggc aacagagtat gagactccat ctcaaaaaaa	900
aaaaaagtac ctacctcaga gttcaaaacta gtgaatatta ggaagtgcct gagacagtga	960
caccaaagtg cacaataaat actcgccagt ttcatatta ttaaagaatc catttgaatg	1020
tcagctcaac acagcctcct ataccgaggc attgtgaacc gcctctccc agcttctcca	1080
ggcttttcca agaatcaggg aactgtagc ctgttggtct cagtgtatga cagacacgga	1140
ggaagcacat ctttagctga tacttaacaa gagacctga gcgcacatac acccgcgac	1200
acatgcatgg agcttcacct tctctgtcat tctgcagtga ccaggagagc aagagctccc	1260
acctcccttc aaaacactgt gccatcccg ggactaagg cctctttaaa gcacggcacc	1320
tccacgaggg agggccacag ccacatacac tccacctggc aggtggacag cgtgagcacg	1380
tggacatag cagggacaag gtgccccggc cagcccaac gccctctgcc gctgacaggg	1440
acagaagccc tctccagctg cgtgtgctgc agaggccatg cgtagcctcc agctgcattc	1500
tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga gctggctccg	1560
gacaacaggg atggaggaaa ggtcccacat tcacattcct gatacgtgga caaggtgagg	1620
ggccgcaatc gctctggcag cattttaaag atggggaagt agcagacacc cacgcgtgaa	1680
ggcaggagag cccaactgt ggtggaaatg gccccagaat ggtagggcca agcctagctc	1740
cagacacccc agagccctgg agaagccaag actgagggag aaagcctgag ggaggagcgc	1800
cccagtcctc agggaccggc ctggtgcaga gctgcagctg atgttccct ctgtgcagcc	1860

ccacctctg cctcgctgag ctccctgctg cgagggcctc gggcgcaagg gggaggcagg 1920

tctctatctc atggagctgt cagatgagac atcgcgatcg gaggcctcag cctcgcttgg 1980

cggcgggggc gggcgctaa gcgggaccgc agtgaaagca ggagactttc tagaaaaaaa 2040

caccagtgtt caaccttggg gcaggcagga atcctgaaga cggacggcac tctcctcct 2100

gctgcctcac cctctggcag cccgtgagaa gtaccggaag cgagggcggg gccgcgggat 2160

ggcgaggagg cggcagggac tgaactctct ccaaaccac cctgacaggg aaatgggccc 2220

cgctgtgtc ttgggaactc agaggctgag gtcaggcatg atggctcacg cctgtaattc 2280

cagcactttg ggaggcagag gcgggtggat cagcagctca ggagttcaag gcaagcctgg 2340

ccaacatggt gaaaccccat ctctactaaa aatgcaaaaa ttagccaggt gtggtggcgg 2400

gcgccttga ggctgaggca agataatcac ttgaacctgg gaggcggagg ttgtagttag 2460

caaaaaaaaa aaaaaaaga aatagctgaa gtcacagtag gagagaagct gctgagcctc 2520

cagcacctgt actctagggc cttggcttta tctctatctg cagtattttt gtgattttta 2580

aaaattcact ttctgttgc ggtgtaactt acacagggtc aaatgcacaa atcatggccc 2640

tgactttaga taaaaatctg cccccacaac cttctgttc cttgccagtt tttaactgc 2700

ctctaaccag gggaaccacc agagctgggt gccttgggag gtttcagccc tcccgcatg 2760

aatggacata gtcacccaa ctgccaaggg agagagctgt gggctctggc cagccccacc 2820

aggtaactcc caaagggcag cccacagca agatgtgacc cagtcattgc ctgagggtct 2880

ctggggctgt gttccaact ttctccccgc tgtgtcccc tggaaggccc catgccagg 2940

ggaggcgcct accggtctcc agactgggtg atgagccggc ggcaggcttc catctgcacg 3000

tccaggccgc gcttcatgct gcacatctcc atgtactcgt gcaggtgccg gttcatgtcg 3060

ttcttggccg tgccaactc cagctgtggt ggggcaggca gggccatcgg ggttaacagg 3120

tggcgttcac agcgctctg ttgccccgc caggaggcca acacccaag agcagtggct 3180

gggccccggg ccaggcagc cattactgaa ggctcaga 3218

<210> 22

<211> 508

<212> DNA

<213> Homo sapiens

<400> 22

agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60

gtacgtgaagg cagcctgtgc gggcgagacc agactgggccc ctcctcctct gcagtgtatt 120

gtttcttctt cttttttaaa tcacgttttc ctgcttttc taggttctag gtaccagcct 180

ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccgt	240
cggcatccaa acaccagca gcgagcttcc aaaaagaac ccaaagttgt cttctcaagt	300
gatgagtcca gtgaggaagg tatgatgtc cgcctgttc ccggccgaga aggcacacag	360
ctagggtgca gagggctggg ttccatagga cctgctgcgg gggcctgagt gtagatgtc	420
tgccccactg ccgcagaagg gcctctctg tacagcttg atttatttc ttctgtcgg	480
tgtgggattg tctcacttg tctctgat	508
<210> 23	
<211> 2650	
<212> DNA	
<213> Homo sapiens	
<400> 23	
atctatttt tcaccatctt tgtgactcag cttttctta ttctttaat tctttgcata	60
gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga	120
gcacgcgctc gcaggcacag atcctaggaa gtctgttctt gtctccctg tgcagggtat	180
cctgtagggt gacctggaat tcgaattctg ttcccttgt aaaatattg tctgtctctt	240
ttttttaaaa aaaaaaagg ccgggcactg tggtcacgc ctgtaatccc agcactttgc	300
gataccaagg cgggtggata acctgaggta gggagtcca gaccagcctg accaactgg	360
agaaacccca tcttactaa aaataaaaaa ttagccgggc gtattggcgt gcgcctgtaa	420
tcccagctac tcaagaggct gaggcaggag aatgcctga acccagagc ggaggttga	480
gtgagccgaa atcacacat tgcactccag ctgggcaac aatagcgaac ctccatctca	540
aattaaaaa aaaatgccta cacgtctttt aaaatgaag gctttctctt aaattagcct	600
aactgaactg cgttgagctg ctccaacttt ggaatatatg ttgccaatc tctttgtttt	660
ctaataaata aatgttttta tatactttta gacattttt cctaagctg tctttgtttc	720
atctttcaca ttagccactg ttcatgcagc agagagaggg ttatcagtgc agagagagat	780
gagtgaagcc agagtccatg ggcctgtccc gggatggcag atgagcttc tgccccgtca	840
ctgccacctt tcccctctca acctctggac cctgcacagt gaccagacag cctctctggg	900
gagaattatg cagtgcctag gctccagatc agtgcctctg aaccgggggc aattttgtct	960
gccagaggac atctgacaac acctggggcc tgtttgttg tcatagccta taggggaaga	1020
atgtaccag catttgtggg aagaggccag ggaatggct caacatcctg cagtgcacag	1080
gatggccct caacaaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa	1140
acctgtctca gaccaagggt tgctttctgc cgtgtgcctc accccacccc cactcgtgtt	1200

ccctaataccc atctccaaag gttggcagca gaccggccca ggctcgtgga agttcagatc	1260
atgatccctt ccagctctgc aggagacaag acctgtctcc cagcattcct cattgttccc	1320
gggtctgcag agggcgtgag ctatgctgca ggcgggctgc cccctgaagc ctgcgcaccc	1380
ctctccagct cctcaagtct tctctgctga gtcaccttcg aaccggaggc tgtgagctgg	1440
ctgtcgtgac cacactgggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc	1500
tccctgggca ctgagctccc tctttgcggg gagaagacag taatgaaaaa tgacaagcat	1560
gaggcagagg ggaagatcac gcttgggttg tgcaggagca tggagggtgct cttaatgctc	1620
tcaatgagaa aggggttaacg gtcctgggtt caggaatagc tgagtcagag gtggggcttc	1680
ctccactccc ccacccacc ctttcacca ttagggacct tcttgcttg ctcttgctac	1740
tctgctctgg gtggtcattg tgaaaagccc gcaccaacca tgccagtggc agccagacga	1800
ggacacagcc tggtcttggg tcccagcagg aaaggcaatc ccagaaaggc agggtcaggg	1860
actggagtcc tgtgggtgct ttttaagcaa agattatcac caggcaggct aaacttagca	1920
accggctttt agctagaagg gcagggggct ggtgtcaggt tatgctgggc cagcaaagag	1980
gccccgggac cccctccat gcacctgctg atgggccaag gccacccac cccacccct	2040
tccttacaag tgttcagcac cctcccatcc cacactcaca aacctggccc tctgcccctc	2100
taccagaaga atggatcccc tgtgggaggg ggcaggggac ctgttccac cgtgtgcca	2160
agacctcttt tccactttt tccctcttct tgactaccc tgcctcaat atccccggc	2220
gcagccagtg aaaggagtc cctggctcct ggctcgctg cacgtcccag ggcggggagg	2280
gacttccgcc ctcagctccc gctcttcgcc ccaggctgga tggaatgaaa ggcacactgt	2340
ctctctccct aggcagcaca gccacaggt ttccaggagt gcctttgtgg gaggcctctg	2400
ggccccacc agccatcctg tctctcgctt ggggccccag cccggagaga gccgctggtg	2460
cacacagggc cgggattgtc tgccctaatt atcaggcca ggctacagg ctgcaggaca	2520
tcgtgacctt ccgtgcagaa acctccccct cccctcaag ccgctcccg agcctccttc	2580
ctctccaggc cccagtgcc cagtggccag tgcccagccc aggcctcggt cccagagatg	2640
ccaggagcca	2650
<210> 24	
<211> 1963	
<212> DNA	
<213> Homo sapiens	
<400> 24	
agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg	60

gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctccctcct gcagtgattt	120
gtttcttctt cttttttaa tcacgttttc ctgccttttc taggttctag gtaccagcct	180
ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccgt	240
cggcatccaa acaccagca gcgagcttcc aaaaagaaac ccaaagttgt ctctcaagt	300
gatgagtcca gtgaggaagg tatgatgtc ccgctgttc ccggccgaga aggcacacag	360
ctagggtgca gagggtgtt ttcatagga cctgctgcgg gggcctgagt gtagatgtc	420
tgccccactg ccgcagaagg gcctctctg tacagcttg attttatttc ttctgtcgg	480
tgtgggattg tctcactgt tctctgatat ctatttttc accatcttg tgactcagct	540
ttttcttatt ctttaattc ttgcataga tctttcagca gagatgacag aagacgagac	600
accaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt	660
ctgttctgt cctccctgtg cagggtatcc ttaggggtga cctggaattc gaattctgtt	720
tcccttgtaa aatattgtc tgtctctttt ttttaaaaa aaaaaaggcc gggcactgtg	780
gtcacgcct gtaatccag cactttgcga taccaaggcg ggtggataac ctgaggtagg	840
gagttcgaga ccagcctgac caacatggag aaacccatc tctactaaaa ataaaaatt	900
agccgggctg attggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa	960
tgcctgaac ccagaggcgg aggttgtagt gagccgaat cacaccattg cactccagct	1020
tgggcaaca tagcgaacct ccatctcaaa ttaaaaaaaaa aatgcctaca cgctctttaa	1080
aatgcaaggc ttctcttaa attagcctaa ctgaactgcg ttgagctgct tcaactttgg	1140
aatatatgtt tgccaatctc ctgttttct aatgaataaa tgtttttata tacttttaga	1200
cattttttcc taagctgtc ttgtttcat ctttcacatt agcccagttt catgcagcag	1260
agagagggtt atcagtgcag agagagatga gtgagcccag agtcttaggg cctgtcccg	1320
gatggcagat gagcttctg cccgtcact gccaccttc cctctcaac ctctggacce	1380
tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc tccagatcag	1440
tgtcttgaa ccgggggcaa tttgtctgc cagaggacat ctgacaacac ctggggcctg	1500
ttttgtgtc atagctata ggggaagaat gctaccagca tttgtgggaa gaggccagg	1560
atgtggctca acatcctgca gtgcacagga tggccctca acaaagaatc acacggccca	1620
caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg ctttctgcc	1680
tgtgcctcac cccacccca ctctgttcc ctaatccat ctccaaaggt tggcagcaga	1740
ccggcccagg ctcttggaag ttcagatcat gatccctcc agctctgcag gagacaagac	1800
ctgtctccca gcatctca ttgttcccg gtctgcagag ggcgtgagct atgtgcagg	1860
cgggctgccc ctgaagcct gcgcaccct ctccagctcc tcaagtcttc tctgtgagt	1920

caccttcgaa ccggaggctg tgagctggct gtcgtgacca cac 1963

<210> 25

<211> 1195

<212> DNA

<213> Homo sapiens

<400> 25

tggtgcctct gctgtcatga caacagcaca ctacgtcagt agtgctccct gggcactgag	60
ctccctcttt gcggggagaa gacagtaatg aaaaatgaca agcatgaggc agaggggaag	120
atcacgcttg ggttggtcag gagcatggag gtgctcttaa tgctctcaat gagaaagggt	180
taacggctct ggittgcagga atagctgagt cagaggtggg gcttcctcca ctccccacc	240
ccaccctttt caccattagg gacctctttg ccttgctctt gctactctgc tctgggtggt	300
cattgtgaaa agcccgacc aacctgcca gtggcagcca gacgaggaca cagcctggct	360
ctgggtccca gcaggaaagg caatcccaga aaggcagggt cagggactgg agtcctgtgg	420
gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg cttttagcta	480
gaagggcagg gggctgggtg caggttatgc tgggccagca aagaggcccg ggatccccct	540
cccatgcacc tgctgatggg ccaaggccac ccacccccac ccccttctt acaagtgttc	600
agcaccttc catccacac tcacaaacct ggccctctgc cctctacca gaagaatgga	660
tcccctgtgg gagggggcag gggacctgtt ccacacctgt gcccaagacc tcttttcca	720
ctttttccct cttcttgact cacctgccc tcaatatccc ccggcgcagc cagtgaagg	780
gagtccttgg ctcttgctc gcctgcacgt ccaggggcgg ggagggactt cgcctcac	840
gtcccgtct tgcgccagg ctggatggaa tgaaagcac actgtctctc tccttaggca	900
gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca	960
tcctgtctc cgctggggc ccagcccgg agagagccgc tggtcacac agggccggga	1020
ttgtctgccc taattatcag gtccaggcta cagggtgca ggacatctg accttccgtg	1080
cagaaacctc ccctccccc tcaagccgc tccgagcct ctttctctc caggecccca	1140
gtgcccagtg ccagtgccc agcccaggcc tcggtcccag agatgccagg agcca	1195

<210> 26

<211> 256

<212> DNA

<213> Homo sapiens

<400> 26

cctctgggcc cccaccagcc atcctgtcct ccgcctgggg cccagcccg gagagagccg 60
ctggtgcaca caggcccggt attgtctgcc ctaattatca ggtccaggct acagggtgc 120
aggacatcgt gaccttccgt gcagaaacct ccccccccc ctcaagccgc ctcccagacc 180
tccttctctt ccaggccccc agtgcccagt gccagtgcc cagcccaggc ctcggtccca 240
gagatgccag gagcca 256

<210> 27
<211> 1941
<212> DNA

<213> Homo sapiens
<400> 27

agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60
gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
gtttcttctt cttttttaa tcacgttttc ctgccttttc taggttctag gtaccagcct 180
ctggcttcta cagcctcaga caatgacttt gtacaccag agccccgcg tactaccgt 240
cggcatcaa acaccagca gcgagcttcc aaaaagaaac ccaaagttgt ctctcaagt 300
gatgagtcca gtgaggaagg tatgatgtc ccgcctgttc ccggccgaga aggcacacag 360
ctagggtgca gagggtggt ttccatagga cctgctcgg gggcctgagt gtagatgtc 420

tgccccactg ccgcagaagg gcctctcctg tacagcttg atttatttc ttctgtcgg 480
tgtgggattg tctcacttgt tctctgatat ctatttttc accatcttg tgactcagct 540
ttttcttatt cttttaattc ttgcataga tctttcagca gagatgacag aagacgagac 600
acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660
ctgttctgt cctccctgtg cagggtatcc ttaggggtga cctggaattc gaaccggagg 720
ctgtgagctg gcgtcgtga ccacactggt gcctctgctg tcatgacaac agcacactac 780
gtcagtagtg ctccctgggc actgagctcc ctctttgcgg ggagaagaca gtaatgaaaa 840

atgacaagca tgaggcagag gggaagatca cgttgggtg gtgcaggagc atggaggtgc 900
tcttaatgct ctcaatgaga aagggttaac ggtcctggtt gcaggaatag ctgagtcaga 960
ggtggggctt cctccactcc cccacccac ccttttacc attagggacc ttcttgctt 1020
gtctttgcta ctctgctctg ggtggtcatt gtgaaaagcc cgcaccaacc atgccagtgg 1080
cagccagacg aggacacagc ctggctctgg gtcccagcag gaaaggcaat ccagaaagg 1140
cagggtcagg gactggagtc ctgtgggtgc tttttaagca aagattatca ccaggcaggc 1200
taaacttagc aaccggcttt tagctagaag ggcagggggc tgggtgcagg ttatgctggg 1260

ccagcaaaga ggccccgggat cccctccca tgcacctgct gatgggcaa ggccacccca 1320
 cccaccccc ttccttaca gtgttcagca cctcccatc ccacactcac aaacctggcc 1380
 ctctgcctc ctaccagaag aatggatccc ctgtgggagg gggcaggga cctgttcca 1440
 ccgtgtgccc aagacctctt tcccacttt tccctcttc ttgactcacc ctgccctcaa 1500
 tatcccccg cgagccagt gaaagggagt cctggctcc tggctcgct gcacgtcca 1560
 gggcggggag ggacttccgc cctcagctcc cgtctctgc cccaggctgg atggaatgaa 1620
 aggcacactg tctctctcc taggcagcac agcccacagg tttccaggag tgcctttgtg 1680

ggaggcctct ggccccccac cagccatcct gtctccgcc tggggccca gcccgagag 1740
 agccgtggt gcacacagg cgggattgt ctgccctaat tatcaggctc aggctacagg 1800
 gctgcaggac atcgtgacct tccgtgcaga aacctcccc tccccctcaa gccgcctccc 1860
 gagcctcctt cctctccagg cccccagtgc ccagtgcga gtgccagcc caggcctcgg 1920
 tcccagagat gccaggagcc a 1941

<210> 28
 <211> 1433
 <212> DNA
 <213> Homo sapiens
 <400> 28

atctatcttt tcaccatctt tgtgactcag cttttctta ttcctttaat tctttgcata 60

gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
 gcatcggctc gcaggcacag atcctaggaa gtctgttctt gtctccctg tgcagggtat 180
 cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
 gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gcactgagct 300
 ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
 cacgcttggg tgggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
 acggtcctgg ttgcaggaat agctgagtca gaggtggggc ttcctccact ccccccccc 480

acccctttca ccattaggga ccttcttgcc ttgctcttgc tactctgctc tgggtggtca 540
 ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
 ggggtccagc aggaaggca atcccagaaa ggcagggtca gggactggag tcctgtgggt 660
 gctttttaag caaagattat caccaggcag gctaaactta gcaaccggct ttagctaga 720
 agggcagggg gctggtgtca ggttatgctg ggccagcaaa gagggccggg atccccctcc 780
 catgcacctg ctgatgggcc aaggccaccc ccccccccc ccttccttac aagtgttcag 840

caccctccca tcccacactc acaaacctgg ccctctgccc tcctaccaga agaatggatc 900

ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc ttttccact 960

ttttccctct tcttgactca ccctgccctc aatatccccc ggcgagcca gtgaaaggga 1020

gtccctggct cciggtcgc ctgcacgtcc cagggcgggg agggacttcc gccctcacgt 1080

cccgtcttcc gcccaggtt ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140

acagcccaca ggtttccagg agtgcctttg tgggaggcct ctgggcccc accagccatc 1200

ctgtctccg ccigggggcc cagcccggag agagccgtg gtgcacacag ggccgggatt 1260

gtctgccta attatcaggt ccaggctaca gggctgcagg acatcgtgac ctccgtgca 1320

gaaacctccc cctccccctc aagccgctc ccgagcctc ttcctctcca gggccccagt 1380

ggccagtgcc cagtgccag cccaggcctc ggtcccagag atgccaggag cca 1433

<210> 29

<211> 17130

<212> DNA

<213> Cricetulus griseus

<400> 29

gtgagtgggt caggggtgact ggcctattg ggttctgtt cagtgaacac agtctgttct 60

tcgagtgtg tgggactctt aacttactgc cgtgtggcc tctactacgc catttaacta 120

gagtgtctt agatcacttt ctatgtccac tccccgtcc taaaaaaaa atttattttg 180

tatacattag tattttgcct gtgtgtatgt ccatgtgaag gtgccagatc tggaacttga 240

gttacaggca gtgtgagct gccacatggg tcttggaac tgaaccagg tcttctgcag 300

gagcagcaag tgccatctc ctagcccca gtaaaagccg atttctagt gagacatggt 360

ctttctgtgt agctctagct ggcttagaat ttgctaccta gaccaggctg gcccacaatt 420

cagattcccc tgctctgcc tctcatgtgc tggtaggaaa ggcgtgagtc acatgcttgg 480

cactagacta gctttgtgt tagagcaaaa ctaagtaagt agagtcccg ttgtccctg 540

tcctctgccc catcatctc atacccttgg gtactgtgat cttacaatct acagtacct 600

ggatgaatga gtgtgagtga caatgaatga cagagcagta cccagagtcc acagtcaca 660

ctgatactgg gttcactgta cagtggcaca tactgtcac cacagtgtca tactggacat 720

tttcatgacc ctaaggccac atcacccctg tgggtgtcact cagtgtgaat ttcacctat 780

atagtattt caggttggct ttttacttag aatataacat ttaaattatg tttcttttt 840

ggttttgagt tttgttttc gagacagggt tttctgtat tgccttggag gttgtgctgg 900

aacttgctct ggagaccagg ctggcttcga actcacagag atcctccac ctctgcctcc 960

cgagtgctgg gattaaaggc gtgcgccacc aatgccccag catctttctg tttctctgtt	1020
tttaaagagg gagataatgt ttccagtgtt gtttttaact ttttcctgtt aggctacat	1080
agatgaattt gagcagaagc ttccggcatg tcataccaga ggcatggatg gaatagagga	1140
gcttgaaact ggccaagaag gcagccagca agccctgtct gccaagaaac cctctgtctg	1200
tatgagagac aagcccacac ggactcccca caaatcatcc ctgtcattta ttggtttttg	1260
tttttaaaat cacatctccc tgccatttct agtctcaagg cgacagcctc tgacttctat	1320
agattcagac aatgactttg tcacacctaa gcctcgacgc accaaccgtc atcgtccaaa	1380
cactcagcaa cgaaagtcca agaagaaagc caaagtgtc ttttcaagt atgagtcag	1440
tgaggatggt atgactcgcc ctcccggcag ctgcttgctt gccttggggt ctccttgggg	1500
cttgcctggt ttagagatca ttcttctggt taatgtgttc cccactgtgc acgttggctc	1560
acctttgcag ctctatgttt taagatccac ttacacat ctttgtgact cagcttttgt	1620
tcctataatt ctttgaaacg aactttcagc ggaaatgacg gaagaagaga ctccaagag	1680
aaccaccccc atccgcagat catctgcccg aagacacagg tcttaggagt tgcctcacct	1740
atcccatcc ctgctgtgca gggatatctta attgattttt aattcatttc ctcccttgta	1800
aaaatgttgc ctcttgttt ttaatatgta aatggctttt gcattcaagg cttttctct	1860
agagcccatc tgactgtgctc ttccactttc caacaaacac ttaaccaatc accatgtttt	1920
ctaagaataa atgtttttat atacttttat tctttgtggt tctgtttcat ctccacctt	1980
agcccagatt caggcagcaa tagagggtta tcaaacaga gagaagatc accagcccag	2040
gcacctctgag gctgtctctg ggtcttcatg ttctctcatt gattcatctt gcacaactga	2100
tgagactcca ctgtagagaa ttctcacttc tcagcctggg ggtactttct acccagagac	2160
ctattttctg catagccttt tgggacacat aaagcttctt tctgcatag aacaccccac	2220
aacagtgtat caggagtgtg caagttgaca aacaatgtc tagacctag gttttcttcc	2280
tctgtgtgcc tcatccaga agagatcatg actcccagga gtcagccttt actatggggt	2340
ctgcaggggc gtccagcccc tcagcggcaa gccatgccc cctccccaa gtccttaatc	2400
tgctgagtca ctiggaacag gagacactga ttctgtgtgc atgacaacag cacattgcca	2460
tagaaatgct cctacctct tacgtgtgtg gtgggggaac agtaatgaca aaccataggc	2520
aggaggcaaa aggggaagac gcacctcaga aacatgtgtt aggttagggc agaactatgg	2580
aggggctcct gagactcttt gatgggaaag ggttaatgct gtcctgaaa cctctgttgg	2640
aaggcagaaa agggacaggg ctgagtcctc gcactgggac catttccatc ctctgcatcc	2700

tgcccccggc tcatggaaag cctgggcatg ggccacacag ctgtcagtct tggctctggg	2760
gccccaaagga ggtagggcaa tcccagaatg gcaaggagcc aggactggat ttgggggtgca	2820
gcccagcctg ctccctgcct tttaagcaaa ggttatcacc aggccagcta aacttagcaa	2880
ttaggtctct cagctaaaag agcagggggc tggctctcaag ttgactgac ctagcaaaga	2940
ggccccagga tccccctgcc cagcacctgt ggctgagctc ccaagccctt cccgagagct	3000
caggatccac ctttccacc ctccctactc ttcagaggag gaacccctt tctcctccc	3060
acttgttggg gggggctggg gccaggctgt tctggcttgg ggtataatac cccctacccc	3120
ttctactttc cctcctctc agacctcacc ctgcctccac gagggcagcc aagaaaggag	3180
agtccttggc tgcagggccca gtaggcacgt cccaggacgg ggagggaactt cgcacctcac	3240
gtccagctct cgcacctggg gctgcagtgg gtgaaagggg cagtgtctcc tagcctgggc	3300
ggtgcaaccc tcaggttccg aggaggaacg ctctgggagg cttctttgcc tctccaacc	3360
caaccacaa ccaggacatt gtctcacc cggggcccca acctagacct taactgagga	3420
acacagaggc cagtttgtaa gtctcaatta tgcaggcat cccgacctgt ggcgtaggga	3480
gcgcccccc aggccgttc cctagcctcc tctggccct cacagcccag gcctctggcc	3540
caagaaatgg aagtgggggt gggggatgga actgcgaatg cgaagggcc cgcaggagg	3600
caaagtgacc cctcccgggc cttttctgt ccgagacttg ttttgcctg tgtcactacc	3660
gaagaaccac gagaagatcc tcaactttt cacagcctt gcataaagg gagagggtcg	3720
gcggtgcagc tgtggcacac acgcacttct gtcaaccgc cccccccg cccccgttc	3780
tgttcttcc caggttctcc ccatttatc ggggcggcaa ctttaggtc cctgggtcct	3840
ggaagtcctt agtacacact cttcgtcctt aagtcctag tctgtattcc ctcggtccta	3900
tcctgtcccc catcaccggg tcacctccc agcgaagcaa tctcagttcc cctcccctc	3960
tcagccccga gccacacgt ttggtgcgtg cacatttcaa aaacgaggcg ggtccaaaga	4020
gagggggtgg ggagggtccg agtgggccag ctactcgcgg ctttacgggt gcacgtagct	4080
caggcctcag cgcccttgag ctgtgactgg atggatgagc ggggcgggag gcggggcgag	4140
cgctctcggc gtccccacc acccagttc ctataaatac ggactgcagc cctccccgt	4200
gctctctgct cctccctgtt ctagagacag ccgcatctt ccgtgcagt ccaggtgaaa	4260
accgagagt gggccgcagg tggccgggga cggtcggaaa cggggaagg gggcgctcag	4320
cccgggactg cgggcgctgg ggcgagctcc actgcccgag cccgggtcc gcattgcaga	4380
ggctggaggg ggaagggatg gggggcgcgg cggggggcac gctgacctct ctttcttctc	4440

tctccctccg cagcctcgct ccggagacgc aatggtgaag gtcggcgtga acgggtgagt	4500
tcgtggctgg gctagggctg ggctccgggt cccgctccgt cgcgtatgca ggtctacccc	4560
accccggggc tctgcgggag cgtggggctg ccggtgggtg gccgcagcac ccaaggagac	4620
ctcaaggtca gcgagccacc tctcccttg cggggatgag cagcgcggag tccnnnnnnn	4680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	4740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	4800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	4860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnaggtcggt	4920
ggcacccggc cgtgcggctg ccgctttagc gcatccatca tctcccaagg gcttccttta	4980
gggtggctgg ccgccccat gttgcaaagc ggaaggaaat gaatgaacca ccgttaggaa	5040
acctcccttc ggccttcctc ctctctagcc cgtgactaac ctccccactc cctccccggg	5100
tggagtcgcc tctgtactgt aagccaggct atgcaaggct tccgtgtctc cgagagagct	5160
ctacctgcc agctgtctca tattattagc ctcaaagcag ccctcaagcc tcatttacct	5220
tgagcatatg atatattttg tagattctct gagaatcgaa gcggacttgg agaggctctgc	5280
ttgtccttct cccagcccaa aggtggtagc tatggcgtag cgccggaggg gggagtgggg	5340
ggggagctga gtcattgttg ttctgaaaag aaaatttcca ccacaaaatg gctccggctg	5400
tagcatcccc ttcccccat aacctctgct tcccatcaca cctgaccca aacctgtag	5460
gccagactgt aaaggtcact aagaggattg agtgtctgag cctcggaacc ctgcccttct	5520
ccccatccca tctctggaa accagatctc ccccgctcca ccctaactctg aggttatatt	5580
tagccggctg acctttcagt atttggggtc tgggcccta cacacatctg ttgtcctgc	5640
tcctgatttt tagctagcaa attcaagtgc ttgcaaatt agagcccagg gattaggggt	5700
tggaagctc agtggtttct tcagtcttct cctttagggg gagggacttg gaggaagcag	5760
gtgggccgac cctgtccta ctattctga cctttaacct tgccctttga gcttgatgat	5820
gctgagtga cgagttcttc ctgtccaggg ggtgtagcct gaagccagc caggctagaa	5880
caaacttccc aggggggtgg ggtagtgaat gccttgtgcc cacacagggg cacactgcca	5940
cctcttggag acttgaaatg actggtgggg gggttgaca aggctttgag cccaatcacc	6000
tcttggacag gaaagtaacc ccactttat ggccctctg taaaagccca gtcaaacctc	6060
atttgtcaa ggaagataga cctcttgggg ctctctaagg ataggggtgt tctatatattg	6120
ggccctgctt ctaagcattc agccagcttt attaaaggaa attcataaca aaacttgaat	6180
ttctgtcttc ttaaatacta atagtgtgct ggatctccat taaaaatgct gtcttgcaca	6240
gtaggctatg gtttctgtgg gctctctaca gctatgggac aactggattc tgttttctga	6300

agggcattgtg tcagctcagt actgactata gacctatgag ttctctgacc ccctaactca	6360
cctttttttt tcttgccca gatttggccg tattggacgc ctggttacca gggctgcctt	6420
cacttctggc aaagtgaag ttgttgccat caatgacccc ttcattgacc tcaactacat	6480
ggtctacatg ttccagtatg actctaccca tggcaagtcc aaaggcacag tcaaggctga	6540
gaatggaaag cttgtcatca acgggaaggc catcaccatc ttccaggagc gagatcccgc	6600
caacatcaaa tggggatgat ctggcgccga gtatgttggt gaatctactg gcgtcttcac	6660
cacatggag aaggctgggg ccacttgaa gggcggggcc aagagggtca tcatctccgc	6720
cccttctgct gatgccccca tgttttgat ggggtgtgaac caagacaagt atgacaactc	6780
cctcaagatt gtcaggtagat gatggcagag ggctgtggca aagtgggcaa gcaggggcaa	6840
ggttacaggt gggcgagcct cctaacctgt ctcttctctt cagcaatgcg tctgcacca	6900
ccaactgctt agccccctg gcccaaggta tccatgacaa ctttggcatt gtggaaggac	6960
tcatggtatg tagttcatct gtttcatcct gccagcagtg ggcgctgtgg tgggggccct	7020
gcaagacctc actccctgcc tctgtgtctt tcagaccacg gtccatgcca tcaactgccac	7080
ccagaagact gtggatggcc cctccgggaa gctgtggcgt gatggccgtg gggctgcccc	7140
gaacatcacc cctgcatcca ctggcgctgc caaggctgtg ggcaaagtca gccagagct	7200
gaacgggaag ctgactggca tggccttcg tgttctacc cccaactgtt ccgttgtgga	7260
tctgacatgt cgctggaga aacctgtatg tctggggtgg gctgagggtt gtctctagt	7320
gtgaggttgg ggcttgagta gtcacattga tttttgcct taataggcca agtatgagga	7380
catcaagaag gtggtgaagc aggcatttga gggccactg aagggcattc tgggctacac	7440
cgaggaccag gtgtctcct gcgacttcaa cagtgactcc cactcttcca cctttgatgc	7500
tggggctggc attgtcttca atgacaactt tgtaaagctc atttcttgt atgtgggac	7560
agaaacagtc ttttaaaagt aacttgggggt ttgtcacccc ttggtgtcta actcaaaaag	7620
taaacccttc ctaactgccc tccaatgtg ttctaggtat gacaatgaat ttggctacag	7680
caacagagtg gtggacctca tggcctacat ggcttccaag gagtaagaag cccaccctgg	7740
accatccacc ccagcaagga ctgagcaag agggaggccc tggtgtctga gcagtcctg	7800
tccaataacc cccacaccga tcatctcct cacagtgtcc atcccagacc cccagaataa	7860
ggaggggctt agggagcctt actctcttga ataccatcaa taaagttcac tgcacccatc	7920
ttccttggcc tticaatgta aggttggggg agggaggctg tgacctagca aggttgggaa	7980
tcctctgtgt cacttttcaa acagggcact agccacatgc cagcccaggt ttcctgtcct	8040
gaacagatga aaattcacct aaggtgtctt ggtgtggga ggagtggggg ttgacaatgg	8100
gaccagtagt atggtcttag aggcctgggc tggactgcat caagtccag ggctgtgtgt	8160

gtgttatctg caaacaagg tcaattgtgt ctggaggcct taggtaaaat tggaggatg	8220
cccaacatag taaaaatgta tcagccaggg gaagtgacta cactgtatct aacctgaaac	8280
agctgagctg taagccagca gctgtcacta tgttcagggtg tggccgctg gttctggggt	8340
ggtcacttgt atccagtttg ttaggaagtg ttgtcattgc ttgttaggaa gacaacacat	8400
ctcagctgg gcagtggtag tgcgtgcctc taatgccagc attcccagca cggaagtggc	8460
agaggcagga ggacagcctg gaataacaac ccagggcaga gccccatct cggaaaagac	8520
aaaaaccaga aagtgcctaa aacattgaca caaagggtgct caaatattcc ttcattgctt	8580
ttagagattc cactgtcagc ttggcatggc ctctagttag acatcctgac ttggtccct	8640
gctttccaag gtcaggagaa tgatagccac agaacgttcc ctgagctgat ggctggagaa	8700
ccgggggtccc tgagccccc cctcacacc catgtgcagg agggagcttc acctttccct	8760
ccgagcagtg tctgccttca ggacaccgct ctgagcccag aactgagtc tttttgtgt	8820
gcatttcccc ggggaaaagc tacacattcc ctgtacagca ggcaggaggg cagctgtgag	8880
ctcatggaca cgggacagaa ggccaggtag cctgcctccc tgcggtaccc cgcctccctt	8940
cgaggcctta ctaggaagca aagccttctc cagggtgtacg tgatgcggaa agtctgcaga	9000
aagtcgggga ctcccagatg agtctgctgc acagtgcctg ggccagttag caccagaaca	9060
aaagtccaag atagccaagg agaacaggca ggtcccacaa tcacattcct taaaaacgga	9120
gaaggaggga tcacaagtg tctggcagcg tttcaaagat agagaagtag caaacagtca	9180
gagtgaagaa ggcaggagtg tccaatgtc atggggatat ccaggaaggg tggggccaag	9240
cctagctcca gacaccagg gagccgggc gaggagacag cccagggcc aggcctcatg	9300
cagtggcaca aaagacccc accccggcct cggtagctc caggagacca ggcactggtc	9360
cctatctcat ggagctatca gatgagacat ggcaatcgga gtcctcagtt tcgcttggcg	9420
ctggcggcgg gtcgctaagc gggaccgcag tgaaagcagg agactttcta gaaaaaaca	9480
ccagatgtca ccttggggca agcaggaagc ccaagcaggc ctgcatgctc gctctcctt	9540
ccatggcctt atcttctggg gtgcagctgg atccaggca ggggaagtggg tgggaagact	9600
tgtccccatt cacctctatg aggaacagtt cccctcttgg aggcctctgag aggcctcagag	9660
atcaaaaatg tctcaagtc cagcatggag atgcttcttg agccacagt gtgctgacac	9720
cagggccagc tcttgtcca gctgtaaagg gggtagggct gtctcactat gttccaagg	9780
ctgaccccca aacttggtg caaagcaatc ctcttgctcc agagtagctg ggatgacaag	9840
caaaggccac taaacctagt gaacatgctg cacactgtgc ctagcttaat gtaattgggg	9900

gagtgggttg gggttttgtt ttgagacagg gtctcagtat gcattcctgg ctggcctgga 9960

attcacagag gagagaccgg cctggtactt gtggcaatct gtagggagac accgtaacca 10020

cgcctcctaa gagctggctc atcaggggtgt ggtcaagggt atgtcagggt gacatgggag 10080

aggtgtaaag ggataaggaa cgcacatggt agctcttttc cctgctgctt tggcttgctc 10140

tgctttcctg ccggttggtc acctaatata tatatcctga ccatcacggg ctatgtatta 10200

gtaattaatc ctaatagccg gtgtctccct acaacaatcc acctgcctgt gacacctatg 10260

tgcttggatg acagatgtac accaccacac ctggcttact gtggttcctt agattttacc 10320

tgttgttgcg ggttgtttta ttttaagaca agtttttgcg attgtagtcc aggatagcct 10380

caaactcacg gcagtccttt gcttcagcct cccaaggctg ggcttgaga catgtacat 10440

catgcctggg ttacacattt ttaatatata atttattttt attctgtgtt aattggtatt 10500

tgcttgata tatgtctgtg taaaggtatc agaagccctg gaactggagt tacagacact 10560

cgtgtgggtg ctgggaattg aaccaggcc ctctggaaga gcatccagtg tcttaacta 10620

atgagccatc tctcagcccc caggtttttt tttttttttt tttactctgt gtgtatgtgt 10680

gtgtgtgcac acgcacgcgc atgctcgac ataccaccgc acacagggtc tcacacaggc 10740

tagacaagta ctctaccact gagctatatt ccccaaacc aaatttactc ttacacaagg 10800

tcaaatactt aaatcatgac ttagactttg gataaaaacc tgccctgtg tctcttcatt 10860

cctagccagg ttccggctct cagacactgg ggactggta ggagtcgtct gaccaccagg 10920

gcaggagag gaagctatca atctgggcca cacctatagg tgactccaaa ggacagctcc 10980

acttgcatg gtgacctgt caccggccac tgggtggga cgagggcaa tcagagcagg 11040

tccttaccgg tccccgact gcgtgatcag ggaggtcct taccgggtccc ccgactgtgt 11100

gatgaggcgg cggcaggtct ccatctgcac gtccaagccc cgtttcatgc tgcacatctc 11160

catgtactcg tgcagggtgc ggttcatgtc attcttggct gtggccagct ccagctgcag 11220

aaaagtaggg cagggcagga tgaggaggtt agcaagcca aggcagacag gggattagag 11280

cccagctggg ccccgaggc caggcagcca tggttaaaag ctgcgccttc aaaagaaagc 11340

aagcagnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11400

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11460

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11520

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11580

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11640

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11700

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11760

11820	11820	11820	11820	11820	11820	11820
11880	11880	11880	11880	11880	11880	11880
11940	11940	11940	11940	11940	11940	11940
12000	12000	12000	12000	12000	12000	12000
12060	12060	12060	12060	12060	12060	12060
12120	12120	12120	12120	12120	12120	12120
12180	12180	12180	12180	12180	12180	12180
12240	12240	12240	12240	12240	12240	12240
12300	12300	12300	12300	12300	12300	12300
12360	12360	12360	12360	12360	12360	12360
12420	12420	12420	12420	12420	12420	12420
12480	12480	12480	12480	12480	12480	12480
12540	12540	12540	12540	12540	12540	12540
12600	12600	12600	12600	12600	12600	12600
12660	12660	12660	12660	12660	12660	12660
12720	12720	12720	12720	12720	12720	12720
12780	12780	12780	12780	12780	12780	12780
12840	12840	12840	12840	12840	12840	12840
12900	12900	12900	12900	12900	12900	12900
12960	12960	12960	12960	12960	12960	12960
13020	13020	13020	13020	13020	13020	13020
13080	13080	13080	13080	13080	13080	13080
13140	13140	13140	13140	13140	13140	13140
13200	13200	13200	13200	13200	13200	13200
13260	13260	13260	13260	13260	13260	13260
13320	13320	13320	13320	13320	13320	13320
13380	13380	13380	13380	13380	13380	13380
13440	13440	13440	13440	13440	13440	13440
13500	13500	13500	13500	13500	13500	13500
13560	13560	13560	13560	13560	13560	13560
13620	13620	13620	13620	13620	13620	13620

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	13920
nnnnnnnnnn nnnnnnnnnnac acacacacac acacacacac acacacttac ttcctgagga	13980
agactgcagg gaacaaggct agatagaacc cattattcca ggagacaaaa gtgaagcaaa	14040
aatggctctct taaagcacia agcagcacta ccaacagtgg tctcttcag cctatatatc	14100
accctagttt acacattgaa gggaaagatg tacattctag aacattttca acaagtcagc	14160
ccccaaattc cagactcagc atgggagatc agagcagtag aggtagatgg aggagataga	14220
ggggccaggg ccatggagga aaactccac ctgtcatcag ctttaccatc caataatggg	14280
cagctctacc tcagtaacac tgagagatta agggccatac agacaaaagg aacttttgcc	14340
acaaagctac caagtgaact ttctttacct ttctctcaa gtcattcttc caggaaggac	14400
agaagtcaca caacttagaa acaactttcc atccatgtgc agtatacata tttttatgtg	14460
tgtggataca gatgtgcatg tgtgtgggta catggggaag cctgatgtcg atgcagggaa	14520
tcttcctagg tggttcttat actctgagtc tgtgtaacgg cattggttgt cctggaactt	14580
gatctgtaga ccagactggc ctcaaactca cagagatcta cctgtctctg cctcccgaat	14640
gctgggatta aaagcgtgtg cctcagccgg acgttggtgg cgcatgcctt taatcccagc	14700
actcgggagg cagaggcagg aggatctctg tgagttctag accagcctgg tctacannnn	14760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	14820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ctagttccag	14880
ggcagcctcc aaagccacag agaaaccctg tctcaaaaaa caaaacaaaa ataaataaat	14940
aaataaataa aatcctgcgc ctccactgcc cggtcagaa tatitttaaa gacaaggatt	15000
atgtagccca ggctggcctc aaacttgta ttagctaaag atgatggtgg cctctgatcc	15060
tcctgcctct acctcccaag tgtaggact ataggctggc actaccacac cgggtttcat	15120
gtgggctagg atcacactca ggcagggtct catgcacatt ggcaggcat tctaccaact	15180
gagttacaca gtcccagccc aggatgtctc accagatcac gggttacagct aggcacctgt	15240
ggtctgaggt ttgaccaca tggcagcccc agtcactgag ctaagggtgaa actcatactt	15300
gactaaaggc acggaatctg aaaggggtgg gtacaacca ggaggaaaat cggcggaagc	15360

aataagaggc gtcaacaggc aatggctact tttccccaga caagtctcag ctggaagcag 15420
cacacgggct gggtatctga aactcagaag tttcctgggt taaaaacaaa agccctgggt 15480
tttcttagga tagtatgttc atttgatgtg tgcctgggaca ctgagaacac aggaaaataa 15540
cttccaaagt ttgtttcctc aatgcagaca gggtgaacct ttacagggct tgcaggagg 15600
ctgatacaag aggtctcctg cctcattcac ccatcagagt ccaagacaca gtcctcacac 15660
agccccccc aaccaactgt cctccgacgg tgacctctc cctacaaacc tgcctgcagc 15720
ccagaaggac agtcagagcc ccagctccat agccatttgt aaaggactcc ccaggctcaa 15780

aaacaaaagg gagactctc caggagctca caggctgagc caaggaggca ggagtgtcc 15840
ggcacttagg gtgtccgggt gataatgcac aatatgtatg atgtgtcaac ctgccagatc 15900
ctgtcctaag cgccacgtgt ctatctgatc atcctaataa cccagcctgc ggggagagac 15960
atgcacatct cttctgaca aataaaggagg caggcccaag cttagcatgg cagtacacat 16020
ctgtaaccca gcactcaaca ggcagagcag gaggattgac tgcctcctat tccaggccag 16080
tttggtaca cagaaagacc ctgtactgat gctaagagag atggctcaga gtttaagagca 16140
cttctgtctc ctggagagga cccaggtttg ctctccagca cccatgtcag gctcacagtc 16200

gcctgtgaac accagctcca ggggatctga tgccctcttc tggcctccag gggcacgcgc 16260
gtgtgcgtc acacacacac acacacacac acacacacac acacacacac acacacacac 16320
acagagtga aagctgaagt gcagctcaca ccataatcc tagcaaattg agacaaaagg 16380
accaggagtt cgaggccaac ctagactaca agacaagtac ttgtaataa atacatatat 16440
acgaaaataa acaaaccaaa aataatgaaa caggcactca gatcaaggga gccagtattg 16500
gtggcacacg ctttaatcc cagcactcag gaggcagagg caggtggatc tctgagtttg 16560
aggccagcct agtctacaga atgagatcca ggactacaca gagaaccct gtctaccct 16620

cctctccctt aaaaaaatt aagaaagaaa agaaagagcc aggggctggt aagcaatgga 16680
tcagcaggta aaggcacaag ccaccaagcc caacaaccta aatttgatcc ctgggacct 16740
catggtagaa ggagagaact cactgcaagc tgcctctga cctccacaca cacaccatgg 16800
caggcacatg ctctccccgc cacatacata catacataca tacatacata catacataca 16860
tacatacatg taacagaaa agaaagtagg aagacaaaag gttcattcta agtggtgaca 16920
gagccaagat ttgaaccac atctatctgc gtctagatcc tgaggctcga agggccatt 16980
agagctacca aaactggggt ggggtgcaact cctcctgtca acaggtggaa ggtcctggtc 17040

ttgggctctg gctgggagga ggccagccag cagcacacct ccagcctggt ttgtgagccc 17100
tggaacaga gccctgcca aacccttac 17130

<211> 3020

<212> DNA

<213> Artificial Sequence

<220><223> amplicon

<400> 30

tacggccggc ttactgtac agtggcacat actgtcacc acagtgtcat actggacatt	60
ttcatgacct taaggccaca tcacccctgt ggtgtcactc agttgtaatt tcacctata	120
tagtgatttc aggttggctt ttactttaga atataacatt taaattatgt ttcttttttg	180
gttttgagtt ttgtttttcg agacagggtt tctctgtatt gctttggagg ttgtgctgga	240
acttgctctg gagaccaggc tggcttcgaa ctacacagaga tcctcccacc tctgcctccc	300
gagtgtctggg attaaaggcg tgcgccacca atgcccagc atctttctgt ttctctgttt	360
ttaaagaggg agataatgtt tccagtgtcg tttttaactt ttttcctgta ggctaccata	420
gatgaatttg agcagaagct tcgggcatgt cataccagag gcatggatgg aatagaggag	480
cttgaaactg gccagaagg cagccagcaa gccctgtctg ccaagaaacc ctctgtctgt	540
atgagagaca agcccacag gactccccac aaatcatccc tgtcatttat tggttttgt	600
ttttaaaatc acatctccct gccatttcta gtctcaaggc gacagcctct gacttctata	660
gattcagaca atgactttgt cacacctaa cctcgacgca ccaaccgtca tcgtccaaac	720
actcagcaac gaaagtccaa gaagaaagcc aaagtgtct tttcaagtga tgagtccagt	780
gaggatggta tgactcgccc tcccggcagc tgcttgcttg ccttggggtc tccttggggc	840
ttgcctgggt gtagagtcat tcttcttgtt aatgtgttcc cactgtgca cgttggctca	900
cctttgcagc tctatgtttt aagatccact ttacaccatc tttgtgactc agcttttgtt	960
cctataattc ttgaaacaga actttcagcg gaaatgacgg aagaagagac tccaagaga	1020
accaccccca tccgcagatc atctgcccga agacacaggt cttaggagtt gcctcaccta	1080
ttcccatccc tgctgtgcag ggtatcttaa ttgattttta attcatttcc tcccttgtaa	1140
aaatgttgcc tecttgtttt taatatgtaa atggctcttg cattcaaggc ctttctccta	1200
gagcccatct gactgtgctt ttactttcc aacaaacact taaccaatca ccatgttttc	1260
taagaataaa tgtttttata tacttttatt ctttgtctgt ctgtttcacc ttccacctta	1320
gccagattc aggcagcaat agagggttat caaaccagag agaagataca ccagcccagg	1380
catcctgagg ctgtctctgg gtcttcatgt ttctcattg attcatcttg cacaactgat	1440
gagactccac tgtagagaat tctcacttct cagcctgggg gtactttcta cccagagacc	1500
tatttctgtc atagccittt gggacacata aagcttcctt cctgcataga acacccaca	1560

acagtgtatc aggagtgtac aagttgacaa acaatgtctt agacctacgg ttttcttcct 1620
ctgtgtgcct catcccgaa gagatcatga ctcccaggag tcagccttta ctatggggtc 1680
tgcaggggcg tccagccct cagcggcaag ccatgcccac cctccccaag tccttaatct 1740
gtgagtcac ttggaacagg agacactgat tctgtgtca tgacaacagc acattgccat 1800
agaaatgtct cctacctctt acgtgtgtgg tgggggaaca gtaatgacaa accataggca 1860
ggaggcaaaa gggaagacgg cacctcagaa acatgtgtta ggtagggca gaactatgga 1920

ggggctcctg agactctttg atgggaaagg gttaatgtct ctcctgaaac ctctgttgga 1980
aggcagaaaa gggacagggc tgagtcctcg cactgggacc atttccatcc tctgcactct 2040
gccccggct catggaaagc ctgggcatgg gccacacagc tgtcagtctt ggctctgggg 2100
ccccaaggag gtagggcaat ccagaaatgg caaggagcca ggactggatt tggggtgcag 2160
cccagctgc tccctgcctt ttaagcaaag gttatcacca ggccagctaa acttagcaat 2220
taggctcttc agctaaaaga gcagggggct ggtctcaagt tgactgacc tagcaaagag 2280
gccccaggat cccctgccc agcacctgtg gctgagctcc caagcccttc ccgagagctc 2340

aggatccacc ctttccacc tccctactct tcagaggagg aacccccctt ctccttccca 2400
cttgttgag ggggctgggg ccaggctgtt ctggcttggg gtataatacc cctaccct 2460
tctactttcc cctcctctca gacctacc cgcctccagc agggcagcca agaaaggaga 2520
gtccctggct gcagggccag taggcacgtc ccaggacggg gagggacttc cgcctcacg 2580
tccagctctc cgcctgggg ctgcagtggg tgaaaggggc agtgtctcct agcctgggcg 2640
gtgcaacct caggttccga ggaggaacgc tctgggagge tttttgcct cctccaacc 2700
aaccacaac caggacattg tctcacc cggggcccaa cctagacctt aactgaggaa 2760

cacagaggcc agtttgtaag tctcaattat gcagggcatc ccgacctgtg gcgtagggag 2820
cgccctcca ggccgttcc ctagectct cctggccctc acagcccagg cctctggccc 2880
aagaaatgga agtgggggtg ggggatggaa ctgcgaatgc gaaggcccc cgcaggagc 2940
aaagtgacc ctcggggcc ttttctgtc cgagacttgt ttttgcctgt gtcactaccg 3000
aagaaccacg gccggcctga 3020

<210> 31

<211> 3537

<212> DNA

<213> Artificial Sequence

<220><223> amplicon

<400> 31

tactcgcgat tcactgtaca gtggcacata cttgtcacca cagtgtcata ctggacattt	60
tcatgacct aaggccacat caccctgtg gtgtcactca gttgtaattt caccctatat	120
agtgatttca ggttggtttt ttacttagaa tataacattt aaattatgtt tcttttttgg	180
ttttgagttt tgtttttcga gacagggttt cctgtattg ctttggaggt tgtgctggaa	240
cttgctctgg agaccaggtt ggcttcgaac tcacagagat cctccacct ctgcctccc	300
agtgctggga ttaaaggcgt gcgccaccaa tgccccagca tctttctgtt tctctgtttt	360
taaagaggga gataatgttt ccagtgtgtt ttttaacttt tttcctgtag gctaccatag	420
atgaatttga gcagaagctt cgggcatgtc ataccagagg catggatgga atagaggagc	480
ttgaaactgg ccaagaaggc agccagcaag cctgtctgc caagaaacct tctgctggta	540
tgagagacaa gccacacgg actccccaca aatcatcctt gtcatttatt ggtttttgtt	600
tttaaaaatca catctcctg ccatttctag tctcaaggcg acagcctctg acttctatag	660
attcagacaa tgactttgtc acacctaagc ctgcagcac caaccgtcat cgtccaaaca	720
ctcagcaacg aaagtccaag aagaaagcca aagtgtctt ttcaagtgat gagtccagt	780
aggatggtat gactgcctt cccgcagct gcttgcttgc cttgggtct cttggggct	840
tcctggttg tagagtcat cttcttggtt atgtgttccc cactgtgcac gttggctcac	900
ctttgcagct ctatgtttta agatccactt tacaccatct ttgtgactca gcttttgtt	960
ctataattct ttgaacagaa ctttcagcgg aaatgacgga agaagagact cccaagagaa	1020
ccaccccat ccgcagatca tctgcccga gacacaggtc ttaggagttg cctcacctat	1080
tccatccct gctgtgcagg gtatcttaat tgatttttaa ttcatttctt cctttgtaa	1140
aatgttgctt cctgtttttt aatatgtaa tggcttttgc attcaaggcc tttctcctag	1200
agcccatctg actgtgtctt tcactttcca acaaacactt aaccaatcac catgttttct	1260
aagaataaat gtttttatat acttttattt tttgtgtgtc tgtttcatct tccaccttag	1320
cccagattca ggcagcaata gagggttatc aaaccagaga gaagatacac cagcccaggc	1380
atcctgaggc tgtctctggg tcttcatgtt tctcattga ttcatttgc acaactgatg	1440
agactccact gtagagaatt ctactttctc agcctggggg tactttctac ccagagacct	1500
atttctgtca tagccttttg ggacacataa agcttccttc ctgcatagaa caccacaaa	1560
cagtgtatca ggagtgtaca agttgacaaa caatgtctta gacctacgtt tttcttctc	1620
tgtgtgcctc atcccagaag agatcatgac tcccaggagt cagcctttac tatgggtctt	1680
gcaggggctt ccagcccctc agcggcaagc catgccacc ctcccaagt ccttaatctg	1740

ctgagtcact tggaaacagga gacactgatt ctgctgtcat gacaacagca cattgccata	1800
gaaatgctcc ctacctctta cgtgtgtggt gggggaacag taatgacaaa ccataggcag	1860
gaggcaaaag ggaagacggc acctcagaaa catgtgttag gttagggcag aactatggag	1920
gggtcctga gactctttga tgggaaaggg ttaatgctgc tctgaaacc tctgttgaa	1980
ggcagaaaag ggacagggtc gagtccccgc actgggacca tttccatcct ctgcatcctg	2040
ccccggctc atggaaagcc tgggcatggg ccacacagct gtcagtcttg gctctggggc	2100
ccaaggagg tagggcaatc ccagaatggc aaggagccag gactggattt ggggtgcagc	2160
ccagcctgct cctgccttt taagcaaagg ttatcaccag gccagctaaa cttagcaatt	2220
aggctcttca gctaaaagag cagggggctg gtctcaagtt gcactgacct agcaaagagg	2280
ccccaggatc cccctgccca gcacctgtgg ctgagctccc aagcccttcc cgagagctca	2340
ggatccaccc tttccacct cctactctt cagaggagga acccccttcc tcttccac	2400
ttgttgagg gggctggggc caggctgttc tggcttgggg tataataccc cctaccctt	2460
ctactttccc ctctctcag acctcacct gcctccacga gggcagccaa gaaaggagag	2520
tccttggtg cagggccagt aggcacgtcc caggacgggg agggacttcc gccctcacgt	2580
ccagctctcc gccctggggc tgcagtgggt gaaaggggca gtgtctccta gcctgggcgg	2640
tgcaacctc aggttccgag gaggaacgt ctgggaggt tctttgctc ctccaacca	2700
accacaacc aggacattgt cctaccccg gggccccaac ctagacctta actgaggaac	2760
acagaggcca gtttgaagt ctcaattatg cagggcaccc cgacctgtgg cgtaggagc	2820
gccccccag gccgttccc tagcctctc ctggccctca cagcccaggc ctctggccca	2880
agaaatggaa gtgggggtgg gggatggaac tgcgaatgcg aaggggcccc gcaggaggca	2940
aagtgacccc tccgggctc tttctgtcc gagacttgt tttgcctgtg tctactaccga	3000
agaaccacga gaagatctc aacttttcca cagcctttgc ataaagggga gagggtcggc	3060
ggtgcagctg tggcacacac gcacttctgc tcaaccgcc cccccgcc cccgttctg	3120
ttcttccca ggttctccc attttatcgg ggcggcaact tttaggtccc tgggtcctgg	3180
aagtcttag tacacactc tcgtcttaa gtccatagtc tgtattcct cggctctatc	3240
ctgtcccca tcaccgggtc acctccccag cgaagcaatc tcagttccc tccccctc	3300
agccccgagc ccacagtgt ggtgctgca catttcaaaa acgaggcggg tccaaagaga	3360
gggggtgggg aggtgccgag tggccagct actcgcggt ttacgggtgc acgtagctca	3420
ggcctcagcg ccttgagct gtgactggat ggatgagcgg ggcgggaggc ggggcgagcg	3480
tcctcggcgc tccccaccac ccagttcct ataaatagca gagtcgttt agtgaac	3537

<210> 32

<211> 537
 <212> DNA
 <213> Artificial Sequence
 <220><223> amplicon
 <400> 32
 tactcgcgaa gaagatcctc aacttttcca cagcctttgc ataaagggga gagggtcggc 60
 ggtgcagctg tggcacacac gcacttctgc tcaaccgcc ccccccgcc cccgttctg 120
 ttccttccca ggttctcccc attttatcgg ggcggcaact tttaggtccc tgggtcctgg 180
 aagtccttag tacacactct tcgtccttaa gtccatagtc tgtattccct cggtcctatc 240

 ctgtccccc tcaccgggtc acctccccag cgaagcaatc tcagttcccc tccccctctc 300
 agccccgagc ccacacgttt ggtgcgtgca catTTcaaaa acgaggcggg tccaaagaga 360
 ggggggtggg aggtgccgag tggcccagct actcgcggt ttacgggtgc acgtagctca 420
 ggcctcagcg cccttgagct gtgactggat ggatgagcgg ggcgggaggc ggggcgagcg 480
 tcctcggcgc tccccaccac ccagttcct ataaatagca gagtcgttt agtgaac 537
 <210> 33
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer GlnPr1896
 <400> 33

 tacggccggc ttactgtac agtggcacat 30
 <210> 34
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer GlnPr1897
 <400> 34
 tcaggccggc cgtggttctt cggtagtgac 30
 <210> 35
 <211> 35
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer GlnPr1901

<400> 35
tactcgcgaa gaagatcctc aacttttcca cagcc 35

<210> 36
<211> 40
<212> DNA

<213> Artificial Sequence
<220><223> Primer GlnPr1902

<400> 36
gttcactaaa cgagctctgc tatttatagg aactggggtg 40

<210> 37
<211> 40
<212> DNA
<213> Artificial Sequence
<220><223> Primer GlnPr1903

<400> 37
caccagtt cctataaata gcagagctcg ttagtgaac 40

<210> 38
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Primer GlnPr1904

<400> 38
cgctagcacc ggtcgatcga 20

<210> 39
<211> 31
<212> DNA
<213> Artificial Sequence
<220><223> Primer GlnPr1905

<400> 39
tactcgcgat tcactgtaca gtggcacata c 31