



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0089314
(43) 공개일자 2020년07월24일

(51) 국제특허분류(Int. Cl.)
C12N 15/81 (2006.01) C12N 15/67 (2006.01)
C12P 21/00 (2006.01)
(52) CPC특허분류
C12N 15/81 (2013.01)
C12N 15/67 (2013.01)
(21) 출원번호 10-2020-7018326
(22) 출원일자(국제) 2017년11월30일
심사청구일자 2020년06월24일
(85) 번역문제출일자 2020년06월24일
(86) 국제출원번호 PCT/CN2017/113941
(87) 국제공개번호 WO 2019/100431
국제공개일자 2019년05월31일
(30) 우선권주장
201711194355.9 2017년11월24일 중국(CN)

(71) 출원인
강마-헬스코드(상하이) 바이오테크 컴퍼니 리미티드
중국 상하이 201321 푸둥 뉴 에어리어 푸룽후아 로드 118 빌딩 12
(72) 발명자
구오, 민
중국 상하이 201321 푸둥 뉴 에어리어 푸룽후아 로드 118 빌딩 12
쥔우, 지지안
중국 상하이 201321 푸둥 뉴 에어리어 푸룽후아 로드 118 빌딩 12
유, 쉬에
중국 상하이 201321 푸둥 뉴 에어리어 푸룽후아 로드 118 빌딩 12
(74) 대리인
특허법인다울

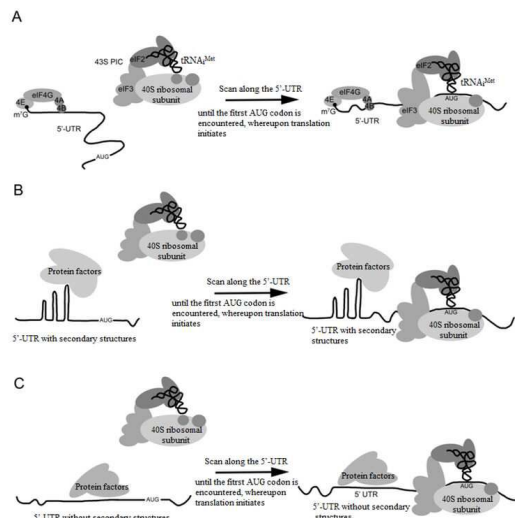
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 단백질 합성 효율을 향상시킬 수 있는 탠덤 DNA 요소

(57) 요약

단백질 합성 효율을 향상시킬 수 있는 탠덤 DNA 요소, 특히, 핵산 구조체는 탠덤의 진핵 세포(예컨대, 효모)로부터 기원한 IRES인핸서(예컨대, ScBOI1, ScFLO8, ScNCE102, ScMSN1, KIFLO8, KINCE102, KIMSN1, KIBOI1), Ω 서열 및 효모-특이적 Kozak 서열에 의해 형성된다. 효모 시험관 내 생합성 시스템(예컨대, 효모 시험관 내 단백질 합성 시스템)에서 핵산 구조체의 사용은 단백질 합성 효율을 현저히 개선시킬 수 있다.

대표도 - 도1



(52) CPC특허분류

C12P 21/00 (2013.01)

명세서

청구범위

청구항 1

식 I의 핵산 서열을 포함하는 핵산 구조체;

Z1-Z2-Z3-Z4-Z5 (I)

상기에서,

Z1~Z5은 각각 구조체의 일부인 요소이고;

각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;

Z1은 IRES 요소를 포함하는 인핸서 요소이고;

Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

Z3은 아데닌 데옥시뉴클레오타이드(adenine deoxynucleotide)의 올리고머 사슬[oligo(A)]_n 이고;

Z4는 번역 개시 코돈이며;

Z5는 세린 코돈이고;

상기 Z3, Z4 및 Z5는Kozak 서열을 구성(constitute)하고, Kozak 서열은 효모로부터 유래된 것임.

청구항 2

제1항에 있어서, 상기 IRES 요소는 ScBOI1, ScFLO8, ScNCE102, ScMSN1, KIFLO8, KINCE102, KIMSN1, KIBOI1 및 이들의 조합으로 구성된 군으로부터 선택되는 것인, 핵산 구조체.

청구항 3

제1항에 있어서, 상기 핵산 구조체의 핵산 서열은 서열번호 2-17인 것인, 핵산 구조체.

청구항 4

5'에서 3'의 화학식 II의 구조를 포함하는 핵산 구조체:

Z1-Z2 Z3-Z4-Z5-Z6 (II)

상기에서,

Z1~Z6은 각각 핵산 구조체의 일부인 요소이고;

각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;

Z1은 IRES 요소를 포함하는 인핸서 요소이며;

Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

Z3은 아데닌 데옥시뉴클레오타이드의 올리고머 사슬[oligo(A)]_n 이고;

Z4는 번역 개시 코돈이며;

Z5는 세린 코돈이고;

Z6은 외래성 단백질의 암호화 서열이며;

상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것임.

청구항 5

제4항에 있어서, 상기 외래성 단백질의 암호화 서열은 루시페린(luciferin), 또는 루시페라아제(luciferases) (예: 반딧불이 루시페라아제), 녹색형광단백질(green fluorescent protein), 황색형광단백질(yellow fluorescent protein), 아미노아실-tRNA합성효소(aminoacyl-tRNA synthetase), 글리세르알데히드-3-인산수소이탈효소(glyceraldehyde-3-phosphate dehydrogenase), 카탈라아제(catalase), 액틴(actin), 항체의 가변영역, 루시페라아제 돌연변이(luciferase mutants), α -아밀라아제(α -amylase), 엔테로신 A(enterocin A), C형 간염 바이러스 E2 당단백질(hepatitis C virus (HCV) E2 glycoprotein), 인슐린 전구체(insulin precursors), 인터페론 α A(interferon α A), 인터루킨-1 β (interleukin-1 β), 라이소자임(lysozyme), 혈청 알부민(serum albumins), 항체의 단쇄가변분절(single-chain variable fragment)(scFv), 트란스타이레틴(transthyretin), 티로시나아제(tyrosinase), 자일라나아제(xylanase) 및 이들의 조합으로 구성된 군으로부터 선택된 외래성 단백질을 코딩하는 것인, 핵산 구조체.

청구항 6

제4항에 있어서, 상기 외래성 단백질은 루시페린, 또는 루시페라아제 (예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 액틴, 항체의 가변영역, 루시페라아제 돌연변이, α -아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 α A, 인터루킨-1 β , 라이소자임, 혈청 알부민, 항체의 단쇄가변분절 (scFv), 트란스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합으로 구성된 군으로부터 선택되는 것인, 핵산 구조체.

청구항 7

5'에서 3'의 화학식 III의 구조를 포함하는 핵산 구조체:

Z0-Z1-Z2 Z3-Z4-Z5-Z6 (III)

상기에서,

Z0~Z6은 각각 핵산 구조체의 일부인 요소이고;

각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;

Z0은 T7 프로모터, T3 프로모터, SP6 프로모터 및 이들의 조합으로 구성된 군으로부터 선택되는 프로모터 요소이고;

Z1은 IRES 요소를 포함하는 인핸서 요소이며;

Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

Z3은 아데닌 데옥시뉴클레오타이드의 올리고머 사슬[oligo(A)]_n 이고;

Z4는 번역 개시 코돈이며;

Z5는 세린 코돈이고;

Z6은 외래성 단백질의 암호화 서열이며;

상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것임.

청구항 8

5'에서 3'의 화학식 IV의 구조를 포함하는 핵산 구조체:

Z0'-Z1-Z2 Z3-Z4-Z5-Z6 (IV)

상기에서,

Z0'~Z6은 각각 핵산 구조체의 일부인 요소이고;

각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;

Z0'은 GAA이고;

Z1은 IRES 요소를 포함하는 인헨서 요소이며;

Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

Z3은 아데닌 데옥시뉴클레오타이드의 올리고머 사슬[oligo(A)]_n 이고;

Z4는 번역 개시 코돈이며;

Z5는 세린 코돈이고;

Z6은 외래성 단백질의 암호화 서열이며;

상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것임.

청구항 9

청구항 제1항 내지 제8항에 따른 핵산 구조체를 포함하는 벡터 또는 벡터 조합물.

청구항 10

게놈이 하나 이상의 지점에서 청구항 제1항 내지 제8항에 따른 핵산 구조체와 통합되거나 청구항 제9항에 따른 벡터 또는 벡터 조합물을 포함하는 유전자 변형 세포.

청구항 11

다음 군으로부터 선택되는 하나 이상의 시약을 포함하는 키트:

- (a) 청구항 제1항 내지 제8항에 따른 핵산 구조체
- (b) 청구항 제9항에 따른 벡터 또는 벡터 조합물; 및
- (c) 청구항 제10항에 따른 유전자 변형 세포.

청구항 12

시험관 내 고수율 단백질 합성에 적용 가능한 청구항 제1항 내지 제8항에 따른 구조체, 청구항 제9항에 따른 벡터 또는 벡터 조합물, 청구항 제10항에 따른 유전자 변형 세포 또는 제11항에 따른 키트의 용도.

청구항 13

다음 단계를 포함하는 외래성 단백질을 시험관 내에서 고수율로 합성하는 방법:

- (i) 진핵 생물-기반 시험관 내 생합성 시스템의 존재하에, 청구항 제1항 내지 제8항에 따른 핵산 구조체를 제공 하는 단계: 및

(ii) 적합한 조건 하에서, 단계 (i)의 진행 생물-기반 시험관 내 생합성 시스템을 T1 기간 동안 인큐베이션하여 외래성 단백질을 합성하는 단계.

청구항 14

제13항에 있어서, 상기 방법은 (iii) 선택적으로, 진행 생물-기반 시험관 내 생합성 시스템으로부터 외래성 단백질을 분리하거나; 또는

진행 생물-기반 시험관 내 생합성 시스템으로부터 외래성 단백질을 검출하는 것을 더 포함하는 것인, 외래성 단백질을 시험관 내에서 고수율로 합성하는 방법.

청구항 15

제13항에 있어서, 상기 진행 생물-기반 시험관 내 생합성 시스템은 효모-기반 시험관 내 생합성 시스템인 것인, 외래성 단백질을 시험관 내에서 고수율로 합성하는 방법.

발명의 설명

기술 분야

[0001] 본 발명은 생명공학 분야와 관련된 것으로, 더 구체적으로, 단백질 합성 효율을 향상시킬 수 있는 탠덤(tandem) DNA 요소에 관한 것이다.

배경 기술

[0002] 단백질은 세포에서 중요한 분자이며, 세포의 거의 모든 기능에 관여한다. 단백질의 서열 및 구조는 상기 기능을 결정한다. 세포에서, 단백질은 다양한 생화학 반응을 촉매하는 효소로서 작용할 수 있으며, 유기체의 다양한 활동을 조절(coordinate)하기 위한 신호 분자로서 작용할 수 있다. 단백질은 생물 형태를 지지할 수 있고, 에너지를 저장할 수 있으며, 분자를 운반할 수 있고, 유기체가 움직일 수 있게 한다. 생물의약품분야에서, 표적 약물인 항체(일종의 단백질)는 암 및 다른 질병을 치료하기 위한 중요한 수단이다.

[0003] 세포에서, 단백질 번역의 조절은 영양 결핍과 같은 외부 압력에 반응할 뿐만 아니라 세포 발달 및 분화와 같은 다양한 과정에서 중요한 역할을 한다. 단백질 번역의 4 단계 과정은 번역 개시, 번역 신장(번역 연장), 번역 종결 및 리보솜 재순환을 포함하고, 그 중에서 번역 개시는 가장 규제된 과정이다. 진행 세포에서 번역 개시는 두 가지 방법이 있다: 전형적인 캡-의존적(cap-dependent) 방법 및 캡-비의존적(cap-independent) 방법(도 1에 도시).

[0004] “캡(cap) 구조” 의존적 번역 개시는 수십 개의 번역 개시 인자 및 리보솜의 40S소단위체가 관련된 매우 복잡한 과정이다. “캡 구조” 비의존적 번역 개시는 주로 mRNA의 5'-비번역부위(5'-untranslated region)에 위치하는 내부 리보솜 진입 부위(Internal Ribosome Entry Sites; IRESs)에 의해 매개된다. IRESs는 1980년대에 바이러스 mRNA에서 처음 발견되었으며, 세포에서의 내인성 IRESs도 이후 널리 보고되었다. 바이러스성 IRESs는 일반적으로 복잡한 2차 및 3 구조를 가지고 있다. 숙주세포 내에서, 바이러스성 IRESs는 숙주세포의 리보솜을 모집하여 숙주 세포의번역 개시 인자에 의존하거나 또는 비의존적인 단백질 번역을 개시한다. 바이러스성 IRESs와 비교하여, 세포에서 내인성 IRESs는 일반적으로 단백질 번역 개시 효율이 낮으며, 이는 공통성이 없는 다양하고 복잡한 메커니즘에 의해 조절된다. 세포에서 상이한 내인성 IRESs는 서열 및 구조에서 공통성을 가지고 있지 않으므로, 예측하기 어렵다.

[0005] 세포 내의 단백질 합성에 대한 상기 이해에 더하여, 단백질 합성은 또한 세포 외부에서 수행될 수 있다. 시험관 내(in vitro) 단백질 합성 시스템은 일반적으로 박테리아, 진균, 식물 세포 또는 동물 세포에 기초한 용혈 시스템(lysis system)을 나타내며, 이는 외래성 단백질의 번역을 신속하고 효율적으로 완료하기 위하여 mRNA 또는 DNA 주형, RNA 중합효소, 아미노산 및 ATP를 포함하는 요소가추가된다. 현재, 자주 이용되고 있는 상용성 시험관 내 단백질 발현 시스템은 대장균추출물(*E. coli* extract, ECE) 시스템, 토끼 망상적혈구(rabbit

reticulocyte lysate, RRL)시스템, 밀배아 추출물(wheat germ extract, WGE)시스템, 곤충 세포 추출물(insect cell extract, ICE) 시스템 및 인간 자원 시스템을 포함한다.

- [0006] 시험관 내에서 합성된 mRNA는 일반적으로 “캡 구조”를 가지고 있지 않으며, “캡 구조”를 갖는 mRNA로의 변형은 많은 시간이 걸리며 비싸다. 결과적으로, 시험관 내 단백질 합성 시스템은 일반적으로 단백질을 위해 캡-비의존적 번역 개시 방법을 사용한다. 그러나, IRES 또는 Ω 서열만을 사용한 번역 개시의 효율은 다소 낮으며, 시험관 내에서 빠르고 효율적이며 고수율(high-throughput)로 단백질을 합성하는 목적을 달성할 수 없다. 현재, 이 두 가지 유형의 번역 개시 요소의 탠덤연결에 관한 연구는 없다.

발명의 내용

해결하려는 과제

- [0007] 따라서, 당업계에서는 단백질 합성 효율을 향상시킬 수 있는 새로운 핵산 구조체(construct)를 개발할 필요가 있으며, 이는, IRES 서열 및 Ω 서열을 탠덤(in tandem) 연결한다.

과제의 해결 수단

- [0008] 본 발명의 하나의 목적은 단백질 번역의 효율을 향상시킬 수 있는 새로운 핵산 구조체를 제공하는 것으로, 이는 IRES 서열 및 Ω 서열을 탠덤 연결한다.
- [0009] 첫 번째 측면에 따르면, 본 발명은 식 I의 핵산 서열을 포함하는 핵산 구조체를 제공한다.
- [0010] Z1-Z2-Z3-Z4-Z5 (I)
- [0011] 상기에서,
- [0012] Z1~Z5는 각각 상기 구조체의 일부인 요소이고;
- [0013] 각각의 “-”는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;
- [0014] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0015] Z2는 담배 모자이크 바이러스의 5'선도서열(leading sequence), 즉 Ω 서열이며;
- [0016] Z3은 아데닌 데옥시뉴클레오타이드(adenine deoxynucleotide)의 올리고머 사슬이며, 또한 [oligo(A)]_n으로 표시되고;
- [0017] Z4는 번역 개시 코돈이며;
- [0018] Z5는 세린 코돈이고;
- [0019] 상기 Z3, Z4 및 Z5는 코작 서열(Kozak sequence)을 구성(constitute)하고, 상기 Kozak 서열은 효모로부터 유래된 것이다.
- [0020] 다른 바람직한 실시예에서, Z1은 효모로부터 유래된 것이다.
- [0021] 다른 바람직한 실시예에서, IRES 요소의 공급원은 원핵세포 및 진핵세포로 구성된 군으로부터 선택된 하나 이상의 유형의 세포이다.
- [0022] 다른 바람직한 실시예에서, 진핵세포는 고등진핵세포를 포함한다.
- [0023] 다른 바람직한 실시예에서, IRES 요소는 내인성 IRES 요소 및 외인성 IRES 요소를 포함한다.
- [0024] 다른 바람직한 실시예에서, IRES 요소의 공급원은 인간(human) 세포, 중국 햄스터 난소(CHO, Chinese hamster ovary) 세포, 곤충(insect) 세포, 밀 배아(wheat germ) 세포 및 토끼 망상적혈구 세포(rabbit reticulocyte)로 구성된 군으로부터 선택된 하나 이상의 유형의 세포이다.
- [0025] 다른 바람직한 실시예에서, IRES 요소는 ScBOI1, ScFLO8, ScNCE102, ScMSN1, KIFLO8, KINCE102, K1MSN1, KIBOI1 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0026] 다른 바람직한 실시예에서, 효모는 사카로마이세스 세레비제(*Saccharomyces cerevisiae*), 클루이베로마이세스(*Kluyveromyces*) 및 이들의 조합으로 구성된 군으로부터 선택된다.

- [0027] 다른 바람직한 실시예에서, 클루이베로마이세스는 클루이베로마이세스 락티스(*Kluyveromyces lactis*), 클루이베로마이세스 막시아누스(*Kluyveromyces marxianus*), 클루이베로마이세스 도브잔스키(*Kluyveromyces dobzhanskii*) 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0028] 다른 바람직한 실시예에서, 번역 개시 코돈은 ATG, ATA, ATT, GTG, TTG 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0029] 다른 바람직한 실시예에서, 번역 개시 코돈은 ATG이다.
- [0030] 다른 바람직한 실시예에서, 세린 코돈은 TCT, TCC, TCA, TCG, AGT, AGC 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0031] 다른 바람직한 실시예에서, 세린 코돈은 TCT이다.
- [0032] 다른 바람직한 실시예에서, n은 6-12, 바람직하게는, 8-10이다.
- [0033] 다른 바람직한 실시예에서, Ω 서열은 직접 반복 모듈(direct repeat modules)(ACAATTAC)_m 및 (CAA)_p을 포함한다.
- [0034] 다른 바람직한 실시예에서, m은 1-6, 바람직하게는, 2-4이다.
- [0035] 다른 바람직한 실시예에서, p는 6-12, 바람직하게는, 8-10이다.
- [0036] 다른 바람직한 실시예에서, (CAA)_p 모듈의 수는 1-5, 바람직하게는, 1-3이다.
- [0037] 다른 바람직한 실시예에서, (CAA)_p 모듈은 최적화된 (CAA)_p 모듈을 더 포함한다.
- [0038] 다른 바람직한 실시예에서, Kozak 서열은 서열번호 84(SEQ ID NO.: 84)에 나타내었다.
- [0039] 다른 바람직한 실시예에서, 핵산 구조체의 서열은 서열번호 2-17(SEQ ID NO.: 2-17) 중 어느 하나일 수 있다.
- [0040] 다른 바람직한 실시예에서, 핵산 구조체의 서열은 서열번호 2-9(SEQ ID NO.: 2-9) 중 어느 하나일 수 있다.
- [0041] 다른 바람직한 실시예에서, 핵산 구조체의 서열은 서열번호 3, 4 또는 6(SEQ ID NO.: 3, 4 또는 6)이다.
- [0042] 본 발명의 두 번째 측면에 따르면, 본 발명은 5'에서 3'의 화학식 II의 구조를 포함하는 핵산 구조체를 제공한다.
- [0043] Z1-Z2 Z3-Z4-Z5-Z6 (II)
- [0044] 상기에서,
- [0045] Z1~Z6은 각각 핵산 구조체의 일부인 요소이고;
- [0046] 각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;
- [0047] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0048] Z2는 담배 모자이크 바이러스의 5'선도서열(leading sequence), 즉 Ω 서열이며;
- [0049] Z3은 아데닌 테옥시뉴클레오타이드의 올리고머 사슬이며, 또한 [oligo(A)]_n으로 표시되고;
- [0050] Z4는 번역 개시 코돈이며;
- [0051] Z5는 세린 코돈이고;
- [0052] Z6은 외래성 단백질의 암호화 서열이며;
- [0053] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.
- [0054] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 원핵생물 또는 진핵생물로부터 유래한다.
- [0055] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 동물, 식물 또는 병원체로부터 유래한다.
- [0056] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 포유동물, 바람직하게는 인간, 마우스 및 쥐를 포함한 영장류 또는 설치류로부터 유래한다.

- [0057] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 루시페린(luciferin), 또는 루시페라아제(luciferases)(예: 반딧불이 루시페라아제), 녹색형광단백질(green fluorescent protein), 황색형광단백질(yellow fluorescent protein), 아미노아실-tRNA합성효소(aminoacyl-tRNA synthetase), 글리세르알데히드-3-인산수소이탈효소(glyceraldehyde-3-phosphate dehydrogenase), 카탈라아제(catalase), 액틴(actin), 항체의 가변영역, 루시페라아제 돌연변이(luciferase mutants), α -아밀라아제(α -amylase), 엔테로신 A(enterocin A), C형 간염 바이러스 E2 당단백질(hepatitis C virus (HCV) E2 glycoprotein), 인슐린 전구체(insulin precursors), 인터페론 α A(interferon α A), 인터루킨-1 β (interleukin-1 β), 라이소자임(lysozyme), 혈청 알부민(serum albumins), 항체의 단쇄가변분절(single-chain variable fragment)(scFv), 트란스타이레틴(transthyretin), 티로시나아제(tyrosinase), 자일라나아제(xylanase) 및 이들의 조합으로 구성된 군으로부터 선택된 외래성 단백질을 코딩한다.
- [0058] 다른 바람직한 실시예에서, 외래성 단백질은 루시페린, 또는 루시페라아제(예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 액틴, 항체의 가변영역, 루시페라아제 돌연변이, α -아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 α A, 인터루킨-1 β , 라이소자임, 혈청 알부민, 항체의 단쇄가변분절(scFv), 트란스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0059] 다른 바람직한 실시예에서, 핵산 구조체의 서열은 서열번호 85-87(SEQ ID NO.: 85-87) 중 어느 하나일 수 있다.
- [0060] 본 발명의 세 번째 측면에 따르면, 본 발명은 5'에서 3'의 화학식 III의 구조를 포함하는 핵산 구조체를 제공한다.
- [0061] Z0-Z1-Z2 Z3-Z4-Z5-Z6 (III)
- [0062] 상기에서,
- [0063] Z0~Z6은 각각 핵산 구조체의 일부인 요소이고;
- [0064] 각각의 "-"는 독립적으로 결합 또는 뉴클레오티드 연결 서열이고;
- [0065] Z0은 프로모터 요소이고, 프로모터 요소는 T7 프로모터, T3 프로모터, SP6 프로모터 및 이들의 조합으로 구성된 군으로부터 선택되고;
- [0066] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0067] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;
- [0068] Z3은 아데닌 테옥시뉴클레오티드의 올리고머 사슬이며, 또한 [oligo(A)]_n으로 표시되고;
- [0069] Z4는 번역 개시 코돈이며;
- [0070] Z5는 세린 코돈이고;
- [0071] Z6은 외래성 단백질의 암호화 서열이며;
- [0072] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.
- [0073] 본 발명의 네 번째 측면에 따르면, 본 발명은 5'에서 3'의 화학식 IV의 구조를 포함하는 핵산 구조체를 제공한다.
- [0074] Z0'-Z1-Z2 Z3-Z4-Z5-Z6 (IV)
- [0075] 상기에서,
- [0076] Z0'~Z6은 각각 구조체의 일부인 요소이고;
- [0077] 각각의 "-"는 독립적으로 결합 또는 뉴클레오티드 연결 서열이고;
- [0078] Z0'은 GAA이고;
- [0079] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0080] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

- [0081] Z3은 아데닌 데옥시뉴클레오타이드의 올리고머 사슬이며, 또한 [oligo(A)]_n으로 표시되고;
- [0082] Z4는 번역 개시 코돈이며;
- [0083] Z5는 세린 코돈이고;
- [0084] Z6은 외래성 단백질의 암호화 서열이며;
- [0085] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.
- [0086] 본 발명의 다섯 번째 측면에 따르면, 본 발명은 벡터 또는 벡터 조합물(vector combination)을 제공하며, 상기 벡터 또는 벡터 조합물은 본 발명의 첫 번째 내지 네 번째 측면에 따른 핵산 구조체를 포함한다.
- [0087] 본 발명의 여섯 번째 측면에 따르면, 본 발명은 유전자 변형 세포(genetically engineered cell)를 제공하며, 상기 유전자 변형 세포의 게놈(genome)은 하나 이상의 위치(site)에 본 발명의 첫 번째 내지 네 번째 측면에 따른 핵산 구조체와 통합되거나, 유전자 변형 세포는 본 발명의 다섯 번째 측면에 따른 벡터 또는 벡터 조합물을 포함한다.
- [0088] 다른 바람직한 실시예에서, 유전자 변형 세포는 원핵 세포 또는 진핵 세포를 포함한다.
- [0089] 다른 바람직한 실시예에서, 진핵 세포는 고등 진핵 세포를 포함한다.
- [0090] 다른 바람직한 실시예에서, 유전자 변형 세포는 인간 세포(예를 들어, HeLa cell), 중국 햄스터 난소 세포(CHO), 곤충 세포, 밀 배아 세포, 토끼 망상적혈구 세포, 효모 세포 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0091] 다른 바람직한 실시예에서, 유전자 변형 세포는 효모 세포이다.
- [0092] 다른 바람직한 실시예에서, 효모 세포는 사카로마이세스 세레비제, 클루이베로마이세스 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0093] 다른 바람직한 실시예에서, 클루이베로마이세스는 클루이베로마이세스 락티스, 클루이베로마이세스 막시아누스, 클루이베로마이세스 도브잔스키 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0094] 본 발명의 일곱 번째 측면에 따르면, 본 발명은 키트를 제공하며, 상기 키트는 하나 이상의 다음 군으로부터 선택된 시약을 포함한다:
- [0095] (a) 본 발명의 첫 번째 내지 네 번째 측면에 따른 핵산 구조체;
- [0096] (b) 본 발명의 다섯 번째 측면에 따른 벡터 또는 벡터 조합물; 및
- [0097] (c) 본 발명의 여섯 번째 측면에 따른 유전자 변형 세포.
- [0098] 다른 바람직한 실시예에서, 키트는 (d)진핵 생물-기반 시험관 내 생합성 시스템(예컨대, 진핵 생물-기반 시험관 내 단백질 합성 시스템)을 더 포함한다.
- [0099] 다른 바람직한 실시예에서, 진핵 생물-기반 시험관 내 생합성 시스템은 효모-기반 시험관 내 생합성 시스템, 중국 햄스터 난소 세포 기반 시험관 내 생합성 시스템, 곤충 세포기반 시험관 내 생합성 시스템, HeLa 세포기반 시험관 내 생합성 시스템 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0100] 다른 바람직한 실시예에서, 진핵 생물-기반 시험관 내 생합성 시스템은 진핵 생물-기반 시험관 내 단백질 합성 시스템을 포함한다.
- [0101] 다른 바람직한 실시예에서, 진핵 생물-기반 시험관 내 단백질 합성 시스템은 효모-기반 시험관 내 단백질 합성 시스템, 중국 햄스터 난소 세포 기반 시험관 내 단백질 합성 시스템, 곤충 세포 기반 시험관 내 단백질 합성 시스템, HeLa 세포 기반 시험관 내 단백질 합성 시스템 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0102] 다른 바람직한 실시예에서, 키트는 (e) 효모-기반 시험관 내 생합성 시스템(예컨대, 효모-기반 시험관 내 단백질 합성 시스템)을 더 포함한다.
- [0103] 다른 바람직한 실시예에서, 효모-기반 시험관 내 생합성 시스템(예컨대, 효모-기반 시험관 내 단백질 합성 시스템)은 클루이베로마이세스-기반 시험관 내 생합성 시스템(예컨대, 클루이베로마이세스-기반 시험관 내 단백질 합성 시스템), 바람직하게는 클루이베로마이세스 락티스 기반 시험관 내 생합성 시스템(예컨대, 클루이베로마이

세스 락티스 기반 시험관 내 단백질 합성 시스템)이다.

- [0104] 본 발명의 여덟 번째 측면에 따르면, 본 발명은 본 발명의 첫 번째 내지 네 번째 측면에 따른 핵산 구조체, 본 발명의 다섯 번째 측면에 따른 벡터 또는 벡터 조합물, 및 여섯 번째 측면에 따른 유전자 변형 세포 또는 본 발명의 일곱 번째 측면에 따른 키트의 용도를 제공하고, 이는 시험관 내 단백질의 고수율 합성을 위해 적용 가능하다.
- [0105] 본 발명의 아홉 번째 측면에 따르면, 본 발명은 다음 단계를 포함하는 시험관 내 외래성 단백질을 고수율로 합성하는 방법을 제공한다:
- [0106] (i) 진핵 생물-기반 시험관 내 생합성 시스템의 존재하에, 본 발명의 첫 번째 내지 네 번째 측면에 따른 핵산 구조체를 제공하는 단계; 및
- [0107] (ii) 적합한 조건 하에서, 단계 (i)의 진핵 생물-기반 시험관 내 생합성 시스템을 T1 기간 동안 인큐베이션하여 외래성 단백질을 합성하는 단계.
- [0108] 다른 바람직한 실시예에서, 상기 방법은 (iii) 진핵 생물-기반 시험관 내 생합성 시스템의 외래성 단백질을 선택적으로 분리 또는 검출하는 단계를 더 포함한다.
- [0109] 다른 바람직한 실시예에서, 진핵 생물-기반 시험관 내 생합성 시스템은 효모-기반 시험관 내 생합성 시스템(예컨대, 효모-기반 시험관 내 단백질 합성 시스템)이다.
- [0110] 다른 바람직한 실시예에서, 효모-기반 시험관 내 생합성 시스템(예컨대, 효모-기반 시험관 내 단백질 합성 시스템)은 클루이베로마이세스 기반 시험관 내 생합성 시스템(예컨대, 클루이베로마이세스 기반 시험관 내 단백질 합성 시스템), 바람직하게는 클루이베로마이세스 락티스 기반 시험관 내 생합성 시스템(예컨대, 클루이베로마이세스 락티스 기반 시험관 내 단백질 합성 시스템)이다.
- [0111] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 원핵 생물 또는 진핵 생물로부터 유래된 것이다.
- [0112] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 동물, 식물 또는 병원체로부터 유래된 것이다.
- [0113] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 포유동물, 바람직하게는 인간, 마우스 및 쥐를 포함한 영장류 또는 설치류로부터 유래된 것이다.
- [0114] 다른 바람직한 실시예에서, 외래성 단백질의 암호화 서열은 루시페린, 또는 루시페라아제 (예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 액틴, 항체의 가변영역, 루시페라아제 돌연변이, α -아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 α A, 인터루킨-1 β , 라이소자임, 혈청 알부민, 항체의 단쇄가변분절 (scFv), 트란스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합으로 구성된 군으로부터 선택된 외래성 단백질을 암호화(encode)한다.
- [0115] 다른 바람직한 실시예에서, 외래성 단백질은 루시페린, 또는 루시페라아제 (예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 액틴, 항체의 가변영역, 루시페라아제 돌연변이, α -아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 α A, 인터루킨-1 β , 라이소자임, 혈청 알부민, 항체의 단쇄가변분절 (scFv), 트란스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0116] 다른 바람직한 실시예에서, 단계 (ii)에서, 반응 온도는 20-37 °C 바람직하게는 22-35 °C이다.
- [0117] 다른 바람직한 실시예에서, 단계 (ii)에서, 반응 시간은 1-10 시간, 바람직하게는 2-8시간이다.
- [0118] 본 발명의 범위 내에서, 상술된 기술적 특징 및 본 발명에서 이하에 설명된(예를들어, 실시예 또는 실험예) 기술적 특징의 구체적인 설명은 서로 조합될 수 있으며, 이에 따라 새로운 또는 바람직한 기술적 솔루션 형성이 가능함을 이해해야 한다. 공간상 한계로 인해 여기에 더 이상 반복하지 않는다.

발명의 효과

- [0119] (1) 첫 번째로, 본 발명은 프로모터, 효모-유래 IRES, Ω 서열, Kozak 서열 및 외래성 단백질의 암호화 서열을 포함하는 핵산 구조체가, 본 발명의 진핵 생물-기반 시험관 내 생합성 시스템(예컨대, 효모-기반 시험관 내 단백질 합성 시스템)에서 이용될 경우, 외래성 단백질 번역의 효율을 현저히 개선시킬 수 있음을 발견하였다.

- [0120] (2) Ω 서열 및 Kozak 서열로 탠덤 연결된 본 발명의 진핵 세포의 내인성 IRESs는 진핵 생물-기반 시험관 내 생합성 시스템에서 단백질 번역 개시의 효율을 향상시킬 수 있다. 번역 개시의 효율성을 향상시키는 측면에서, 탠덤 DNA요소는 마지막의 2개만을 연결하는 탠덤 시퀀스 (Ω-10A)보다 유리하다. 상기, 클루이베로마이세스 락티스 기반 시험관 내 생합성 시스템에서, KINCE102-Ω-10A를 사용하여 합성을 개시할 때 반딧불이 루시페라아제 (Fluc)의 상대적 발광 수치(RLU) 값은 1.67×10^9 에 도달하였으며, 이는 Ω-10A 서열을 이용할 때의 1.65배였다.
- [0121] (3) 사카로마이세스 세레비제와 비교하여 클루이베로마이세스 락티스는 안정성 및 고효율성으로 인해 식품 및 제약 분야에서 단백질 생산에 이용될 수 있고, 시험관 내 생합성 시스템의 장점(예컨대, 고수율 단백질 합성 및 스크리닝에 대한 적합성, 독성 단백질 합성의 능력, 최소 시간, 최소 비용 등)도 있으며, 따라서 클루이베로마이세스 락티스로부터 유래된 시험관 내 생합성은 단백질 합성 분야에서도 널리 이용될 수 있다.
- [0122] (4) 본 발명에 의해 제공되는 핵산 구조체는 진핵 생물-기반 시험관 내 생합성 시스템의 단백질 번역 개시의 효율을 향상시킬 뿐만 아니라, 더욱 중요하게는, 상이한 단백질의 합성을 위한 클루이베로마이세스 락티스 기반 시험관 내 생합성의 가능성을 증가시킨다.
- [0123] (5) 본 발명의 핵산 구조체는 단백질 번역 개시의 효율을 향상시킬 뿐만 아니라 진핵 생물-기반 세포 시험관 내 생합성 시스템에 이용되는 DNA요소를 설계하기 위한 새로운 개념 및 방법을 제공하고, 이는 과학적 연구 및 산업 생산 분야에서 관련 시스템의 적용을 크게 증가시킬 수 있다.
- [0124] (6) 본 발명은 강한 프로모터(예를 들어, T7 프로모터, T3 프로모터, SP6 프로모터)를 본 발명의 핵산 구조체와 조합하여 매우 높은 단백질 합성 효율을 달성할 수 있음을 처음으로 발견하였다.
- [0125] (7) 본 발명은 본 발명의 핵산 구조체의 업스트림에GAA의 3개의 잔기를 배치하는 것 또한 매우 높은 단백질 합성 효율을 달성할 수 있음을 발견하였다.

도면의 간단한 설명

- [0126] 도 1은 생합성에서 단백질 번역 개시에서의 5'-UTR 염기서열의 중요한 역할을 나타낸다. 5'-UTR는 주로 단백질 번역 개시를 조절하고 세포에서 mRNA를 안정화시키는 역할을 한다. (A) 캡-의존적 단백질 번역 개시의 과정에서, 5'-UTR은 번역 개시 인자 및 43S전사 개시전 복합체(pre-initiation complex, PIC)를 모집하는데 중요한 역할을 하고, 43S PIC 및 번역 개시의 스캐닝(scanning)을 조절할 수도있다. 캡-비의존적 단백질 번역 개시의 과정에서, (B) 2차 구조를 갖는 5'-UTR 및 (C) 2차 구조를 갖지 않는 5'-UTR은 일부 단백질 인자의 도움이 필요하고, 이는 43S PIC를 모집하는데 매우 중요하다.

도2는 시험관 내 단백질 합성 시스템에서 진핵 생물의 내인성 IRESs의 단백질 합성 개시의 효율을 나타낸다. 사카로마이세스 세레비제 및 클루이베로마이세스 락티스로부터 유래된 IRESs로부터 선택된 8개의 상동성 IRES를 효모-기반 시험관 내 단백질 합성 시스템에 사용하였다. 기존의 Ω 서열과 비교하여, 6개의 IRESs(ScFLO8, ScMSN1, ScNCE102, K1FLO8, K1MSN1 및 KINCE102)를 사용하여 합성을 시작할 때 루시페라아제의 상대적 발광 수치(RLU) 값은 Ω 서열을 사용할 때의 값을 초과하였다.

도 3은 시험관 내 단백질 합성 시스템에서 단백질 합성 개시 효율에서의 탠덤 DNA 요소를 비교한 것을 나타낸다. Ω 서열의 업스트림(upstream) 또는 다운스트림(downstream)에 각각 위치한 8개의 IRESs는 각각 Ω 서열 및 클루이베로마이세스락티스-특이적 Kozak 서열과 탠덤연결되고, 16개의 탠덤 DNA 요소를 구성하였다. 16개의 탠덤 DNA요소는 시험관 내 단백질 합성에서 사용하였고, 여기서 KINCE102-Ω-10A를 사용하여 합성을 개시할 때 루시페라아제의 상대적 발광 수치(RLU) 값은 Ω 서열 및 Kozak 서열이 연결된 탠덤 요소 Ω-10A를 사용할 때보다 초과하고, RLU 값은 Ω-10A를 사용할 때의 1.65-배(fold)였다. 두 번째로, 다른 2개의 탠덤 요소인 ScFLO8-Ω-10A 및 K1MSN1-Ω-10A를 사용할 때의 상대적 발광 수치(RLU) 값은 각각 Ω-10A를 사용할 때의 RLU 값에 근접하였으며, Ω-10A를 사용할 때의 RLU 값의 81.68 % 및 85.35 %에 도달하였다.

도 4는 본 발명에서 KINCE102-Ω-10A 서열에 의한 시험관 내 단백질 번역 개시의 효율을 향상시키는 메커니즘을 나타낸다. (A) Ω 서열은 43S PIC를 모집하고 단백질 번역을 개시하기 위하여 번역 개시 인자 eIF4G와의 결합해야 한다. (B) KINCE102는 폴리A 결합 단백질(poly (A) binding protein) Pab1을 모집할 수 있는 A-리치(rich) RNA 서열이다. Pab1는 eIF4G와 상호작용하여 43S PIC에 대한 모집 효과를 향상시키고 번역 개시 효율을 향상시킬 수 있다. KINCE102 및 Ω 서열은 시너지 효과를 형성하여 번역 개시 효율을 향상시킨다.

도 5는 Ω-10A 서열 및 GAA 서열이 K10NCE102의 시험관 내 단백질 번역 효율을 향상 시켰음을 보여준다. K10NCE102-Ω-10A 탠덤 요소의 시험관 내 번역 효율은 K10NCE102의 1.23배 이고, GAA-K10NCE102-Ω-10A 탠덤 요소의 시험관 내 번역 효율은 K10NCE102-Ω-10A의 1.28배, Ω-10A의 2.52배이다.

발명을 실시하기 위한 구체적인 내용

- [0127] 많은 스크리닝 및 탐색을 통해, 광범위하고 심도있는 연구를 수행 한 후, 시험관 내 단백질 번역의 효율성을 향상시킬 수 있는 새로운 핵산 구조체를 처음으로 예상치 못하게 발견하였다. 본 발명의 핵산 구조체는 진핵 세포(예를 들어, 효모)로부터 유래된 IRES 인핸서(예를 들어, ScBOI1, ScFLO8, ScNCE102, ScMSN1, K1FLO8, K1NCE102, K1MSN1, K1BOI1), Ω 서열 및 효모 특이적(예컨대, 클루이베로마이세스, 바람직하게는 클루이베로마이세스 락티스) Kozak 서열을 연결시킴으로써 설계되었다. 본 발명의 핵산 구조체는 효모-기반 시험관 내 생합성 시스템(예를 들어, 효모-기반 시험관 내 단백질 합성 시스템)에 사용되고, 합성된 루시페라아제의 활성을 나타내기 위한 상대적 발광 수치(RLU) 값은 매우 높으며, 단일 Ω 서열과 Kozak 서열로 연결된 탠덤서열(Ω-10A)의 1.65 배에 도달하거나 Ω-10A와 거의 동일 할 수 있다.
- [0128] 또한, 본 발명자들은 놀랍게도 본 발명의 핵산 구조체의 업스트림에 GAA의 3개의 잔기를 배치하는 것이 단백질 합성의 효율을 향상시킬 수 있으며, 시험관 내 합성을 매개하기 위한 본 발명의 핵산 구조체를 사용할 때 루시페라아제의 상대적 발광 수치 값은 약 0.28배 증가하였다(K1NCE102-Ω-10A에 비해 GAA-K1NCE102-Ω-10A). Ω-10A를 사용할 때와 비교하면, GAA-K1NCE102-Ω-10A를 사용할 때의 루시페라아제의 상대적 발광 수치 값은 1.52배 증가하였다(즉, GAA-K1NCE102-Ω-10A를 사용할 때의 상대적 발광 수치 값은 Ω-10A를 사용할 때의 값의 2.52배였다)(도 5에 도시된 바와 같이). 상기에 기초하여, 본 발명자들은 본 발명을 완성하였다.
- [0129] **진핵 생물 기반 시험관 내 생합성 시스템**
- [0130] 진핵 생물-기반 시험관 내 생합성은 진핵 세포 기반의 전사-번역 공역(transcription-translation coupled) 시스템으로, 이는 DNA 주형으로부터 RNA 합성을 시작하거나, 주형으로서 DNA 또는 RNA를 이용하여 시험관 내에서 단백질 합성을 완료할 수 있다. 진핵 세포는 효모 세포, 토기 망상적혈구, 밀배아 세포, 곤충 세포, 인간 세포 및 이와 같은 세포를 포함한다. 진핵 생물-기반 시험관 내 생합성 시스템은 RNA 또는 복잡한 구조를 갖는 단백질을 합성할 수 있는 장점을 갖고, 단백질 등의 번역 후 변형(post-translational modification)의 장점을 갖는다.
- [0131] 본 발명에 있어서, 진핵 생물-기반 시험관 내 생합성 시스템은 특별히 제한되지 않는다. 바람직한 진핵 생물-기반 시험관 내 생합성 시스템은 효모-기반 시험관 내 생합성 시스템, 바람직하게는, 효모-기반 시험관 내 단백질 합성 시스템, 더 바람직하게는 클루이베로마이세스-기반 발현 시스템(더 바람직하게는, 클루이베로마이세스 락티스 기반 발현 시스템)을 포함한다.
- [0132] 효모(yeast)는 단순한 배양, 효율적인 단백질 접힘(protein folding) 및 번역 후 변형의 장점을 갖는다. 효모들 중에서, 사카로마이세스 세레비제(*Saccharomyces cerevisiae*) 및 피히아 파스토리스(*Pichia pastoris*)는 복잡한 진핵생물 단백질을 발현하는 유기체 모델이다. 효모는 또한 시험관 내 번역 시스템의 준비를 위한 원료로서 사용될 수 있다.
- [0133] 클루이베로마이세스(*Kluyveromyces*)는 자낭포자형성효모(ascosporogenous yeast)이다. 클루이베로마이세스 중에서 클루이베로마이세스 막시아누스(*Kluyveromyces marxianus*) 및 클루이베로마이세스 락티스(*Kluyveromyces lactis*)는 산업에서 널리 사용되고 있는 효모이다. 다른 효모들과 비교할 때, 클루이베로마이세스 락티스는 뛰어난 분비 능력, 더 나은 대규모 발효 특성, 식품 안전성 수준에 부합(conforming)하고, 단백질의 번역 후 변형의 능력을 갖는 것과 같은 많은 장점을 가지고 있다.
- [0134] 본 발명에 있어서, 진핵 생물-기반 시험관 내 생합성 시스템은 다음을 포함한다:
- [0135] (a) 진핵 세포 추출물;
- [0136] (b) 폴리에틸렌 글리콜;
- [0137] (c) 선택적(optional) 외래성 수크로오스; 및
- [0138] (d) 선택적(optional) 용매, 이것은 물 또는 수용성 용매.
- [0139] 가장 바람직한 실시예에서, 본 발명에 의한 시험관 내 생합성 시스템은 진핵 세포 추출물, 4-하이드록시 피페라

진에탄설펜산(4-hydroxyethyl piperazineethanesulfonic acid), 아세트산 칼륨, 아세트산 마그네슘, 아데노신 3인산(ATP), 구아노신 3인산(GTP), 사이티딘 3인산(CTP) 및 티미딘 3인산(TTP), 아미노산 혼합물, 포스포크레아틴(phosphocreatine), 디티오쓰레이톨(DTT), 크레아틴포스포키네이스(creatine phosphokinase), 리보핵산가수분해효소 억제제, 루시페린, 루시페라아제의 DNA 및 RNA 중합효소를 포함한다.

- [0140] 본 발명에 있어서, RNA 중합효소는 특별히 제한되지 않으며, 하나 이상의 RNA 중합효소일 수 있다. 대표적인 RNA 중합효소는 T7 RNA 중합효소이다.
- [0141] 본 발명에 있어서, 시험관 내 생합성 시스템에서 진행 세포 추출물의 비율은 특별히 제한되지 않으며; 일반적으로 시험관 내 시스템에서 진행 세포 추출물은 20-70%, 바람직하게는 30-60%, 더 바람직하게는 40-50%를 차지한다.
- [0142] 본 발명에 있어서, 진행 세포 추출물은 온전한 기능을 갖춘 세포(intact cells)를 포함하지 않는다. 대표적인 진행 세포 추출물은 RNA합성을 위한 다양한 유형의 RNA 중합효소, 리보솜, 운반 RNA(tRNA), 아미노아실-tRNA합성효소를 포함하는 단백질 번역을 위한 인자, 및 개시 인자, 신장 인자, 종결 인자를 포함하는 단백질 합성에 요구되는 인자를 포함한다. 또한, 진행 세포 추출물은 진행 세포의 세포질로부터 유래된 몇몇의 다른 단백질, 특히 수용성 단백질을 포함한다.
- [0143] 본 발명에 있어서, 진행 세포 추출물의 단백질 함량은 20-100 mg/mL, 바람직하게는 50-100 mg/mL이다. 단백질 함량을 측정하는 방법은쿠마시브릴리언트블루(Coomassie Brilliant Blue) 분석법이다.
- [0144] 본 발명에 있어서, 진행 세포 추출물의 제조방법은 제한되지 않는다. 바람직한 제조 방법은 다음 단계를 포함한다:
- [0145] (i) 진행 세포 준비하는 단계;
- [0146] (ii)진행 세포를 세척하여 세척된 진행 세포를 수득하는 단계;
- [0147] (iii) 세척된 진행 세포를 세포 용해 처리제로 처리하여 진행 세포 조추출물을 수득하는 단계; 및
- [0148] (iv) 진행 세포 조추출물을 고액분리(solid-liquid separation)를 통해 액상층(즉, 진행 세포 추출물)을 수득하는 단계.
- [0149] 본 발명에 있어서, 고액분리 방법은 특별히 제한되지 않는다. 바람직한 방법은 원심분리이다.
- [0150] 바람직한 실시예에서, 원심분리는 액체 상태에서 수행한다.
- [0151] 본 발명에 있어서, 원심분리 조건은 특별히 제한되지 않는다. 바람직한 원심분리 조건은 5000g - 10000 g, 더 바람직하게는 8000 g - 30000 g이다.
- [0152] 본 발명에 있어서, 원심분리 시간은 특별히 제한되지 않는다. 바람직한 원심분리 시간은 0.5분 내지 2시간, 바람직하게는 20분 내지 50분이다.
- [0153] 본 발명에 있어서, 원심분리의 온도는 특별히 제한되지 않는다. 바람직하게는, 원심분리는 1-10℃, 바람직하게는 2-6℃에서 수행한다.
- [0154] 본 발명에 있어서, 세척 처리 방법은 특별히 제한되지 않는다. 바람직한 세척 처리 방법은 pH7-8(바람직하게는, pH7.4)의 세척 용매를 이용하여 처리하여 수행한다. 세척 용매는 특별히 제한되지 않는다. 대표적인 세척 용매는 포타슘4-하이드록시에틸피페라진 에탄설펜산(potassium 4-hydroxyethylpiperazine ethanesulfonate), 아세트산 칼륨, 아세트산 마그네슘 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0155] 본 발명에 있어서, 세포 용해 처리 방법은 특별히 제한되지 않는다. 바람직한 세포 용해 처리 방법은 고압용해 및 동결해동(예를 들어, 저온의 액화질소에서 처리함)용해 등을 포함한다.
- [0156] 시험관 내 생합성 시스템에서 뉴클레오사이드삼인산(nucleoside triphosphates)의 혼합물은 아데노신 3인산, 구아노신 3인산, 사이티딘 3인산및 우리딘3인산을 포함한다. 본 발명에서, 다양한 단일 뉴클레오티드의 농도는 특별히 제한되지 않는다. 각 단일 뉴클레오티드의 농도는 일반적으로 0.5-5 mM, 바람직하게는 1.0-2.0 mM이다.
- [0157] 시험관 내 생합성 시스템에서 아미노산 혼합물은 천연 또는 비천연 아미노산을 포함할 수 있고, D-형 또는 L-형의 아미노산을 포함할 수 있다. 대표적인 아미노산은 글리신, 알라닌, 발린, 류신, 이소류신, 페닐알라닌, 프롤린, 트립토판, 세린, 트립신, 시스테인, 메티오닌, 아스파라진, 글루타민, 트레오닌, 아스파르트산, 글루탐산,

라이신, 아르기닌 및 히스티딘과 같은 20가지 유형의 천연 아미노산을 포함하지만 이제 제한되지 않는다. 각 타입의 아미노산 농도는 일반적으로 0.01-0.5 mM, 바람직하게는 0.02-0.2 mM, 예컨대, 0.05 mM, 0.06 mM, 0.07 mM 및 0.08 mM이다.

[0158] 바람직한 실시예에서, 시험관 내 생합성 시스템은 폴리에틸렌 글리콜(PEG) 등을 더 포함한다. 폴리에틸렌 글리콜 등의 농도는 특별히 제한되지 않는다. 일반적으로, 생합성 시스템의 총 중량을 기준으로, PEG 등의 농도(w/v)는 0.1-8%, 바람직하게는 0.5-4% 및 더 바람직하게는 1-2%이다. PEG의 대표적인 실시예는 PEG3000, PEG8000, PEG6000 및 PEG3350을 포함하지만 이에 제한되지는 않는다. 본 발명의 시스템은 또한 다른 다양한 분자량의 폴리에틸렌 글리콜을 포함할 수 있음을 이해해야한다(예컨대, PEG 200, 400, 1500, 2000, 4000, 6000, 8000, 10000 등)

[0159] 바람직한 실시예에서, 시험관 내 생합성 시스템은 수크로오스를 더 포함한다.

[0160] 수크로오스의 농도는 특별히 제한되지 않는다. 일반적으로, 단백질 생합성 시스템의 총 중량 기준으로, 수크로오스의 농도는 0.03-40wt%, 바람직하게는 0.08-10wt%, 및 더 바람직하게는 0.1-5wt%이다.

[0161] 가장 바람직한 실험관 내 생합성 시스템은 진핵 생물 세포 추출물에 추가적으로 다음과 같은 성분을 포함한다: pH7.4의 4-하이드록시 피페라진에탄설포산 22 mM, 아세트산 칼륨 30-150 mM, 아세트산 마그네슘 1.0-5.0 mM, 뉴클레오사이드 삼인산 혼합물 1.5-4 mM, 아미노산 혼합물 0.08 -0.24 mM, 포스포크레아틴 25mM, 디티오쓰레이톨 1.7mM, 크레아틴포스포키나제 0.27 mg/mL, PEG 1%-4%, 수크로오스 0.5%-2%, 반딧불이 루시페라아제의 DNA 8-20 ng/μL 및 T7 RNA 중합효소 0.027-0.054 mg/mL.

[0162] Ω 서열

[0163] 본 발명에서 사용된 용어 "Ω 서열"(즉, 오메가 서열)은 담배 모자이크 바이러스 게놈의 5' 선도 서열(5' 리더 서열)을 지칭하며, 이는 바이러스의 번역인핸서이다. Ω의 DNA 서열은 68 염기쌍을 포함하고, 8-염기쌍(ACAATTAC)의 직접 반복 모듈을 1-6(바람직하게는 2-4, 더 바람직하게는 3)개 포함하고, (CAA)_p 모듈을 1-5(바람직하게는 1-3, 더 바람직하게는 1)개 포함하며, 상기 p는 6-12, 바람직하게는 8-10이다. 이러한 두 모듈은 Ω 서열의 번역-인핸서 기능의 핵심이다. 본 발명의 효모-기반 시험관 내 단백질 합성 시스템에서, Ω 서열은 캡-비의존적 단백질 번역을 시작할 수 있으며, 이는 번역 개시 인자인 eIF4G를 모집함으로써 달성할 수 있다. 그러나, Ω 서열에 의해 개시된 단백질 번역의 효율은 상대적으로 낮다. Ω 서열의 구조적 서열을 최적화하고, 또한 다른 DNA 요소 또는 단백질 번역의 효율을 강화하기 위한 단백질과 함께 협력할 필요가 있다.

[0164] 코작 서열(Kozak 서열)

[0165] 공지된 진핵 생물의 mRNA 분자에서 번역 개시 코돈(AUG)의 업스트림 및 다운스트림 서열을 확인하는 것은 "Kozak 서열"이라고 하는 공통서열(consensus sequence)을 찾는 데 도움을 준다. Kozak 서열은 mRNA의 번역 개시 효율을 향상시키는 것으로 확인되었다. 다른 종의 Kozak 서열은 종종 다르다. 예를 들어, 사카로마이세스 세레비제(*Saccharomyces cerevisiae*)의 Kozak 서열 및 포유류 세포의 Kozak 서열은 현저히 상이하다.

[0166] 본 발명에서 사용된 Kozak 서열은 아데닌데옥시뉴클레오티드의 6-12개(바람직하게는 8-10개)의 올리고머 사슬, 번역 개시 코돈(예컨대, ATG, ATA, ATT, GTG, TTG 등, 바람직하게는 ATG), 및 세린 코돈(예컨대, TCT, TCC, TCA, TCG, AGT, AGC 등, 바람직하게는 TCT)을 포함하고, 클루이베로마이세스(*Kluyveromyces*) (바람직하게는 클루이베로마이세스 락티스(*Kluyveromyces lactis*)로부터)로부터 유래된 것이다.

[0167] 외래성 암호화 서열(외래성 DNA)

[0168] 본 발명에서 사용된 용어 "외래성 암호화 서열" 및 "외래성 DNA"는 상호 교환적으로 사용되며, 둘 다 RNA 또는 단백질 합성을 안내하는데 사용되는 외래성 DNA 분자를 지칭한다. 일반적으로, DNA 분자는 선형 또는 원형이다. DNA 분자는 외래성 RNA 또는 외래성 단백질을 암호화(encoding)하는 서열을 포함한다.

[0169] 본 발명에 있어서, 외래성 암호화 서열의 예는 게놈 서열 및 cDNA 서열을 포함하지만 이에 제한되지 않는다. 외래성 단백질을 암호화 하는 서열은 또한 프로모터 서열, 5' 비번역 서열 및/또는 3' 비번역 서열을 포함한다.

[0170] 본 발명에 있어서, 외래성 DNA의 선택은 특별히 제한되지 않는다. 일반적으로, 외래성 DNA는 작은 비암호화 RNA(small non-coding RNA)(sncRNA), 긴 비암호화 RNA(long non-coding RNA)(lncRNA), 운반 RNA(tRNA), 글루코사민-6-포스페이트 신타아제(glucosamine-6-phosphate synthase, glmS)를 포함한 리보자임, 소핵 RNA(small nuclear RNA)(snRNA), RNA 및 단백질의 복합체(예컨대, 스플라이소솜(Spliceosome)), 다양한 비암호화 RNA 및

이들의 조합으로 구성된 군으로부터 선택된다.

[0171] 외래성 DNA는 또한 루시페린, 또는 루시페라아제(예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 액틴 및 항체의 가변영역, 및 루시페라아제 돌연변이 암호화(encode)하는 외래성 DNA, 및 이들의 조합을들로 구성된 군으로부터 선택될 수 있다.

[0172] 외래성 DNA는 또한 α -아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 α A, 인터루킨-1 β , 라이소자임, 혈청 알부민, 항체의 단쇄가변분절(scFv), 트란스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합을 암호화하는 외래성 DNA들로 구성된 군으로부터 선택될 수 있다.

[0173] 다른 바람직한 실시예에서, 외래성 DNA는 녹색형광단백질(enhanced GFP, eGFP), 황색형광단백질(YFP), 대장균 β 갈락토시다아제(β galactosidase, LacZ), 인간 라이신-tRNA 합성 효소(lysine-tRNA synthetase), 인간 류신-tRNA 합성 효소(leucine-tRNA synthetase), 애기장대글리세르알데하이드-3-포스페이트 탈수소효소(Arabidopsis thaliana glyceraldehyde-3-phosphate dehydrogenase), 쥐 카탈라아제(murine catalase) 및 이들의 조합으로 구성된 군으로부터 선택된 단백질을 암호화(encode)한다.

[0174] 핵산 구조체

[0175] 본 발명은 식 I의 핵산 서열을 포함하는 핵산 구조체를 제공한다.

[0176] Z1-Z2-Z3-Z4-Z5 (I)

[0177] 상기에서,

[0178] Z1~Z5은 각각 구조체의 일부인 요소이고;

[0179] 각각의 "-"는 독립적으로 결합 또는 뉴클레오티드 연결 서열이고;

[0180] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;

[0181] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

[0182] Z3은 아데닌 데옥시뉴클레오티드의 올리고머 사슬[oligo (A)]_n이고;

[0183] Z4는 번역 개시 코돈이며;

[0184] Z5는 세린 코돈이고;

[0185] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.

[0186] 본 발명은 5'에서 3'의 화학식 II의 구조를 포함하는 핵산 구조체를 제공한다.

[0187] Z1-Z2 Z3-Z4-Z5-Z6 (II)

[0188] 상기에서,

[0189] Z1~Z6은 각각 핵산 구조체의 일부인 요소이고;

[0190] 각각의 "-"는 독립적으로 결합 또는 뉴클레오티드 연결 서열이고;

[0191] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;

[0192] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 Ω 서열이며;

[0193] Z3은 아데닌 데옥시뉴클레오티드의 올리고머 사슬[oligo(A)]_n이고;

[0194] Z4는 번역 개시 코돈이며;

[0195] Z5는 세린 코돈이고;

[0196] Z6은 외래성 단백질의 암호화 서열이며;

[0197] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.

[0198] 본 발명은 5'에서 3'의 화학식 III의 구조를 포함하는 핵산 구조체를 제공한다.

- [0199] Z0-Z1-Z2 Z3-Z4-Z5-Z6 (III)
- [0200] 상기에서,
- [0201] Z0~Z6은 각각 핵산 구조체의 일부인 요소이고;
- [0202] 각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;
- [0203] Z0은 프로모터 요소이고, 프로모터 요소는 T7 프로모터, T3 프로모터, SP6 프로모터 및 이들의 조합으로 구성된 군으로부터 선택되고;
- [0204] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0205] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 오 서열이며;
- [0206] Z3은 아데닌 테옥시뉴클레오타이드의 올리고머 사슬 [oligo(A)]_n이고;
- [0207] Z4는 번역 개시 코돈이며;
- [0208] Z5는 세린 코돈이고;
- [0209] Z6은 외래성 단백질의 암호화 서열이며;
- [0210] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.
- [0211] 본 발명은 5'에서 3'의 화학식 IV의 구조를 포함하는 핵산 구조체를 제공한다.
- [0212] Z0'-Z1-Z2 Z3-Z4-Z5-Z6 (IV)
- [0213] 상기에서,
- [0214] Z0'~Z6은 각각 구조체의 일부인 요소이고;
- [0215] 각각의 "-"는 독립적으로 결합 또는 뉴클레오타이드 연결 서열이고;
- [0216] Z0'은 GAA이고;
- [0217] Z1은 인핸서 요소이고, 인핸서 요소는 IRES 요소를 포함하며;
- [0218] Z2는 담배 모자이크 바이러스의 5' 선도서열(leading sequence), 즉 오 서열이며;
- [0219] Z3은 아데닌 테옥시뉴클레오타이드의 올리고머 사슬 [oligo(A)]_n이고;
- [0220] Z4는 번역 개시 코돈이며;
- [0221] Z5는 세린 코돈이고;
- [0222] Z6은 외래성 단백질의 암호화 서열이며;
- [0223] 상기 Z3, Z4 및 Z5는 Kozak 서열을 구성하고, Kozak 서열은 효모로부터 유래된 것이다.
- [0224] 본 발명에 있어서, 외래성 단백질의 암호화 서열의 선택은 특별히 제한되지 않는다. 일반적으로, 외래성 단백질의 암호화 서열은 루시페린, 또는 루시페라아제(예: 반딧불이 루시페라아제), 녹색형광단백질, 황색형광단백질, 아미노아실-tRNA합성효소, 글리세르알데히드-3-인산수소이탈효소, 카탈라아제, 및 액틴 및 항체의 가변영역의 암호화하는 외래성DNA, 루시페라아제 돌연변이의 DNA, 및 이들의 조합을들로 구성된 군으로부터 선택된다.
- [0225] 외래성 단백질의 암호화 서열은 또한 α-아밀라아제, 엔테로신 A, C형 간염 바이러스 E2 당단백질, 인슐린 전구체, 인터페론 αA, 인터루킨-1β, 라이소자임, 혈청 알부민, 항체의 단쇄가변분절(scFv), 트랜스타이레틴, 티로시나아제, 자일라나아제 및 이들의 조합으로 구성된 군으로부터 선택된 단백질을 암호화(encode)할 수 있다.
- [0226] 또한, 본 발명의 핵산 구조체는 선형 또는 원형일 수 있다. 본 발명의 핵산 구조체는 단일-가닥 또는 이중-가닥일 수 있다. 본 발명의 핵산 구조체는 DNA, RNA, 또는 DNA/RNA 잡종(DNA/RNA hybrid)일 수 있다.
- [0227] 바람직한 실시예에서, 본 발명의 핵산 구조체의 서열은 서열번호 2-17(SEQ ID NO.: 2-17) 중 어느 하나이다.
- [0228] 바람직한 실시예에서, 산 구조체의 서열은 서열번호 2-9(SEQ ID NO.: 2-9) 중 어느 하나이다.

- [0229] 바람직한 실시예에서, 핵산 구조체의 서열은 서열번호 3, 4 또는 6(SEQ ID NO.: 3, 4 or 6)이다.
- [0230] 바람직한 실시예에서, 본 발명의 핵산 구조체의 서열은 서열번호 85-87(SEQ ID NO.: 85-87) 중 어느 하나이다.
- [0231] 다른 바람직한 실시예에서, 구조체는 프로모터, 터미네이터(terminator), poly(A)요소, 운반 요소(transport elements), 유전자 표적화 요소(gene targeting elements), 선별마커 유전자(selection marker genes), 인헨서, 저항성 유전자 및 전위효소-암호화 유전자(transposase-encoding genes)로 구성된 군에서 선택된 요소 또는 조합을 더 포함한다.
- [0232] 본 발명에서 다양한 선별 마커(selectable marker) 유전자가 이용될 수 있으며, 영양요구성 마커(auxotrophic markers), 저항성 마커(resistance markers), 및 리포터 유전자(reporter gene) 마커를 포함할 수 있으나, 이에 제한되지 않는다. 선별 마커의 적용은 수용체 세포를 형질전환 되지 않은 세포(untransformed cell)와 현저하게 구별될 수 있도록 재조합 세포(recons)의 스크리닝(screening)에 역할을 한다. 영양요구성 마커는 전달된 마커 유전자의 도움으로 수용체 세포의 돌연변이 유전자에 상보적일 수 있어서, 수용체 세포는 야생형(wild-type) 성장을 나타낸다. 저항성 마커는 저항성 유전자가 수용체 세포 내로 전달되고, 전달된 유전자는 수용체 세포가 특정 약물 농도에서 약물 내성을 나타내도록 한다. 본 발명의 바람직한 형태로서, 저항성 마커는 재조합 세포의 편리한 스크리닝을 달성하기 위해 이용된다.
- [0233] 본 발명에 있어서, 본 발명의 효모-기반 시험관 내 생합성 시스템에서(예컨대, 효모-기반 단백질 생합성 시스템) 본 발명의 핵산 구조체의 적용은 외래성 단백질 번역의 효율을 현저히 개선시킬 수 있다. 특히, 본 발명의 핵산 구조체를 이용함으로써 합성된 루시페라아제의 활성을 나타내는 상대적 발광 수치 값은 매우 높으며, 본 발명의 핵산 구조체를 이용할 때(예컨대, KINCE102-Ω-10A) 상기 상대적 발광 수치 값은 Ω-10A 서열을 이용할 때의 1.65배였다.
- [0234] **벡터, 유전자 변형 세포**
- [0235] 본 발명은 또한 벡터 또는 본 발명의 핵산 구조체를 포함하는 벡터 조합물을 제공한다. 바람직하게는, 벡터는 세균 플라스미드, 파지(즉, 박테리오파지), 효모 플라스미드, 동물 세포 벡터 및 트랜스포존 벡터(transposon vector)인 셔틀 벡터(shuttle vectors)로 구성된 군에서 선택된다. 재조합 벡터의 제조 방법은 당업자에게 널리 공지되어 있다. 숙주에서 복제되고 안정될 수 있는 한 임의의 플라스미드 또는 벡터가 사용될 수 있다.
- [0236] 당업자는 공지된 방법을 사용하여 본 발명의 프로모터 및/또는 목적 유전자 서열을 포함하는 발현 벡터를 구축할 수 있다. 이러한 방법은 시험관 내 재조합 DNA 기술, DNA 합성 기술, 시험관 내 재조합 기술 등을 포함한다.
- [0237] 본 발명은 또한 유전자 변형 세포를 제공한다. 유전자 변형 세포는 구조체, 벡터 및 벡터 조합물을 포함하고; 또는 유전자 변형 세포의 염색체는 구조체 또는 벡터와 통합된다. 다른 바람직한 실시예에서, 유전자 변형 세포는 전위효소-암호화 유전자를 포함하는 벡터를 포함하고, 또는 유전자 변형 세포의 염색체는 전위효소 유전자와 통합된 것이다.
- [0238] 바람직하게는, 유전자 변형 세포는 진핵 세포이다.
- [0239] 다른 바람직한 실시예에서, 진핵 세포는 인간 세포, 중국 햄스터 난소 세포, 곤충 세포, 밀 배아 세포, 토끼 망상적혈구 세포 및 다른 고등 진핵 세포를 포함할 수 있으나, 이에 제한되지 않는다.
- [0240] 다른 바람직한 실시예에서, 진핵 세포는 효모 세포(바람직하게는, 클루이베로마이세스(*Kluyveromyces*) 세포, 더 바람직하게는 클루이베로마이세스 락티스(*Kluyveromyces lactis*) 세포를 포함할 수 있으나, 이에 제한되지 않는다.
- [0241] 본 발명의 구조체 또는 벡터는 적절한 유전자 변형 세포를 형질전환하는데 이용할 수 있다. 유전자 변형 세포는 원핵 세포, 예컨대 대장균(*E. coli*), 스트렙토마이세스(*Streptomyces*), 아그로박테리움(*Agrobacterium*); 또는 하등 진핵 세포(lower eukaryotic cell), 예컨대 효모 세포; 또는 고등 동물 세포, 예컨대 곤충 세포일 수 있다. 적절한 벡터 및 유전자 조작된 세포를 선택하는 방법은 당업자에게 알려져 있다. 재조합 DNA로 유전자 조작된 세포의 형질 전환은 당업자에게 잘 알려진 통상적인 기술을 이용하여 수행될 수 있다. 숙주가 원핵 생물(예컨대, 대장균)일 때, CaCl₂ 방법 또는 전기천공 방법을 이용하여 처리할 수 있다. 숙주가 진핵 생물일 때, DNA 형질 주입 방법은 인산 칼슘 공침(calcium phosphate coprecipitation) 방법 및 종래의 기계적 방법(예컨대, 미세주입, 전기천공, 리포솜 패키징(liposome packaging))이 선택적으로 이용될 수 있다. 식물 세포의 형질전환은 아그로박테리움-매개 형질전환 또는 유전자 총 형질전환, 예를 들어 리프 디스크(leaf disc) 방법, 미성숙 배아

세포 형질전환 방법, 플라워 버드소킹(flower bud soaking)방법 등과 같은 방법을 이용함으로써 수행될 수 있다.

[0242] **시험관 내 고수율 단백질 합성 방법**

[0243] 본 발명은 다음 단계를 포함하는 시험관 내 고수율 단백질 합성 방법을 제공한다:

[0244] (i) 진핵 생물-기반 시험관 내 생합성 시스템의 존재하에, 본 발명의 첫 번째 측면 내지 네 번째 측면으로부터 선택된 핵산 구조체를 제공하는 단계; 및

[0245] (ii)적합한 조건 하에서, 단계 (i)의진핵 생물-기반 시험관 내 생합성 시스템을 T1 기간 동안 인큐베이션하여 외래성 단백질을 합성하는 단계.

[0246] 다른 바람직한 실시예에서, 상기 방법은 (iii) 진핵 생물-기반 시험관 내 생합성 시스템의 외래성 단백질을 선택적으로 분리 또는 검출하는 단계를 더 포함한다.

[0247] 본 발명은 특정 실시예와 함께 하기에 추가로 설명한다. 이들 실시예는 단지 본 발명을 설명하기 위해 사용되며 본 발명의 범위를 제한하는 것이 아님을 이해해야한다. 하기 실시예에서 구체적으로 기술된 조건 이없는 실험 방법과 관련하여, 당업자는(one person) 일반적으로 Sambrook et al., Molecular Cloning : A Laboratory Manual (New York : Cold Spring Harbor Laboratory Press, 1989)에 설명된 조건과 같은 조건을 따르거나 제조 업체가 권장하는 조건을 따를 수 있다. 달리 언급되지 않는 한, 백분율 및 부분은 중량 백분율 및 중량부를 지칭한다.

[0248] 달리 명시되지 않는 한, 본 발명의 실시예에서 사용된 물질 및 시약은 모두 시판 제품이다.

[0249] **실시예 1: 진핵 세포의내인성 IRES, Ω 서열 및 클루이베로마이세스 락티스-특이적(specific) Kozak 서열을 탠덤연결하는 DNA 요소의 설계**

[0250] 1.1 사카로마이세스 세레비제 및 클루이베로마이세스 락티스(*Kluyveromyces lactis*)에서 내인성 IRESs의 확인: 클루이베로마이세스 락티스의4 개의 내인성 IRESs 및 사카로마이세스 세레비제(*Saccharomyces cerevisiae*)에서 상동성 단백질에 해당하는 IRES (표 1 참조)는 시험관 내 단백질 합성을 개시할 수 있다. IRESs 중 합성을 개시한 Fluc, 6개의 IRESs(K1FLO8, K1MSN1, K1NCE102, ScFLO8, ScMSN1 및 ScNCE102)는 종래의 Ω 서열보다 높은 상대적 발광 수치(RLU)값을 나타내고, 다른 두 개 (K1BOI1 및 ScBOI1)는 Ω 서열보다 낮은 RLU 값을 나타낸다. 이와 같은 8개의 IRESs는 Ω 서열 및 Kozak 서열이 탠덤 연결된 것으로 확인되었다.

[0251] 1.2 16개의 탠덤 요소 확인: 8개의 내인성 IRESs를 Ω 서열 및 클루이베로마이세스 락티스-특이적 Kozak 서열과 탠덤연결시키고, Ω 서열과 관련된 IRES의 업스트림 또는 다운스트림의 위치를 변화시켜; 결과적으로 총 16개의 탠덤 DNA 요소를 설계하였다(K1FLO8, K1MSN1, K1NCE102, K1BOI1, ScFLO8, ScMSN1, ScNCE102 및 ScBOI1-Ω-10A, 뿐만 아니라, Ω-K1FLO8, K1MSN1, K1NCE102, K1BOI1, ScFLO8, ScMSN1, ScNCE102 및 ScBOI1-10A). 서열 번호 2-17(SEQ ID No.:2-17)에 나타난 탠덤 요소의 서열은 각각 상기 탠덤 DNA 요소(K1FLO8, K1MSN1, K1NCE102, K1BOI1, ScFLO8, ScMSN1, ScNCE102 및 ScBOI1-Ω-10A, 뿐만 아니라, Ω-K1FLO8, K1MSN1, K1NCE102, K1BOI1, ScFLO8, ScMSN1, ScNCE102 및 ScBOI1-10A)의 서열에 상응한다.

[0252] 1.3 탠덤 요소 및 리porter 단백질 유전자(reporter protein gene)(Fluc)의 설계;상기 설계된 16개의 탠덤 요소를 기존의 Ω-10A-Fluc 플라스미드(KangMa-Healthcode(Shanghai) Biotech Co., Ltd에서 입수, 및 Ω-10A 서열은 서열번호 1(SEQ ID No.:1)에 나타내었다)에 삽입하여 Ω-10A를 대체하고 16개의 새로운 플라스미드를 형성하고;이에 따라 생성된 16개의 새로운 플라스미드는 각각 K1FLO8-Ω-10A-Fluc, K1MSN1-Ω-10A-Fluc, K1NCE102-Ω-10A-Fluc, K1BOI1-Ω-10A-Fluc, ScFLO8-Ω-10A-Fluc, ScMSN1-Ω-10A-Fluc, ScNCE102-Ω-10A-Fluc, ScBOI1-Ω-10A-Fluc, Ω-K1FLO8-10A-Fluc, Ω-K1MSN1-10A-Fluc, Ω-K1NCE102-10A-Fluc, Ω-K1BOI1-10A-Fluc, Ω-ScFLO8-10A-Fluc, Ω-ScMSN1-10A-Fluc, Ω-ScNCE102-10A-Fluc 및 Ω-ScBOI1-10A-Fluc이다. 이들 중 K1NCE102-Ω-10A-Fluc, ScFLO8-Ω-10A-Fluc 및 K1MSN1-Ω-10A-Fluc의 서열은 각각 서열 번호 85-87에 나타내었다.

표 1

[0253] 사카로마이세스 세레비제 및 클루이베로마이세스 락티스 관련 유전자

유전자명	오픈리딩프레임(ORFs)
<i>ScBOI1</i>	YBL085w
<i>ScFLO8</i>	YER109c

<i>ScNCE102</i>	YPR149w
<i>ScMSN1</i>	YOL116w
<i>K1BOI1</i>	KLLA0E20879g
<i>K1FLO8</i>	KLLA0E20725g
<i>K1NCE102</i>	KLLA0D16280g
<i>K1MSN1</i>	KLLA0A07337g

[0254] 실시예 2: 시험관 내 단백질 합성 시스템에 대한 탠덤 DNA 요소를 포함하는 플라스미드의 설계(construction)

[0255] 2.1 플라스미드의 설계: 8개의 내인성 IRESs를 각각 Ω-10A-Fluc 플라스미드에 삽입하고, 이는 오서열의 업스트림 또는 다운스트림에 각각 위치한다. 구체적으로 사용한 프라이머는 표 2에 나타내었다.

[0256] 하나의 플라스미드에 대한 구체적인 설계 과정은 다음과 같다:

[0257] 삽입될 IRES 단편 및 Ω-10A-Fluc 벡터 플라스미드를 2 쌍의 프라이머를 사용하여 PCR에 의해 각각 증폭시키고, 각각의 PCR 증폭 산물 10 μL를 취합하여 함께 혼합하였다. 이어서, 1 μLDpnI를 증폭 산물의 20 μL 혼합물에 첨가한 후 37 °C에서 6 시간 동안 인큐베이션하였다. 그 후, DpnI-처리된 생성물 4 μL를 DH5 α 수용 세포(competent cell) 50 μL에 첨가 하였다. 혼합물을 얼음상에 30분 동안 두고 42°C에서 45초 동안 열충격(heat-shock)시켰다. 그 후, 혼합물을 Amp 항생제를 포함하는 LB 고체 배지에 코팅하고 밤새 배양하였다. 증식 배양(proliferation culture)을 수행하기 위해 6 개의 단일 클론 콜로니를 골랐다. 시퀀싱(sequencing)으로 정확성을 확인한 후, 플라스미드를 추출하여 보관하였다.

표 2

[0258] PCR 증폭을 위한 프라이머

플라스미드 명	벡터 증폭 프라이머 명	SEQ ID NO.:	IRES 증폭 프라이머 명	SEQID NO.:
K1FLO8-Ω-10A-Fluc	PF_FLO8KL_Omega	18	PF_T7pro_FLO8KL	20
	PR_FLO8KL_T7pro	19	PR_Omega_FLO8KL	21
K1MSN1-Ω-10A-Fluc	PF_MSN1KL_Omega	22	PF_T7pro_MSN1KL	24
	PR_MSN1KL_T7pro	23	PR_Omega_MSN1KL	25
K1NCE102-Ω-10A-Fluc	PF_KLNCE102_Omega	26	PF_T7pro_KLNCE102	28
	PR_KLNCE102_T7pro	27	PR_Omega_KLNCE102	29
K1BOI1-Ω-10A-Fluc	PF_BOI1KL_Omega	30	PF_T7pro_BOI1KL	32
	PR_BOI1KL_T7pro	31	PR_Omega_BOI1KL	33
ScFLO8-Ω-10A-Fluc	PF_FLO8_Omega	34	PF_T7pro_FLO8	36
	PR_FLO8_T7pro	35	PR_Omega_FLO8	37
ScMSN1-Ω-10A-Fluc	PF_MSN1_Omega	38	PF_T7pro_MSN1	40
	PR_MSN1_T7pro	39	PR_Omega_MSN1	41
ScNCE102-Ω-10A-Fluc	PF_NCE102_Omega	42	PF_T7pro_NCE102	44
	PR_NCE102_T7pro	43	PR_Omega_NCE102	45
ScBOI1-Ω-10A-Fluc	PF_BOI1_Omega	46	PF_T7pro_BOI1	48
	PR_BOI1_T7pro	47	PR_Omega_BOI1	49
Ω-K1FLO8-10A-Fluc	O-Flo8(KL)_PF-2	50	Flo8(KL)_F	52
	O-Flo8(KL)-PR-2	51	Flo8(KL)_R	53
Ω-K1MSN1-10A-Fluc	PF_pET21a_KLMSN1_10A	54	PF_KLMSN1_Omega_pET21a	56
	PR_pET21a_KLMSN1_Omega	55	PR_KLMSN1_10A_pET21a	57
Ω-K1NCE102-10A-Fluc	O-IRES-PF	58	NCE102(KL)-F	60
	O-IRES-PR	59	NCE102(KL)-R	61
Ω-K1BOI1-10A-Fluc	PF_pET21a_10A_KLBOI	62	PF_KLBOI_Omega_pET21a	64
	PR_pET21a_Omega_KLBOI	63	PR_KLBOI_10A_pET21a	65
Ω-ScFLO8-10A-Fluc	O-FLO8-PF-2	66	Flo8_F	68
	O-FLO8-PR-2	67	Flo8_R	69
Ω-ScMSN1-10A-Fluc	PF_pET21a_MSN1_10A	70	PF_MSN1_Omega_pET21a	72
	PR_pET21a_MSN1_Omega	71	PR_MSN1_10A_pET21a	73
Ω-ScNCE102-10A-Fluc	O-IRES-PF2	74	NCE102-F	76
	O-IRES-PR2	75	NCE102-R	77

Ω-ScBOI1-10A-F luc	PF_pET21a_10A_BOI1	78	PF_BOI1_Omega_pET21a	80
	PR_pET21a_Omega_BOI1	79	PR_BOI1_10A_pET21a	81

[0259] **실시예 3: 효모-기반 시험관 내 단백질 합성 시스템에서 탠덤DNA 요소의 적용**

[0260] 3.1 T7 전사 개시 서열 및 종결 서열 사이에 있는 모든 플라스미드에서 탠덤 DNA 요소 및 Fluc를 포함하는 단편이 T7_pET21a_F:CGCGAAATTAATACGACTCACTATAGG (서열 번호 : 82) 및 T7ter_pET21a_R : TCCGGTCAGTT (TCCTTTCAGTT)(서열 번호 : 83)를 프라이머로 사용하여 PCR 법에 의해 증폭되었다.

[0261] 증폭된 DNA 단편을 에탄올 침전법을 이용하여 정제 및 농축하였다: 1/10 부피의 3M 아세트산 나트륨(sodium acetate)(pH 5.2)을 PCR 생성물에 첨가하고, 이어서 부피의 2.5-3배의 95% 에탄올을 첨가하고(아세트산 나트륨 첨가 후의 부피에 비해), 15 분 동안 얼음상에서 인큐베이션하였다. 그 후, 혼합물을 실온에서 14000g 보다 높은 속도로 30분 동안 원심분리하고, 상층액(supernatant)을 제거하였다. 침전물을 70 % 에탄올로 세척한 다음, 15 분 동안 원심 분리한 후 상층액을 제거하였다. 침전물을 초순수(ultrapure water)로 용해시켜 DNA 농도를 측정하였다.

[0262] 3.2 사용 지침에 따라, 정제된 DNA 단편을자가-제조 클루이베로마이세스 락티스 기반 시험관 내 단백질 합성 시스템에 첨가하였다. 상기 반응 시스템을 25-30℃의 환경에 두고, 혼합물(mixture stand)을 약 2-6시간 동안 인큐베이션 시켰다. 완료 후, Fluc용 기질로서 동일한 부피의 루시페린(luciferin)을 96-웰 또는 384-웰 화이트 플레이트의 웰에 첨가하고, 이를 즉시 Envision 2120 다기능 마이크로 플레이트리더(multifunctional microplate reader) (Perkin Elmer)에 넣고, Fluc 활성을 검출하기 위해 흡광도를 판독(read)하였으며, 여기서 활성의 단위는 도 2에 도시된 상대적 발광 수치(Relative Light Unit, RLU) 값이다.

[0263] 3.3 진핵 세포의 내인성 IRES 서열을 제외한 DNA 단편(Ω-10A-Fluc) 그룹은 대조군으로 이용하였고, DNA주형을 첨가하지 않은 반응 그룹은 음성 대조군(Negative control, NC)으로 이용하였다. 각 그룹마다 3개의 독립적인 샘플을 사용하였다.

[0264] **실험 결과**

[0265] 1. 진핵 세포의 내인성 IRES 및 클루이베로마이세스 락티스-특이적 Kozak 서열을 포함하는 Ω 서열을 탠덤 연결하는 DNA 요소의 설계

[0266] 총 16개의 탠덤 DNA 요소가 설계되었고, 16개의 요소 모두가 Ω-10A-Fluc 플라스미드에 각각 삽입되어 Ω-10A 서열을 대체하고, 시험관 내 단백질 합성에 사용된 16개의 플라스미드가 형성되었다.

[0267] 2. 시험관 내 단백질 합성 시스템을 위한 탠덤 DNA요소를 포함하는 플라스미드의 설계

[0268] 많은 시도 후에, 시험관 내 단백질 합성 시스템을 위한 총 16개의 플라스미드가 성공적으로 설계되었다.

[0269] 3. 효모-기반 시험관 내 단백질 합성 시스템에서 탠덤DNA 요소의 적용

[0270] 도 3에 나타난 바와 같이, 효모-기반 시험관 내 단백질 합성 시스템에서 반딧불이 루시페라아제(firefly luciferase, Fluc)의 상대적 발광 수치(RLU) 값과 관련하여, 선택된 3개의 탠덤 DNA 요소 중, KINCE102-Ω-10A에 의한 RLU 값만이 Ω-10A 서열에 의한 RLU 값을 초과하였다; 상기에서, KINCE102-Ω-10A에 의한 상대적 발광 수치 값은 1.67×10^9 에 도달하였으며, 이는 Ω-10A 서열에 비해 1.65 배이다(Ω-10A 서열에 의한 상대적 발광 수치 값은 1.01×10^9 임). 또한, ScFLO8-Ω-10A 및 KIMSN1-Ω-10A에 의한 상대적 발광 수치 값은 Ω-10A 서열과 비슷하고, Ω-10A 서열에 의한 RLU 값의 81.68 % 및 85.35 %에 각각 도달하였다.

[0271] 검출 장치를 통한 상대적 발광 수치(RLU) 값 및 단백질 농도 사이의 선형 관계의 범위 내에서, 가장 높은 활성을 나타내는 KINCE102-Ω-10A에 의한 상대적 발광 수치값은 Ω-10A 서열에 의한 것의 1.65배였으며, 이는 탠덤 DNA 요소가 단백질 합성을 약 1.65 배로 향상시킬 수 있음을 나타낸다.

[0272] 본 발명에서 나타내는 결과는: 진핵 세포의 내인성 IRES 및 클루이베로마이세스 락티스-특이적 Kozak 서열을 포함하는 Ω 서열의 탠덤 연결은 단백질 합성의 효율을 향상시킬 수 있고; 이 둘은 서로 상호작용 할 수 있는 번역 개시 인자, Pab1 및 eIF4G를 모집하여 단백질 합성 효율을 증진시키는데 시너지 효과를 얻을 수 있으며(도 4 참조), 및 단백질 합성 개시 효율이 일반적으로 사용되는 Ω-10A 서열 및 Kozak 서열을 포함하는 탠덤DNA 요소를 초과할 수 있는 효모-기반 시험관 내 단백질 합성 시스템에 적용할 수 있다. KINCE102-Ω-10A에 의해 합성이

개시되어 합성된 단백질의 양은 Ω -10A 서열에 의한 것의 1.65 배이고; KINCE102- Ω -10A은 효모-기반 시험관 내 단백질 합성 시스템의 단백질 번역 효율을 높였으며, 클루이베로마이세스 락티스 기반 시험관 내 합성 시스템이 단백질 합성을 개시하도록 번역 요소의 선택성을 증가시켰고, 클루이베로마이세스 락티스 기반 시험관 내 단백질 합성 시스템의 유용성을 크게 향상시켰다.

[0273] 더욱이, 실험 결과에 따르면, 본 발명은 또한 탠덤 DNA 요소의 업스트림에 GAA의 3개의 잔기를 배치하는 것(전사된 mRNA의 5' 말단이 GAA임)이 또한 단백질 합성 효율을 향상시킬 수 있음을 발견하였다. 시험관 내 합성을 매개하기 위해 본 발명의 핵산 구조체 및 GAA의 탠덤을 이용한 루시페라아제의 상대적 발광 수치 값은 본 발명의 핵산 구조체를 단독으로 사용할 때의 RLU 값의 1.28배 였다(도 5 참조).

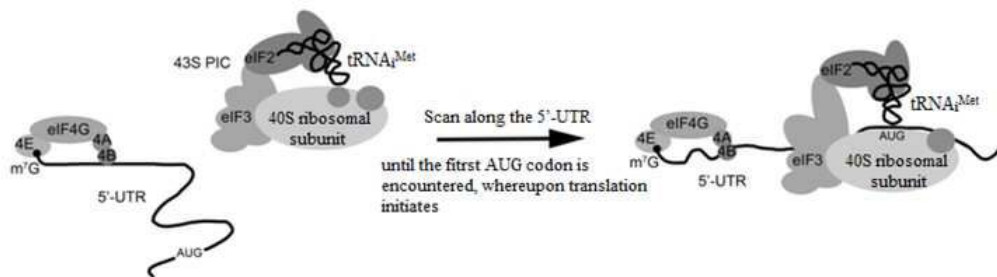
[0274] 게다가, 실험 결과에 따르면, 본 발명은 본 발명의 핵산 구조체(KINCE102- Ω -10A)에 의한 상대적 발광 수치 값은 KINCE102에 의한 것보다 1.23배였다(도 5 참조).

[0275] 본 발명에서 언급된 모든 문서는 각각의 문서가 개별적으로 참조로 포함되는 것처럼, 본 발명은 본 출원에서 참조로 포함된다. 또한, 당업자는 전술한 본 발명의 내용을 읽은 후 본 발명을 다양하게 변경 또는 수정할 수 있으며, 이러한 유사한 형태는 본 발명의 첨부된 청구항에 의해 정의된 범위 내에 속한다는 것을 이해해야 한다.

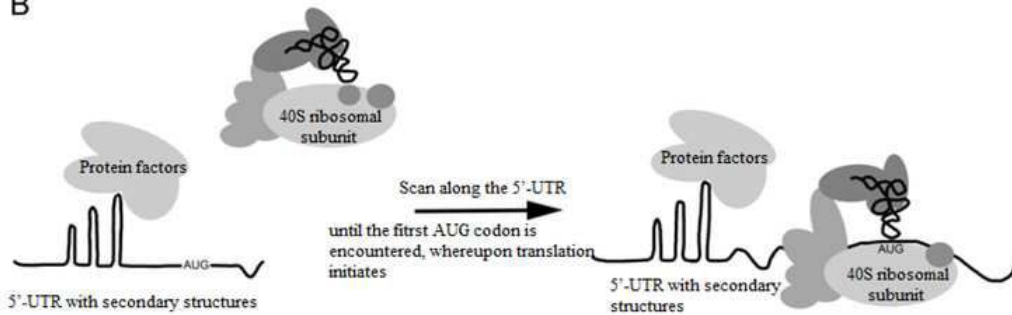
도면

도면1

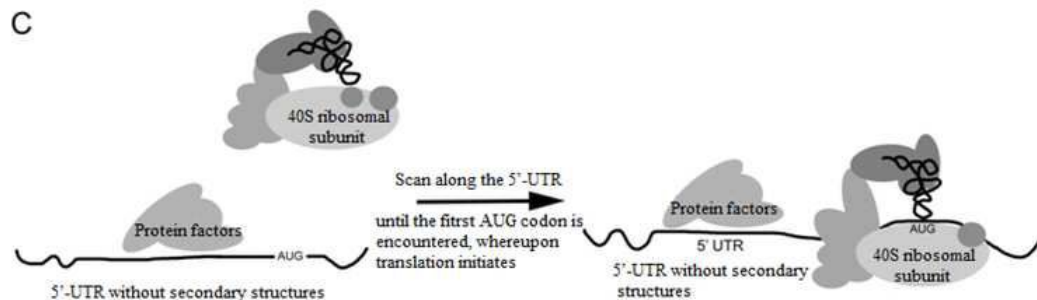
A



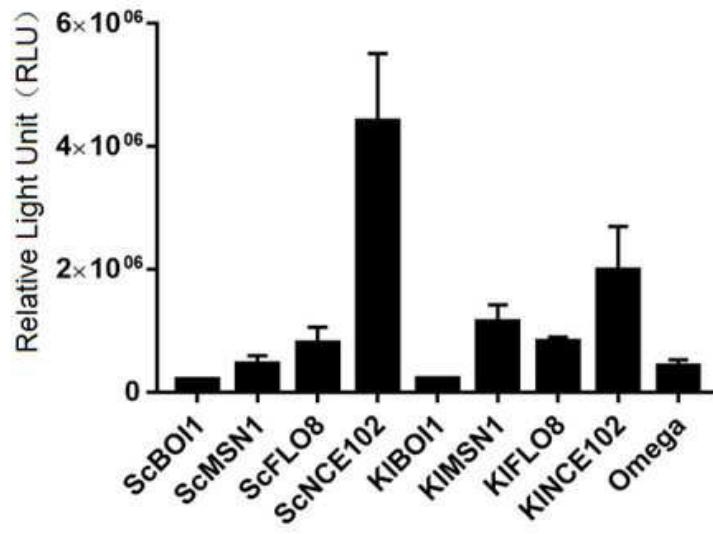
B



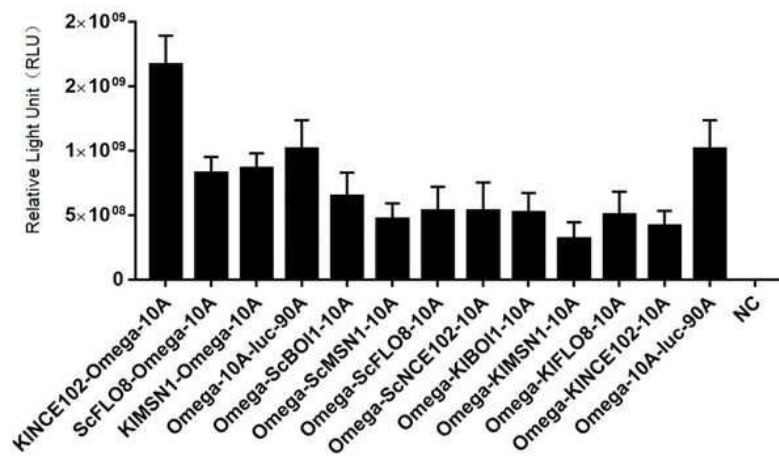
C



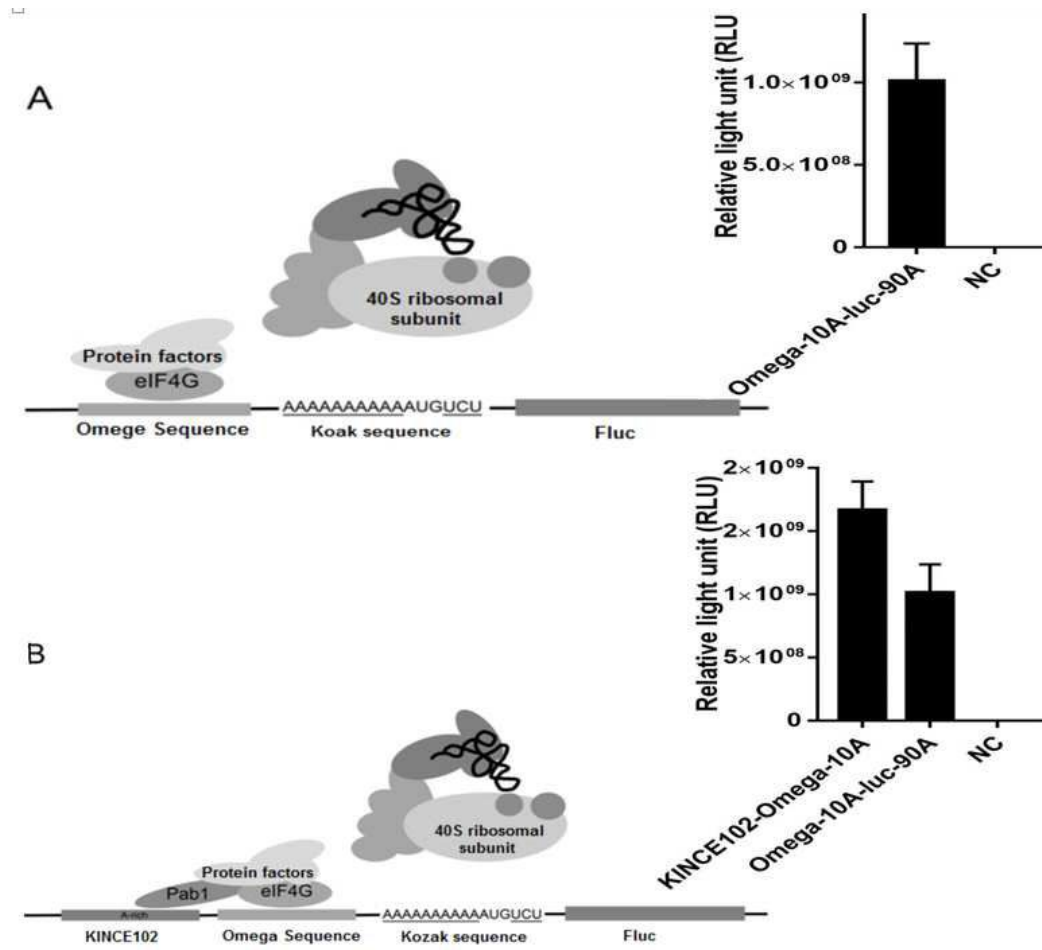
도면2



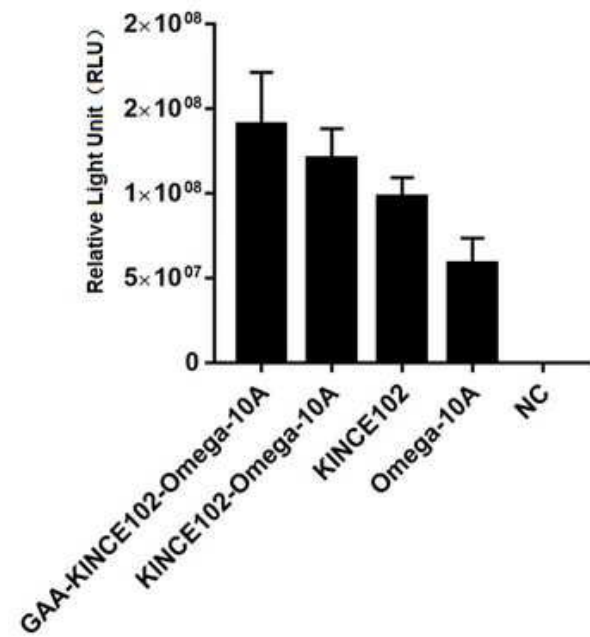
도면3



도면4



도면5



서열 목록

<110> KANGMA-HEALTHCODE (SHANGHAI) BIOTECH CO., LTD

<120> Tandem DNA Element Capable of Enhancing Protein Synthesis
Efficiency

<130> IP2004IPW

<150> CN 201711194355.9

<151> 2017-11-24

<160> 87

<170> KoPatentIn 3.0

<210> 1

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 1

ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattt 60

acaattacaa aaaaaaaaaa tgtct 85

<210> 2

<211> 146

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 2

gagatagaga gagagattcc gtgtttttgt ctctgttct ttacacacgt attttcaacg 60

cgggtattttt acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120

tacaattaca aaaaaaaaaa atgtct 146

<210> 3

<211> 147

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 3

aacatagaaa tcatcatttt aattgattcg gtgttttcga gtctagtatt gattgtttaa 60

tcggtattttt tacaacaatt accaacaaca acaacaaca aacaacatta caattactat 120

ttacaattac aaaaaaaaaa aatgtct 147

<210> 4
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 4
 aaaagaaatc tctcaagctg aaattaaacc aaaactctaa tataagaaaa aaaaatagaa 60
 aggtattttt acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
 tacaattaca aaaaaaaaaa atgtct 146

<210> 5
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 5
 catttagagc atctcagccg tatcataccg ttcgagttcc agaaatacct acaaaatcag 60
 tgggtattttt acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
 tacaattaca aaaaaaaaaa atgtct 146

<210> 6
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 6
 aaaaaataaa cacgaagacg tttatagaca taaataaaga ggaaacgcat tccgtggtag 60
 aggtattttt acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
 tacaattaca aaaaaaaaaa atgtct 146

<210> 7
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 7

ccttgcttat aagaaaagaa accaaatcag aaaaggagat tatttcaagg taggcacga 60
aggtatTTTT acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
tacaattaca aaaaaaaaaa atgtct 146

<210> 8
<211> 146
<212> DNA
<213> Artificial Sequence
<220><223> Artificial sequence
<400> 8

gaaaaatcgg ttaaaaaaac ttttcttctc aaagcatacc taataacaat ataatcccat 60
aggtatTTTT acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
tacaattaca aaaaaaaaaa atgtct 146

<210> 9
<211> 146
<212> DNA
<213> Artificial Sequence
<220><223> Artificial sequence
<400> 9

aatttcaaca aagttctaac tcgaggtgac cggaggccac tgtaataata aaaaatagaa 60

gggtatTTTT acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
tacaattaca aaaaaaaaaa atgtct 146

<210> 10
<211> 146
<212> DNA
<213> Artificial Sequence
<220><223> Artificial sequence
<400> 10

ggtatTTTT caacaattac caacaacaac aaacaacaaa caacattaca attactattt 60
acaattacag agatagagag agagattccg tgtttttgtc tccgtttctt tacacacgta 120
ttttcaacgc aaaaaaaaaa atgtct 146

<210> 11
<211> 147

<212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 11
 ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattd 60
 acaattacaa acatagaaat catcatttta attgattcgg tgttttcgag tctagtattg 120
 attgtttaat caaaaaaaaaa atgtct 147
 <210> 12
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 12
 ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattd 60
 acaattacaa aaagaaatct ctcaagctga aattaaacca aaactctaataaagaaaaa 120
 aaaatagaaa aaaaaaaaaa atgtct 146
 <210> 13
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 13
 ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattd 60
 acaattacac atttagagca tctcagccgt atcataccgt tcgagttcca gaaataccta 120
 caaatcagt aaaaaaaaaa atgtct 146
 <210> 14
 <211> 146
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 14
 ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattd 60
 acaattacaa aaaaataaac acgaagacgt ttatagacat aaataaagag gaaacgcatt 120

ccgtggtaga aaaaaaaaa atgtct	146
<210> 15	
<211> 146	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 15	
ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattt	60
acaattacac ctgtcttata agaaaagaaa ccaaatcaga aaaggagatt atttcaagg	120
aggcatcgaa aaaaaaaaa atgtct	146
<210> 16	
<211> 146	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 16	
ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattt	60
acaattacag aaaaatcggg taaaaaaac tttcttctca aagcatacct aataacaata	120
taatccata aaaaaaaaa atgtct	146
<210> 17	
<211> 146	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 17	
ggtattttta caacaattac caacaacaac aaacaacaaa caacattaca attactattt	60
acaattacaa atttcaacaa agttctaact cgagggtgacc ggaggccact gtaataataa	120
aaaatagaag aaaaaaaaa atgtct	146
<210> 18	
<211> 34	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	

<400>	18	
cgtattttca acgcggtatt tttacaacaa ttac		34
<210>	19	
<211>	35	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	19	
ctctctctct atctccctat agtgagtcgt attaa		35
<210>	20	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	20	
actcactata gggagataga gagagagatt cc		32
<210>	21	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	21	
tgtaaaaata ccgcgttgaa aatacgtgtg ta		32
<210>	22	
<211>	36	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	22	
gattgtttaa tcggtatttt tacaacaatt accaac		36
<210>	23	
<211>	32	
<212>	DNA	

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 23

ctatgttcct atagtgagtc gtattaattt cg 32

<210> 24

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 24

acgactcact ataggaacat agaaatcatc attttaa 37

<210> 25

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 25

gttgtaaaaa taccgattaa acaatcaata ctagactc 38

<210> 26

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 26

tagaaaggta tttttacaac aattaccaac aacaac 36

<210> 27

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 27

gagatttctt ttctatagt gagtcgtatt aatttcgc 38

<210> 28

<211> 33
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 28
 gactcactat aggaaaagaa atctctcaag ctg 33
 <210> 29
 <211> 42

 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 29
 gtaattgttg taaaaatacc tttctat ttttcttat at 42
 <210> 30
 <211> 35
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 30
 cctacaaaat cagtggatt tttacaacaa ttacc 35
 <210> 31
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 31
 gctctaatg cctatagtga gtcgtattaa tt 32

 <210> 32
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 32

tcactatagg catttagagc atctcagccg	30
<210> 33	
<211> 38	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 33	
ttgttgtaaa aataccactg atttttagg tatttctg	38
<210> 34	
<211> 30	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 34	
ccgtggtaga ggtattttta caacaattac	30
<210> 35	
<211> 34	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 35	
cggtgttatt ttttctata gtgagtcgta ttaa	34
<210> 36	
<211> 31	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 36	
ctcactatag gaaaaaataa acacgaagac g	31
<210> 37	
<211> 30	
<212> DNA	
<213> Artificial Sequence	

<220><223> Artificial sequence

<400> 37

taaaaatacc tctaccacgg aatgcgtttc 30

<210> 38

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 38

ggtaggcatac gaaggtatatt ttacaacaat tac 33

<210> 39

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 39

agcaaggcct atagtgagtc gtattaattt c 31

<210> 40

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 40

gactcactat aggccttgct tataagaaaa gaaac 35

<210> 41

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 41

gttgtaaaaa taccttcgat gcctaccttg aaat 34

<210> 42

<211> 36

<212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 42
 caatataatc ccataggtat ttttacaaca attacc 36
 <210> 43
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 43
 ccgatttttc cctatagtga gtcgtattaa tttc 34
 <210> 44
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 44
 cgactcacta tagggaaaaa tcggttaaaa aaac 34

 <210> 45
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 45
 gttgtaaaaa tacctatggg attatattgt tattagg 37
 <210> 46
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 46
 taataaaaaa tagaagggtat tttttacaac aattaccaac 40
 <210> 47

<211>	34	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	47	
ctttgttgaa attcctatag tgagtcgtat taat		34
<210>	48	
<211>	36	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	48	
tacgactcac tataggaatt tcaacaaagt tctaac		36
<210>	49	
<211>	40	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	49	
gtaattgttg taaaaatacc cttctatfff ttattattac		40
<210>	50	
<211>	38	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	50	
caacgcaaaa aaaaaaatgt ctgaagacgc caaaaaca		38
<210>	51	
<211>	41	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Artificial sequence	
<400>	51	

gaatctctct ctctatctct gtaattgtaa atagtaattg t	41
<210> 52	
<211> 44	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 52	
acaattacta tttaacaatta cagagataga gagagagatt ccgt	44
<210> 53	
<211> 55	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 53	
gtttttggcg tcttcagaca tttttttttt tgcgttgaaa atacgtgtgt aaaga	55
<210> 54	
<211> 33	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 54	
gtttaatcaa aaaaaaaaaat gtctgaagac gcc	33
<210> 55	
<211> 45	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 55	
gatttctatg ttgttaattg taaatagtaa ttgtaatgtt gtttg	45
<210> 56	
<211> 42	
<212> DNA	
<213> Artificial Sequence	

<220><223> Artificial sequence

<400> 56

caattactat ttacaattac aaacatagaa atcatcattt ta 42

<210> 57

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 57

cagacatttt tttttttgat taaacaatca atactag 37

<210> 58

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 58

atgtctgaag acgccaaaaa ca 22

<210> 59

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 59

tgtaattgta aatagtaatt gt 22

<210> 60

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 60

aattactatt tacaattaca aaaagaaatc tctcaagctg aa 42

<210> 61

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 61

gtttttggcg tcttcagaca tttttttttt ttttctattt ttttttctta tatt 54

<210> 62

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 62

cagtaaaaaa aaaatgtctg aagacgcca 29

<210> 63

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 63

gagatgctct aaatgtgtaa ttgtaaatag taattgtaat g 41

<210> 64

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 64

ctatttacia ttacacattt agagcatctc agccg 35

<210> 65

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 65

cagacatttt tttttttact gattttgtag gtatttctgg 40

<210> 66
 <211> 43
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 66
 tccgtggtag aaaaaaaaaa atgtctgaa gacgccaaaa aca 43
 <210> 67
 <211> 42
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 67
 cgtcttcgtg tttatttttt tgtaattgta aatagtaatt gt 42
 <210> 68
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 68
 acaattacta ttacaatta caaaaaata aacacgaaga cggt 44
 <210> 69
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 69
 tgtttttggc gtcttcagac attttttttt tttctaccac ggaatgcgtt tcc 53
 <210> 70
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 70

atcgaaaaaa aaaaaaatgt ctgaagacgc caaa	34
<210> 71	
<211> 39	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 71	
ataagcaagg tgtaattgta aatagtaatt gtaatgttg	39
<210> 72	
<211> 41	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 72	
caattactat ttacaattac accttgctta taagaaaaga a	41
<210> 73	
<211> 39	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 73	
cagacatfff ttttttttc gatgcctacc ttgaaataa	39
<210> 74	
<211> 31	
<212> DNA	
<213> Artificial Sequence	
<220><223> Artificial sequence	
<400> 74	
atgtctgaag acgccaaaa cataaagaaa g	31
<210> 75	
<211> 41	
<212> DNA	
<213> Artificial Sequence	

<220><223> Artificial sequence

<400> 75

tgtaattgta aatagtaatt gtaatgttgt ttgttgttg t 41

<210> 76

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 76

acaattacta tttaacaatta cagaaaaatc ggtaaaaaa actt 44

<210> 77

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 77

tgtttttggc gtcttcagac attttttttt ttatgggat tatattgta ttag 54

<210> 78

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 78

tagaagaaaa aaaaaaatgt ctgaagacgc c 31

<210> 79

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 79

gttgaaattt gtaattgtaa atagtaattg taatgttg 38

<210> 80

<211> 41

<212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 80
 caattactat ttacaattac aaatttcaac aaagttctaa c 41
 <210> 81
 <211> 41

 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 81
 tcttcagaca tttttttttt tcttctattt ttattattata c 41
 <210> 82
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 82
 cgcgaaatta atacgactca ctatagg 27
 <210> 83
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 83
 tccggatata gttcctcctt tcag 24

 <210> 84
 <211> 16
 <212> RNA
 <213> Artificial Sequence
 <220><223> Artificial sequence
 <400> 84
 aaaaaaaaaa augucu 16

<210> 85

<211> 1799

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 85

aaaagaaatc tctcaagctg aaattaaacc aaaactctaa tataagaaaa aaaaatagaa	60
aggtatTTTT acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt	120
tacaattaca aaaaaaaaaa atgtctgaag acgccaaaaa cataaagaaa ggccccggcgc	180
cattctatcc tctagaggat ggaaccgctg gagagcaact gcataaggct atgaagagat	240
acgccctggt tcttgaaca attgctTTTA cagatgcaca tatcgaggtg aacatcacgt	300
acgcggaata cticgaaatg tccgttcggt tggcagaagc tatgaaacga tatgggctga	360
atacaaatca cagaatcgtc gtatgcagtg aaaactctct tcaattcttt atgccggtgt	420
tgggcgcgtt atttatcgga gttgcagttg cggccgcgaa cgacatttat aatgaacgtg	480
aattgtcaa cagtatgaac atttcgcagc ctaccgtagt gtttgtttcc aaaaaggggt	540
tgcaaaaaat tttgaacgtg caaaaaaaat taccaataat ccagaaaatt attatcatgg	600
attctaaaac ggattaccag ggatttcagt cgatgtacac gttcgtcaca tctcatctac	660
ctccccgttt taatgaatac gatTTTgtac cagagtcctt tgatcgtgac aaaacaattg	720
caatgataat gaattcctct ggatctactg ggttacctaa ggggtgtggcc ctccgcata	780
gaactgcctg cgtcagattc tcgcatgcca gagatcctat ttttggcaat caaatcattc	840
cggatactgc gatTTtaagt gttgttccat tccatcacgg ttttggaaTG tttactacac	900
tcggatattt gatatgtgga tttcgagtcg tcttaatgta tagatttgaa gaagagctgt	960
ttttacgac ccttcaggat tacaaaattc aaagtgcgtt gctagtacca accctatTTT	1020
cattcttcgc caaaagcact ctgattgaca aatacgattt atctaattta cacgaaattg	1080
cttctggggg cgcacctctt tcgaaagaag tcggggaagc ggttgcaaaa cgcttccatc	1140
ttccagggat acgacaagga tatgggctca ctgagactac atcagctatt ctgattacac	1200
ccgaggggga tgataaacCG ggcgcggtcg gtaaaTgt tccatTTTT gaagcgaagg	1260
ttgtggatct ggataccggg aaaacgctgg gcgttaatca gagaggcgaa ttatgtgtca	1320
gaggacctat gattatgtcc ggttatgtaa acaatccgga agcgaccaac gccttgattg	1380
acaaggatgg atggctacat tctggagaca tagcttactg ggacgaagac gaacacttct	1440
tcatagtTga ccgcttgaag tctTTaatta aatacaaagg atatcaggTg gccccgctg	1500

aattggaatc gatattgtta caacacccca acatcttcga cgcgggcgtg gcaggcttc 1560
 ccgacgatga cgccggtgaa cttcccgccg ccgttgttgt tttggagcac ggaaagacga 1620
 tgacggaaaa agagatcgtg gattacgtcg ccagtcaagt aacaaccgag aaaaagttgc 1680
 gcggaggagt tigtgtttgtg gacgaagtac cgaaaggtct taccggaaaa ctgcagcaa 1740
 gaaaaatcag agagatcctc ataaaggcca agaaggcgg aaagtcctaa ttggtttaa 1799
 <210> 86
 <211> 1799
 <212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 86

aaaaataaa cacgaagacg tttatagaca taaataaaga ggaaacgcat tccgtggtag 60
 aggtatTTTT acaacaatta ccaacaacaa caaacaacaa acaacattac aattactatt 120
 tacaattaca aaaaaaaaa atgtctgaag acgcaaaaa cataaagaaa ggcccggcgc 180
 cattctatcc tctagaggat ggaaccgctg gagagcaact gcataaggct atgaagagat 240
 acgcctggt tctggaaca attgctttta cagatgcaca tatcgagggt aacatcacgt 300
 acgcggaata cttcgaaatg tccgttcggt tggcagaagc tatgaaacga tatgggctga 360

 atacaaatca cagaatcgtc gtatgcagtg aaaactctct tcaattcttt atgccggtgt 420
 tgggcgcgtt atttatcgga gttgcagttg cggccgcgaa cgacatttat aatgaacgtg 480
 aattgctcaa cagtatgaac atttcgcagc ctaccgtagt gttgtttcc aaaaaggggt 540
 tgcaaaaaat ttgaacgtg caaaaaaat taccaataat ccagaaaatt attatcatgg 600
 attctaaaac ggattaccag ggatttcagt cagatgtacac gttcgtcaca tctcatctac 660
 ctcccggttt taatgaatac gattttgtac cagagtcctt tgatcgtgac aaaacaattg 720
 cactgataat gaattcctct ggatctactg ggttacctaa gggtgtggcc cttccgcata 780

 gaactgcctg cgtcagattc tcgcatgcca gagatcctat ttttggaat caaatcattc 840
 cggatactgc gatTTTaaGT gttgttccat tccatcacgg ttttggaatg ttactacac 900
 tcggatattt gatattgtga tttcgagtcg tcttaatgta tagatttgaa gaagagctgt 960
 ttttacgac ccttcaggat taaaaaatt aaagtgcgtt gctagtacca accctatTTT 1020
 cattcttcgc caaaagcact ctgattgaca aatacattt atctaattta cacgaaattg 1080
 cttctggggg cgcacctctt tcgaaagaag tcggggaagc ggttgcaaaa cgcttccatc 1140
 ttccagggat acgacaagga tatgggctca ctgagactac atcagctatt ctgattacac 1200

ccgaggggga tgataaacgg ggcgcggtcg gtaaagttgt tccatttttt gaagcgaagg 1260

ttgtggatct ggataccggg aaaacgctgg gcgttaatca gagaggcgaa ttatgtgtca 1320

gaggacctat gattatgtcc ggttatgtaa acaatccgga agcgaccaac gccttgattg 1380

acaaggatgg atggctacat tctggagaca tagcttactg ggacgaagac gaacacttct 1440

tcatagttga ccgcttgaag tctttaatta aatacaaagg atatcaggtg gcccccgctg 1500

aattggaatc gatattgtta caacacccca acatcttcga cgcgggcgtg gcaggcttct 1560

ccgacgatga cgccggtgaa ctcccccgcc cggttgttgt tttggagcac ggaaagacga 1620

tgacggaaaa agagatcgtg gattacgtcg ccagtcaagt aacaaccgag aaaaagttgc 1680

gcggaggagt tgtgtttgtg gacgaagtac cgaagggtct taccggaaaa ctgcacgcaa 1740

gaaaaatcag agagatcctc ataaaggcca agaaggcgga aaagtccaaa ttggtttaa 1799

<210> 87

<211> 1800

<212> DNA

<213> Artificial Sequence

<220><223> Artificial sequence

<400> 87

aacatagaaa tcatcatttt aattgattcg gtgttttcga gtctagtatt gattgtttta 60

tcggtatttt tacaacaatt accaacaaca acaacaaca aacaacatta caattactat 120

ttacaattac aaaaaaaaaa aatgtctgaa gacgccaaaa acataaagaa aggccccggc 180

ccattctatc ctctagagga tggaaccgct ggagagcaac tgcataaggc tatgaagaga 240

tacgccttgg ttcttgaac aattgctttt acagatgcac atatcgaggt gaacatcacg 300

tacgcggaat acttcgaaat gtccgttcgg ttggcagaag ctatgaaacg atatgggctg 360

aatacaaatc acagaatcgt cgtaatgcagt gaaaactctc ttcaattctt tatgccggtg 420

ttgggcgcgt tatttatcgg agttgcagtt gcgccgcga acgacattta taatgaacgt 480

gaattgctca acagtatgaa catttcgcag cctaccgtag tgtttgttcc caaaaagggg 540

ttgcaaaaaa ttttgaacgt gcaaaaaaaa ttaccaataa tccagaaaat tattatcatg 600

gatttcaaaa cggattacca gggatttcag tcgatgtaca cgttcgtcac atctcatcta 660

cctcccgggt ttaatgaata cgattttgta ccagagtcct ttgatcgtga caaaacaatt 720

gcactgataa tgaattcctc tggatctact gggttaccta aggggtgtggc ccttccgcat 780

agaactgcct gcgtcagatt ctgcgatgcc agagatccta tttttggcaa tcaaatcatt 840

ccggatactg cgattttaag tgtttgtcca ttccatcacg gttttggaat gtttactaca 900

ctcggatatt tgatatgtgg atttcgagtc gtcttaatgt atagatttga agaagagctg	960
tttttacgat cccttcagga ttacaaaatt caaagtgcgt tgctagtacc aaccctat	1020
tcattcttcg ccaaaagcac tctgattgac aaatacgatt tatctaattt acacgaaatt	1080
gcttctgggg gcgcacctct ttcgaaagaa gtcggggaag cggttgcaaa acgcttccat	1140
cttcaggga tacgacaagg atatgggctc actgagacta catcagctat tctgattaca	1200
cccgaggggg atgataaacc gggcgcggtc ggtaaagttg ttccatTTTT tgaagcgaag	1260
gttgtggatc tggataccgg gaaaacgctg ggcgttaatc agagaggcga attatgtgtc	1320
agaggaccta tgattatgtc cggttatgta aacaatccgg aagcgaccaa cgccttgatt	1380
gacaaggatg gatggctaca ttctggagac atagcttact gggacgaaga cgaacacttc	1440
ttcatagttg accgcttgaa gtctttaatt aaatacaaag gatacaggt ggcccccgct	1500
gaattggaat cgatattgtt acaacacccc aacatcttcg acgcgggcgt ggcaggtctt	1560
cccgacgatg acgccggtga acttcccgc gccgttgttg ttttggagca cggaaagacg	1620
atgacggaaa aagagatcgt ggattacgtc gccagtcaag taacaaccgc gaaaaagttg	1680
cgcggaggag ttgtgtttgt ggacgaagta ccgaaaggtc ttaccggaaa actcgacgca	1740
agaaaaatca gagagatcct cataaaggcc aagaaggcg gaaagtccaa attggtttaa	1800
	1800