

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2024-0167690
(43) 공개일자 2024년11월27일

(51) 국제특허분류(Int. Cl.)
C07K 14/32 (2006.01) C12N 15/75 (2006.01)
C12P 21/02 (2006.01)
(52) CPC특허분류
C07K 14/32 (2013.01)
C12N 15/75 (2013.01)
(21) 출원번호 10-2024-7035890
(22) 출원일자(국제) 2023년03월30일
심사청구일자 없음
(85) 번역문제출일자 2024년10월28일
(86) 국제출원번호 PCT/US2023/065162
(87) 국제공개번호 WO 2023/192953
국제공개일자 2023년10월05일
(30) 우선권주장
63/326,615 2022년04월01일 미국(US)

(71) 출원인
다니스코 유에스 인크.
미합중국 캘리포니아 (우편번호 94304) 팔로 알토
페이지 밀 로드 925
(72) 발명자
본조르니 크리스티나
미국 94304 캘리포니아주 팔로 알토 페이지 밀 로
드 925
피데게부르 프리츠
네덜란드 2342베헤이 우흐스트헤스트 빌럼 에인트
호번스트라트 4
(뒷면에 계속)
(74) 대리인
양영준, 김영

전체 청구항 수 : 총 17 항

(54) 발명의 명칭 **그람 양성 박테리아 세포에서 단백질 생산을 강화하는 프로-영역 돌연변이**

(57) 요약

본 발명은 일반적으로 신규한 프로-영역 DNA 서열을 포함하는 재조합 폴리뉴클레오타이드에 관한 것이다. 본 발명의 특정 양태는 관심 단백질을 암호화하는 DNA 서열에 작동가능하게 연결된 신규한 프로-영역 DNA 서열을 포함하는 하나 이상의 도입된 폴리뉴클레오타이드를 포함하는 재조합 그람 양성 박테리아 균주에 관한 것이다.

대표도

아래형 및 변이체 프로-영역 서열
아래형 B, 란투스 프로-영역 서열 (서열번호 15)
AEEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 A (서열번호 9)
AEEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 B (서열번호 11)
AEEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 C (서열번호 14)
AGKEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 D (서열번호 29)
AGKEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 E (서열번호 30)
AGKEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 F (서열번호 31)
AGKEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 G (서열번호 32)
AEEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM
변이체 프로-영역 서열 H (서열번호 33)
AEEAAKERYLIGFNEQAVSFFVQVEANDGVAILSEEEVEIELLHFFETIPVLSVELSPEDVDALDPAISYIEEDAEVTTM

(52) CPC특허분류

C12P 21/02 (2013.01)

(72) 발명자

프리슈 라이언 엘

미국 19805 델라웨어주 윌밍턴 센터 로드 974

물더 함

네덜란드 2342베헤이 우흐스트헤스트 빌럼 에인트
호번스트라트 4

명세서

청구범위

청구항 1

1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하는 변이체 프로-영역 서열로서, 상기 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링되는, 변이체 프로-영역 서열.

청구항 2

제1항에 있어서, 서열번호 15와 적어도 약 70%, 약 75%, 약 80%, 약 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

청구항 3

위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하는 변이체 프로-영역 서열로서, 상기 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링되는, 변이체 프로-영역 서열.

청구항 4

제3항에 있어서, 서열번호 14와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

청구항 5

위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 73에 이소류신(I)에서 발린(V)으로의 치환(I72V)을 포함하는 변이체 프로-영역 서열(상기 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링됨) 또는 위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 80에 글루탐산(E)에서 이소류신(I)으로의 치환(E80I)을 포함하는 변이체 프로-영역 서열(상기 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링됨).

청구항 6

제5항에 있어서, 서열번호 15와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

청구항 7

제1항 내지 제6항 중 어느 한 항의 변이체 프로-영역 핵산을 포함하는 폴리뉴클레오티드.

청구항 8

관심 단백질(POI)을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 제1항 내지 제6항 중 어느 한 항의 변이체 프로-영역을 암호화하는 업스트림 핵산을 포함하는 폴리뉴클레오티드.

청구항 9

관심 단백질(POI)을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 제1항 내지 제6항 중 어느 한 항의 변이체 프로-영역을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 신호 서열을 암호화하는 업스트림 핵산을 포함하는 폴리뉴클레오티드.

청구항 10

제7항 내지 제9항 중 어느 한 항의 다운스트림 폴리뉴클레오타이드에 작동가능하게 연결된 업스트림 프로모터를 포함하는 발현 카세트.

청구항 11

제10항의 도입된 카세트를 포함하는 그람 양성 숙주 세포.

청구항 12

그람 양성 박테리아 세포에서 관심 단백질(POI)을 생산하는 방법으로서,

(a) POI를 암호화하는 다운스트림 핵산에 작동가능하게 연결된, 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하는 변이체 프로-영역 서열을 암호화하는 다운스트림 핵산에 작동가능하게 연결된 업스트림 프로모터를 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(상기 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링됨); 및

(b) 상기 변형된 세포를 POI의 생산 조건하에 배양하는 단계

를 포함하는 방법.

청구항 13

그람 양성 박테리아 세포에서 관심 단백질(POI)을 생산하는 방법으로서,

(a) POI를 암호화하는 다운스트림 핵산에 작동가능하게 연결된, 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하는 변이체 프로-영역 서열을 암호화하는 다운스트림 핵산에 작동가능하게 연결된 업스트림 프로모터 서열을 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(상기 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링됨); 및

(b) 상기 변형된 세포를 POI의 생산 조건하에 배양하는 단계

를 포함하는 방법.

청구항 14

제12항 또는 제13항에 있어서, 상기 카세트는, 상기 업스트림 프로모터와 다운스트림 변이체 프로-영역 사이에 위치하고 작동가능하게 연결된 신호 서열을 암호화하는 핵산을 추가로 포함하는, 방법.

청구항 15

제12항 또는 제13항에 있어서, 상기 변형된 세포는, 동일한 POI를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 서열번호 15의 프로-영역 서열을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 동일한 업스트림 프로모터 서열을 포함하는 발현 카세트가 도입된 대조군 그람 양성 세포에 비해 증가된 양의 POI를 생산하는, 방법.

청구항 16

제12항 또는 제13항에 있어서, 상기 POI의 증가된 양은 대조군 세포에 비해 적어도 약 0.1% 증가된 것인, 방법.

청구항 17

제12항 또는 제13항에 있어서, 상기 POI는 아세틸 에스테라제, 아미노펩티다제, 아밀라제, 아라비나제, 아라비노푸라노시다제, 탄산무수화효소, 카복시펩티다제, 카탈라제, 셀룰라제, 키티나제, 키모신, 큐티나제, 데옥시리보뉴클레아제, 에피머라제, 에스테라제, α -갈락토시다제, β -갈락토시다제, α -글루카나제, 글루칸 라이사제, 엔도- β -글루카나제, 글루코아밀라제, 글루코스 옥시다제, α -글루코시다제, β -글루코시다제, 글루쿠로니다제, 글리코실 하이드롤라제, 헤미셀룰라제, 헥소스 옥시다제, 하이드롤라제, 인버타제, 이소머라제, 라카제, 리가제, 리파제, 리아제, 만노시다제, 옥시다제, 옥시도리덕타제, 펙테이트 리아제, 펙틴 아세틸 에스테라제,

펙틴 해중합효소, 펙틴 메틸 에스테라제, 펙틴분해 효소, 피하이드롤라제, 폴리올 옥시다제, 피옥시다제, 페놀 옥시다제, 피타제, 폴리갈락투로나제, 프로테아제, 펩티다제, 람노-갈락투로나제, 리보뉴클레아제, 트랜스퍼라제, 수송 단백질, 트랜스글루타미나제, 자일라나제, 헥소스 옥시다제, 및 이들의 조합으로 이루어진 군으로부터 선택되는, 방법.

발명의 설명

기술 분야

[0001] 본 발명은 일반적으로 미생물 숙주 세포, 분자 생물학, 단백질 공학, 발효, 단백질 생산 등의 분야에 관한 것이다. 본 발명의 특정 양태는 신규한 프로-영역 핵산(DNA) 서열, 신규한 프로-영역 DNA 서열을 포함하는 재조합 폴리뉴클레오티드, 관심 단백질을 암호화하는 핵산(DNA) 서열에 작동가능하게 연결된 신규한 프로-영역 DNA 서열을 포함하는 하나 이상의 도입된 폴리뉴클레오티드를 포함하는 유전자 변형(재조합) 그람 양성 박테리아 균주 등에 관한 것이다.

[0002] **관련 출원에 대한 상호 참조**

[0003] 본 출원은 2022년 4월 1일에 출원된 미국 가출원 63/326,615호의 이익을 주장하며, 그 전문은 본원에 참조로 포함된다.

[0004] **서열 목록에 대한 참조**

[0005] "NB41862WOPCT_SequenceListing.xml"이라는 파일명의 서열 목록 텍스트 파일의 전자 제출 내용은 2023년 3월 29일에 생성되었으며 크기는 51 KB이고, 전체가 본원에 참조로 포함된다.

배경 기술

[0006] 그람 양성 미생물은 발효 생성물을 배양 배지로 분비할 수 있어 대규모 산업 발효에 자주 사용된다. 분비된 단백질은 세포막과 세포벽을 통과하여 유출된 후 외부 배지로 방출된다. 예를 들어, 대규모 산업 발효 및 이중 폴리펩티드 분비는 산업에서 널리 사용되는 기술로, 발현될 이중 폴리펩티드를 암호화하는 핵산을 통해 미생물 세포가 형질 전환된다. 단백질 생산 방법의 다양한 발전에도 불구하고, 여러 산업에서 활용되는 관심 단백질의 생산 강화를 위한 단백질 발현에 대한 보다 효율적인 방법을 제공해야 할 필요성이 당업계에 여전히 존재한다.

발명의 내용

[0007] 일반적으로 본원에 기재된 바와 같이, 본 발명은 특히 그람 양성 박테리아(숙주) 세포에서 관심 단백질을 생산하기 위한 조성물 및 방법을 제공한다. 특정 구현예는 신규한 프로-영역 핵산(DNA) 서열, 신규한 프로-영역 DNA 서열을 포함하는 재조합 폴리뉴클레오티드(예를 들어 벡터, 플라스미드, 발현 카세트 등), 다운스트림(3') 유전자 암호화 서열에 작동가능하게 연결된 신규한 프로-영역 서열을 포함하는 재조합 폴리뉴클레오티드, 전단백질 신호(분비) 서열을 암호화하는 업스트림(5') DNA 서열에 작동가능하게 연결된 신규한 프로-영역 서열을 포함하는 재조합 폴리뉴클레오티드 등에 관한 것이다.

[0008] 특정 양태에서, 본 발명은 관심 단백질을 암호화하는 하나 이상의 도입된 폴리뉴클레오티드를 발현하는 재조합 그람 양성 박테리아 균주를 제공한다. 특정 구현예에서, 하나 이상의 도입된 폴리뉴클레오티드는 관심 단백질을 암호화하는 DNA 서열에 작동가능하게 연결된 신규한 프로-영역 DNA 서열을 포함한다(관심 단백질을 암호화하는 DNA 서열은 이에 작동가능하게 연결된 업스트림(5') 단백질 신호 서열을 포함할 수 있음). 다른 특정 양태에서, 본 발명은 관심 단백질을 암호화하는 하나 이상의 도입된 신규한 폴리뉴클레오티드 구성체를 발현하는 재조합 그람 양성 박테리아 균주의 설계/구성을 위한 조성물 및 방법, 관심 단백질을 발현하는 재조합 균주를 배양하기 위한 조성물 및 방법, 관심 단백질의 생산 강화를 위한 조성물 및 방법 등을 제공한다.

[0009] 예를 들어, 하나 이상의 구현예에서 본 발명은 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 32, 및 서열번호 33의 하나 이상의 아미노산 서열 및/또는 이의 조합에 제시된 신규한 변이체 프로-영역 서열을 제공한다. 하나 이상의 다른 구현예에서, 본 발명은 도 1, 도 2, 또는 도 3에 제시된 신규한 변이체 프로-영역 서열 및/또는 이의 조합을 제공한다. 하나 이상의 또 다른 구현예에서, 본 발명은 표 1~5 중 하나에 제시되거나 기재된 신규한 변이체 프로-영역 서열 및/또는 이의 조합을 제공한다.

[0010] 따라서 하나 이상의 특정 구현예는 하나 이상의 아미노산 치환, 하나 이상의 아미노산 삽입 등을 포함하는 변이

체(돌연변이) 프로-영역 서열에 관한 것이다. 특정 구현예에서, 변이체 프로-영역 서열은 위치 30에 아미노산 치환을 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링된다. 다른 구현예에서, 변이체 프로-영역 서열은 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링된다. 하나 이상의 다른 구현예에서, 변이체 프로-영역 서열은 서열번호 15의 위치 1~84에 대해 적어도 약 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 서열 동일성을 포함하는 모체 또는 기준 폴리펩티드로부터 유래된다. 다른 특정 구현예에서, 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입 및 위치 3에 라이신(K)의 아미노산 삽입을 적어도 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링된다. 하나 이상의 다른 구현예에서, 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 적어도 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링된다. 일부 다른 구현예에서, 변이체 프로-영역은 서열번호 14의 위치 1~86에 대해 적어도 약 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 포함하는 모체 또는 기준 폴리펩티드로부터 유래된다. 또 다른 구현예에서, 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 위치 4에 알라닌(A)의 아미노산 삽입을 적어도 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 30에 따라 넘버링된다. 특정 양태에서, 변이체 프로-영역은 서열번호 30의 위치 1~87과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된다. 하나 이상의 다른 특정 구현예에서, 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 위치 4에 세린(S)의 아미노산 삽입을 적어도 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 29에 따라 넘버링된다. 일부 다른 구현예에서, 변이체 프로-영역은 서열번호 29의 위치 1~87과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된다. 하나 이상의 다른 구현예에서, 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 위치 4에 알라닌(A)의 아미노산 삽입, 및 위치 5에 알라닌(A)의 아미노산 삽입을 적어도 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 31에 따라 넘버링된다. 하나 이상의 특정 구현예 또는 양태에서, 변이체 프로-영역은 서열번호 31의 위치 1~88과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된다. 다른 구현예에서, 변이체 프로-영역 서열은 위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 73에 이소류신(I)에서 발린(V)으로의 치환(I72V)을 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 32에 따라 넘버링된다. 하나 이상의 다른 양태 또는 구현예에서, 변이체 프로-영역 서열은 위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 80에 글루탐산(E)에서 이소류신(I)으로의 치환(E80I)을 포함하며, 변이체 프로-영역의 아미노산 위치는 서열번호 33에 따라 넘버링된다. 특정 구현예에서, 변이체 프로-영역은 서열번호 15, 서열번호 32, 및/또는 서열번호 33의 위치 1~84와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된다. 다른 양태에서, 본 발명의 변이체 프로-영역은 표 1~5, 도 1, 도 2, 도 3, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 32, 서열번호 33, 및 이들의 조합 중 어느 하나에 제시된 아미노산 변형을 포함한다. 하나 이상의 구현예에서, 표 1~5, 도 1, 도 2, 도 3, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 32, 서열번호 33, 및 이들의 조합 중 어느 하나에 제시된 아미노산 변형을 포함하는 변이체 프로-영역은 서열번호 15와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된다.

[0011] 따라서, 하나 이상의 다른 특정 구현예는 본 발명의 신규한 변이체 프로-영역 서열을 암호화하는 재조합 핵산을 제공한다. 따라서 특정 구현예는 변이체 프로-영역 핵산 등을 포함하는 폴리뉴클레오티드에 관한 것이다. 하나 이상의 구현예에서, 본 발명은 이중성 관심 단백질(POI)을 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된 본 발명의 변이체 프로-영역을 암호화하는 업스트림(5') 핵산(서열)을 포함하는 폴리뉴클레오티드를 제공한다. 다른 특정 양태에서, 본 발명은 관심 단백질(POI)을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 본 발명의 변이체 프로-영역을 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된 전단백

질 신호(분비) 서열을 암호화하는 업스트림(5') 핵산을 포함하는 폴리뉴클레오티드를 제공한다. 다른 특정 구현예에는 본 발명의 다운스트림(3') 폴리뉴클레오티드에 작동가능하게 연결된 업스트림(5') 프로모터 영역 서열을 포함하는 발현 카세트에 관한 것이다.

[0012] 하나 이상의 다른 특정 구현예에서, 본 발명은, 본 발명의 하나 이상의 도입된 폴리뉴클레오티드 또는 발현 카세트를 포함하는 재조합(변형) 그람 양성 박테리아 세포/균주에 관한 것이다. 따라서, 본 발명의 다른 구현예는 본원에 제시되고 기재된 재조합 그람 양성 세포에서 이중성 관심 단백질을 생산하는 방법/프로세스에 관한 것이다.

[0013] 하나 이상의 일부 구현예에서, 본 발명은 그람 양성 박테리아 세포에서 이중성 POI를 생산하는 방법으로서, 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하는 변이체 프로-영역 서열을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 업스트림 프로모터 영역을 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(변이체 프로-영역 핵산(서열)은 POI를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결되고, 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링됨), 및 변형된 세포를 POI의 생산 조건하에 성장/배양/발효시키는 단계를 포함하는 방법을 제공한다. 다른 특정 구현예에서, 본 발명은 그람 양성 박테리아 세포에서 이중성 POI를 생산하는 방법으로서, 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하는 변이체 프로-영역을 암호화하는 다운스트림 핵산(서열)에 작동가능하게 연결된 업스트림 프로모터 영역 서열을 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(변이체 프로-영역 핵산(서열)은 POI를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결되고, 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링됨), 및 변형된 세포를 POI의 생산 조건하에 성장/배양/발효시키는 단계를 포함하는 방법을 제공한다. 또 다른 구현예에서, 본 발명은 그람 양성 박테리아 세포에서 이중성 POI를 생산하는 방법으로서, 본 발명의 발현 카세트를 그람 양성 세포에 도입하는 단계 및 변형된 세포를 POI의 생산 조건하에 성장/배양/발효시키는 단계를 포함하는 방법에 관한 것이다. 이러한 방법의 하나 이상의 특정 구현예에서, 카세트는, 프로모터와 변이체 프로-영역 서열 사이에 위치하고 작동가능하게 연결된 전단백질 신호(분비) 서열을 암호화하는 핵산(서열)을 포함한다. 이러한 방법의 다른 일부 구현예에서, 변형된 세포는 대조군 그람 양성 세포에 비해 증가된 양의 POI를 생산한다. 예를 들어, 대조군 그람 양성 세포는, 동일한 POI(즉, 변형된 세포와 동일한 POI)를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 서열번호 15를 포함하는 프로-영역 서열을 암호화하는 다운스트림 핵산에 작동가능하게 연결된 동일한 업스트림 프로모터 서열(즉, 변형된 세포와 동일)을 포함하는 도입된 발현 카세트를 포함한다. 이러한 방법의 특정 구현예에서, 도입된 카세트는 세포의 게놈에 통합된다. 이러한 방법의 특정 구현예에서, 적어도 두 개의 카세트가 세포에 도입된다. 본 발명의 하나 이상의 다른 특정 구현예가 본원에 기재되고 하기에 제시된다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 야생형(기준) *B. 켈투스* 프로-영역 및 특정 SEL 변이체 프로-영역 서열의 아미노산 서열을 나타낸 것이다. 도 1에 도시된 바와 같이, 기준 *B. 켈투스* 프로-영역 서열은 서열번호 15에 제시된 84개 아미노산(잔기) 위치를 포함하고(위치 30의 글루탐산(E) 잔기는 밑줄 표시), 변이체 프로-영역 서열 A는 서열번호 9에 제시된 84개 아미노산 위치를 포함하고(위치 30의 히스티딘(H) 잔기는 볼드체 표시), 변이체 프로-영역 서열 B는 서열번호 11에 제시된 84개 아미노산 위치를 포함하고(위치 30의 글리신(G) 잔기는 볼드체 표시), 변이체 프로-영역 서열 C는 서열번호 14에 제시된 86개 아미노산 위치를 포함하고(위치 2와 3에 각각 삽입된 글리신(G)과 라이신(K) 잔기는 볼드체 표시(GK), 및 위치 32의 E 잔기는 밑줄 표시), 변이체 프로-영역 서열 D는 서열번호 29에 제시된 87개 아미노산 위치를 포함하고(위치 2~4에 각각 삽입된 글리신(G), 라이신(K), 및 세린(S) 잔기는 볼드체 표시(GKS)), 변이체 프로-영역 서열 E는 서열번호 30에 제시된 87개 아미노산 위치를 포함하고(위치 2~4에 각각 삽입된 글리신(G), 라이신(K), 및 알라닌(A) 잔기는 볼드체 표시(GKAA)), 변이체 프로-영역 서열 F는 서열번호 31에 제시된 88개 아미노산 위치를 포함하고(위치 2~5에 각각 삽입된 글리신(G), 라이신(K), 알라닌(A), 및 알라닌(A) 잔기는 볼드체 표시(GKAA)), 변이체 프로-영역 서열 G는 서열번호 32에 제시된 84개 아미노산 위치를 포함하고(위치 30의 글리신(G) 잔기는 볼드체 표시, L68K 및 I72V 치환은 밑줄 표시), 변이체 프로-영역 서열 H는 서열번호 33에 제시된 84개 아미노산 위치를 포함한다(위치 30의 글리신(G) 잔기는 볼드체 표시, L68K 및 E80I 치환은 밑줄 표시).

도 2는 야생형(WT) 프로-영역 서열(SEQ 15)과 정렬된 변이체 프로-영역 서열 A(SEQ 09), 변이체 프로-영역 서열

B(SEQ 11), 변이체 프로-영역 서열 C(SEQ 14), 변이체 프로-영역 서열 D(SEQ 29), 변이체 프로-영역 서열 E(SEQ 30), 변이체 프로-영역 서열 F(SEQ 31), 변이체 프로-영역 서열 G(SEQ 32), 및 변이체 프로-영역 서열 H(SEQ 33)의 아미노산 서열을 나타낸 것이다. 도 2에 제시된 바와 같이, WT 프로-영역은 위치 30에 글루탐산(E) 잔기를 포함하고(E; SEQ 15), 변이체 프로-영역 서열 A는 위치 30에 히스티딘(H) 잔기를 포함하고(H; SEQ 09), 변이체 프로-영역 서열 B는 위치 30에 글리신(G) 잔기를 포함하고(G; SEQ 11), 변이체 프로-영역 서열 C는 글리신(G)과 라이신(K)의 2개 아미노산 삽입을 포함하고(GK 삽입; SEQ 14), 변이체 프로-영역 서열 D는 글리신(G), 라이신(K), 및 세린(S)의 3개 아미노산 삽입을 포함하고(GKS 삽입; SEQ 29), 변이체 프로-영역 서열 E는 글리신(G), 라이신(K), 및 알라닌(A)의 3개 아미노산 삽입을 포함하고(GKA 삽입; SEQ 30), 변이체 프로-영역 서열 F는 글리신(G), 라이신(K), 알라닌(A), 알라닌(A)의 4개 아미노산 삽입을 포함하고(GKAA 삽입; SEQ 31), 변이체 프로-영역 서열 G는 위치 30에 글리신(G) 잔기, 위치 68에 라이신(K) 잔기, 및 위치 72에 발린(V) 잔기를 포함하고, 변이체 프로-영역 서열 H는 위치 30에 글리신(G) 잔기, 위치 68에 라이신(K) 잔기, 및 위치 80에 이소류신(I) 잔기를 포함한다. 예를 들어, WT 프로-영역 서열(SEQ 15)과 변이체 프로-영역 서열 C(SEQ 14)의 정렬에 도시된 바와 같이, 2개 아미노산 잔기 GK가 WT 프로-영역(서열번호 15)의 위치 1에 있는 알라닌(A)과 위치 2에 있는 글루탐산(E) 사이에 삽입되어, 서열번호 14의 86개 아미노산 변이체 프로-영역 서열 C가 생성되었다. 도 2에 제시된 바와 같이, SEQ 15, SEQ 09, SEQ 11, SEQ 32, 및 SEQ 33에 도시된 2개의 하이픈(--)은 SEQ 14에 도시된 변이체 프로-영역 서열 C 대비 2개의 아미노산 위치 간격을 나타낸다.

도 3은 기준(야생형) *B. 켈투스* 프로-영역 아미노산 서열(E₃₀, 서열번호 15), 변이체 프로-영역 서열 A(H₃₀; 서열번호 09), 변이체 프로-영역 서열 B(G₃₀; 서열번호 11), 변이체 프로-영역 서열 C(G₂K₃ 삽입, 서열번호 14), 변이체 프로-영역 서열 D(G₂K₃S₄ 삽입, 서열번호 29), 변이체 프로-영역 서열 E(G₂K₃A₄ 삽입, 서열번호 30), 변이체 프로-영역 서열 F(G₂K₃A₄A₅ 삽입, 서열번호 31), 변이체 프로-영역 서열 G(E30G, L68K, 및 I72V 치환 포함; 서열번호 32), 및 변이체 프로-영역 서열 H(E30G, L68K, 및 E80I 치환 포함; 서열번호 33)의 아미노산(잔기) 위치 넘버링을 나타낸 것으로, 아미노산 위치는 5'(N-말단)에서 3'(C-말단) 방향으로 아래 첨자 숫자로 표시된다(예를 들어, 서열번호 15의 위치 1에 알라닌(A₁), 서열번호 15의 위치 2에 글루탐산(E₂) 등). 예를 들어, 도 3에 제시된 바와 같이, 본 발명의 돌연변이체 프로-영역 서열은 야생형(기준) 프로-영역 서열 넘버링(서열번호 15, E30), 변이체(기준) 프로-영역 서열 A 넘버링(서열번호 9, H30), 변이체(기준) 프로-영역 서열 B 넘버링(서열번호 11, G30), 변이체(기준) 프로-영역 서열 C 넘버링(서열번호 14, GK 삽입), 변이체(기준) 프로-영역 서열 D 넘버링(서열번호 29, GKS 삽입), 변이체(기준) 프로-영역 서열 E 넘버링(서열번호 20, GKA 삽입), 변이체(기준) 프로-영역 서열 F 넘버링(서열번호 31, GKAA 삽입), 변이체(기준) 프로-영역 서열 G 넘버링(서열번호 32, L68K/I72V 치환), 변이체(기준) 프로-영역 서열 H 넘버링(서열번호 33, L68K/E80I 치환), 또는 이들의 조합을 나타낼 수 있다.

생물학적 서열의 간단한 설명

서열번호 1은 업스트림(5') *aprE* 유전자 플랭킹 영역, *B. 서브틸리스* *rrnI*-P2 프로모터의 변이체, 및 5'-*aprE* UTR 영역을 포함하는 핵산(DNA) 서열이다.

서열번호 2는 "BG46"으로 명명된 야생형 *바실러스 지브소니* 서브틸리신의 아미노산 서열이다.

서열번호 3은 "BG46_변이체 1"로 명명된 변이체 *B. 지브소니* BG46 서브틸리신의 아미노산 서열이다.

서열번호 4는 야생형 *B. 서브틸리스* *AprE* 단백질 신호 서열을 암호화하는 DNA 서열이다.

서열번호 5는 변이체 프로-영역 서열 A(30H, 서열번호 9)를 암호화하는 DNA 서열이다.

서열번호 6은 야생형 *B. 아밀로리퀘팍시엔스* BPN' 종결자의 DNA 서열이다.

서열번호 7은 카나마이신(*kan*) 유전자 발현 카세트의 DNA 서열이다.

서열번호 8은 "BG46_변이체 2"로 명명된 변이체 *B. 지브소니* BG46 서브틸리신의 아미노산 서열이다.

서열번호 9는 변이체 프로-영역 A DNA 서열(서열번호 5)에 의해 암호화된 아미노산 서열이다.

서열번호 10은 변이체 프로-영역 서열 B(30G, 서열번호 11)를 암호화하는 DNA 서열이다.

서열번호 11은 변이체 프로-영역 B DNA 서열(서열번호 10)에 의해 암호화된 아미노산 서열이다.

서열번호 12는 *B. 아밀로리케파시엔스* BPN' 프로-영역 서열의 아미노산 서열이다.

서열번호 13은 변이체 프로-영역 서열 C(GK 삽입, 서열번호 14)를 암호화하는 DNA 서열이다.

서열번호 14는 변이체 프로-영역 C DNA 서열(서열번호 13)에 의해 암호화된 아미노산 서열이다.

서열번호 15는 야생형(WT) *B. 켈투스* 프로-영역 서열의 아미노산 서열이다.

서열번호 16은 *B. 리체니포르미스* *serA* 유전자 5' FR의 DNA 서열이다.

서열번호 17은 *rrnI*-p3 프로모터 영역을 포함하는 DNA 서열이다.

서열번호 18은 *B. 서브틸리스* *aprE* 5'-UTR의 DNA 서열이다.

서열번호 19는 *B. 리체니포르미스* *amyL* 종결자의 DNA 서열이다.

서열번호 20은 *B. 리체니포르미스* *serA* 유전자 3' FR의 DNA 서열이다.

서열번호 21은 *B. 리체니포르미스* *lysA* 유전자 5' FR의 DNA 서열이다.

서열번호 22는 *B. 리체니포르미스* *lysA* 유전자 3' FR의 DNA 서열이다.

서열번호 23은 변이체 프로-영역 서열 C(GKA 삽입, 서열번호 14)를 암호화하는 DNA 서열이다.

서열번호 24는 변이체 프로-영역 C DNA 서열(서열번호 23)에 의해 암호화되는 아미노산 서열이다.

서열번호 25는 변이체 프로-영역 서열 C(GKAA 삽입, 서열번호 14)를 암호화하는 DNA 서열이다.

서열번호 26은 변이체 프로-영역 C DNA 서열(서열번호 25)에 의해 암호화되는 아미노산 서열이다.

서열번호 27은 변이체 프로-영역 서열 C(GKAS 삽입, 서열번호 14)를 암호화하는 DNA 서열이다.

서열번호 28은 변이체 프로-영역 C DNA 서열(서열번호 27)에 의해 암호화되는 아미노산 서열이다.

서열번호 29는 변이체 프로-영역 서열 D의 아미노산 서열이다.

서열번호 30은 변이체 프로-영역 서열 E의 아미노산 서열이다.

서열번호 31은 변이체 프로-영역 서열 F의 아미노산 서열이다.

서열번호 32는 변이체 프로-영역 서열 G의 아미노산 서열이다.

서열번호 33은 변이체 프로-영역 서열 H의 아미노산 서열이다.

발명을 실시하기 위한 구체적인 내용

[0015] 간략하게 전술하고 이하에 상세히 기재된 바와 같이, 본 발명은 특히 신규한 프로-영역 핵산(DNA) 서열, 신규한 프로-영역 DNA 서열을 포함하는 재조합 폴리뉴클레오티드(예를 들어 벡터, 플라스미드, 발현 카세트 등), 다운 스트림(3') 유전자 암호화 서열에 작동가능하게 연결된 신규한 프로-영역 서열을 포함하는 재조합 폴리뉴클레오티드, 단백질 신호(분비) 서열을 암호화하는 업스트림(5') DNA 서열에 작동가능하게 연결된 신규한 프로-영역 서열을 포함하는 재조합 폴리뉴클레오티드 등을 제공한다. 특정 양태에서, 본 발명은 관심 단백질을 암호화하는 하나 이상의 도입된 폴리뉴클레오티드를 발현하는 재조합 그람 양성 박테리아 균주를 제공한다. 특정 구현예에서, 하나 이상의 도입된 폴리뉴클레오티드는 관심 단백질을 암호화하는 DNA 서열에 작동가능하게 연결된 신규한 프로-영역 DNA 서열을 포함한다(관심 단백질을 암호화하는 DNA 서열은 이에 작동가능하게 연결된 업스트림(5') 단백질 신호 서열을 포함할 수 있음). 따라서, 본 발명의 다른 특정 양태는 특히 관심 단백질을 암호화하는 하나 이상의 도입된 신규한 폴리뉴클레오티드 구성체를 발현하는 재조합 그람 양성 박테리아 균주의 설계/구성을 위한 조성물 및 방법, 관심 단백질을 발현하는 재조합 균주를 배양하기 위한 조성물 및 방법, 관심 단백질의 생산 강화를 위한 조성물 및 방법 등을 제공한다.

[0016] I. 정의

[0017] 본 발명의 재조합(변형) 균주 및 본원에 기재된 이의 방법과 관련하여, 다음의 용어 및 문구가 정의된다. 본원에 정의되지 않은 용어는 당업계에서 사용되는 통상적인 의미를 따른다.

[0018] 달리 정의되지 않는 한, 본원에서 사용되는 모든 기술 용어 및 과학 용어는 본 발명의 조성물 및 방법이 적용되

는 기술 분야의 당업자가 일반적으로 이해하는 것과 동일한 의미를 갖는다. 본원에 기재된 것과 유사하거나 동등한 임의의 방법 및 재료가 본 발명의 조성물 및 방법의 실시 또는 시험에 사용될 수도 있지만, 대표적인 예시적 방법 및 재료가 이제 기술된다. 본원에 인용된 모든 간행물 및 특허는 그 전체가 참조로 포함된다.

[0019] 또한, 임의의 선택적 요소를 제외하도록 청구범위가 작성될 수 있음에 유의한다. 따라서, 이러한 서술은 청구 요소의 설명과 관련하여 "전적으로", "단지", "~를 제외한", "~를 포함하지 않는" 등과 같은 배타적 용어를 사용하거나, "부정적" 제한이나 그 단서를 사용하기 위한 선행 근거로서의 역할을 하기 위한 것이다.

[0020] 본원에 기재되고 예시된 개별 구현에 각각은 본원에 기재된 본 발명의 조성물 및 방법의 범위 또는 사상을 벗어나지 않으면서 임의의 다른 여러 구현예의 특징과 쉽게 분리되거나 조합될 수 있는 별개의 구성요소 및 특징을 가지며, 이는 본 발명을 읽을 때 당업자에게 명백할 것이다. 언급된 임의의 방법은 열거된 사건의 순서로 또는 논리적으로 가능한 다른 순서로 수행될 수 있다.

[0021] 본원에서 사용되는 바와 같이, "그램 양성 박테리아", "그램 양성 세포", "그램 양성 박테리아 균주", 및/또는 "그램 양성 박테리아 세포"라는 문구는 당업계에서 사용되는 것과 동일한 의미를 갖는다. 예를 들어, 그람 양성 박테리아 세포는 악티노박테리아(Actinobacteria) 및 피르미쿠테스(Firmicutes)의 모든 균주를 포함한다. 특정 구현예에서, 이러한 그람 양성 박테리아는 간균강(class Bacilli), 클로스트리디움강(class Clostridia) 및 몰리쿠테스강(class Mollicutes)의 것이다.

[0022] 본원에서 사용되는 바와 같이, "바실러스" 속은 *B. 서브틸리스(subtilis)*, *B. 리체니포르미스(licheniformis)*, *B. 렌투스(lentus)*, *B. 브레비스(brevis)*, *B. 스테아로써모필루스(stearothermophilus)*, *B. 알칼로필루스(alkalophilus)*, *B. 아밀로리퀘파시엔스(amyloliquefaciens)*, *B. 클라우시(clausii)*, *B. 할로두란스(halodurans)*, *B. 메가테리움(megaterium)*, *B. 코아굴란스(coagulans)*, *B. 서큘란스(circulans)*, *B. 라우투스(lautus)*, 및 *B. 투링기엔시스(thuringiensis)*를 비롯해(이에 한정되지 않음), 당업계에 알려진 "바실러스" 속 내의 모든 종을 포함한다. 바실러스 속은 계속해서 분류학적 재편성을 거치는 것으로 알려져 있다. 따라서, 상기 속은 현재 "*지오바실러스 스테아로써모필루스(Geobacillus stearothermophilus)*"로 명명된 *B. 스테아로써모필루스*와 같은 유기체를 비롯해(이에 한정되지 않음) 재분류된 종을 포함한다.

[0023] 본원에서 사용되는 바와 같이, 용어 "재조합" 또는 "비천연"은 적어도 하나의 조작된 유전적 변형을 갖거나 또는 이중성 핵산 분자의 도입에 의해 변형된 유기체, 미생물, 세포, 핵산 분자, 또는 벡터를 의미하거나, 이중성 또는 내인성 핵산 분자 또는 유전자의 발현이 제어될 수 있도록 변경된 세포(예를 들어, 미생물 세포)를 의미한다. 재조합은 또한 비천연 세포로부터 유래되거나 또는 하나 이상의 이러한 변형을 갖는 비천연 세포의 자손인 세포를 의미한다. 유전적 변형은, 예를 들어 단백질을 암호화하는 발현가능한 핵산 분자를 도입하는 변형, 또는 다른 핵산 분자 추가, 결실, 치환, 또는 세포의 유전적 물질의 다른 기능적 변형을 포함한다. 예를 들어, 재조합 세포는 천연(야생형) 세포 내에서 동일하거나 상동적인 형태로 발견되지 않는 유전자 또는 다른 핵산 분자를 발현할 수 있거나(예를 들어, 융합 또는 키메라 단백질), 또는 과발현되거나, 저발현되거나, 최소로 발현되거나, 또는 전혀 발현되지 않는 것과 같은 내인성 유전자의 변경된 발현 패턴을 제공할 수 있다. "재조합" 또는 "재조합된" 핵산을 생성하는 것은 일반적으로 2개 이상의 핵산 단편을 조립하는 것이며, 조립을 통해 키메라 유전자가 생성된다.

[0024] 용어 "유래된"은 용어 "기인된", "수득된", "수득 가능한" 및 "생성된"을 포함하며, 일반적으로 하나의 명시된 물질 또는 조성물의 기원이 또 다른 명시된 물질 또는 조성물에서 발견되거나, 또 다른 명시된 물질 또는 조성물을 참조하여 기술될 수 있는 특징을 가진다는 것을 나타낸다. 예를 들어, 본 발명의 재조합 그람 양성 박테리아 세포는 임의의 알려진 그람 양성 박테리아 균주로부터 유래/수득될 수 있다.

[0025] 본원에서 사용되는 바와 같이, "핵산"은 뉴클레오티드 서열 또는 폴리뉴클레오티드 서열, 및 이의 단편 또는 일부를 지칭할 뿐만 아니라, 센스 가닥 또는 안티센스 가닥을 나타내는지와 상관없이, 이중가닥 또는 단일가닥일 수 있는 게놈 또는 합성 기원의 DNA, cDNA, 및 RNA를 지칭한다. 유전자 암호의 축퇴성으로 인해, 다수의 뉴클레오티드 서열이 주어진 단백질을 암호화할 수 있다는 것을 이해할 수 있을 것이다.

[0026] 본원에 기술된 폴리뉴클레오티드(또는 핵산 분자)는 "유전자", "벡터", 및 "플라스미드"를 포함하는 것이 이해된다.

[0027] 따라서, 용어 "유전자"는 단백질 암호화 서열의 전부 또는 일부를 포함하는 아미노산의 특정 서열을 암호화하는 폴리뉴클레오티드를 나타내며, 예를 들어 유전자가 발현되는 조건을 결정하는 프로모터 서열과 같은 조절(비전사) DNA 서열을 포함할 수 있다. 유전자의 전사 영역은 5'-비번역 영역(UTR), 및 3'-UTR을 포함하는 비번역 영

역(UTR)뿐만 아니라 암호화 서열도 포함할 수 있다.

- [0028] 본원에서 사용되는 바와 같이, "내인성 유전자"는 유기체의 게놈 내의 자연적 위치에 있는 유전자를 지칭한다.
- [0029] 본원에서 사용되는 바와 같이, "이종성" 유전자, "비내인성" 유전자, 또는 "외래" 유전자는 숙주 유기체에서 정상적으로는 발견되지 않고 유전자 전달에 의해 숙주 유기체 내로 도입되는 유전자를 지칭한다. 용어 "외래" 유전자(들)는 비천연 유기체 내로 삽입된 천연 유전자 및/또는 천연 또는 비천연 유기체 내로 삽입된 키메라 유전자를 포함한다.
- [0030] 본원에서 사용되는 바와 같이, "이종성 제어 서열"은 자연에서는 관심 유전자의 발현을 조절하는(제어하는) 기능을 하지 않는 유전자 발현 제어 서열(예를 들어, 프로모터, 인핸서, 종결자 등)을 나타낸다. 일반적으로, 이종성 핵산은 이들이 존재하는 세포 또는 게놈의 일부에 대해 내인성이 아니며(천연이 아니며), 감염, 전달감염, 형질도입, 형질전환, 미세주입, 전기천공 등에 의해 세포에 부가되었다. "이종성" 핵산 구성체는 천연 숙주 세포에서 발견되는 제어 서열/DNA 암호화 서열의 조합과 동일하거나 이와는 상이한 제어 서열/DNA 암호화(ORF) 서열의 조합을 함유할 수 있다.
- [0031] 본원에서 사용되는 용어 "발현"은 본 발명의 핵산 분자로부터 유래된 센스 RNA(mRNA) 또는 안티센스 RNA의 전사 및 안정적 축적을 나타낸다. 발현은 또한 mRNA의 폴리펩티드로의 번역을 나타낼 수 있다. 따라서, 용어 "발현"은 전사, 전사 후 변형, 번역, 번역 후 변형, 분비 등을 포함하는(이에 한정되지 않음), 폴리펩티드의 생산에 관여된 임의의 단계를 포함한다.
- [0032] 본원에서 사용되는 바와 같이, 용어 "암호화 서열"은 (암호화된) 단백질 산물의 아미노산 서열을 직접 특정하는 뉴클레오티드 서열을 나타낸다. 암호화 서열의 경계는 일반적으로 오픈 리딩 프레임(이하, "ORF")에 의해 결정되며, 이는 대개 ATG 시작 코돈으로 시작된다. 암호화 서열은 일반적으로 DNA, cDNA, 및 재조합 뉴클레오티드 서열을 포함한다.
- [0033] 본원에서 사용되는 바와 같이, 용어 "프로모터", "프로모터 요소", "프로모터 서열" 등은 프로모터 영역 서열이 업스트림(5')에 위치하고 다운스트림(3') 유전자 CDS에 작동가능하게 연결되는 경우 유전자 암호화 서열(CDS)의 메신저 RNA(mRNA)로의 전사를 제어할 수 있는 핵산(DNA) 서열을 지칭한다. 당업자가 일반적으로 이해하는 바와 같이, 프로모터는 일반적으로 RNA 폴리머라제에 의한 특정 결합 및 전사 개시를 위한 부위를 제공한다. 특정 양태에서, 용어 "프로모터"는 전사 개시에 필요한 프로모터 핵산 서열의 최소 부분(즉, RNA 폴리머라제 결합 부위 포함)을 나타낸다. 예를 들어, 프로모터는 일반적으로 -10(공통 서열) 요소 및 -35(공통 서열) 요소를 포함하며, 이들은 업스트림(5')에 있고 전사될 유전자 CDS의 +1 전사 개시 부위(TSS)에 대한 것이다. 코어 프로모터 -10 및 -35 요소는 각각 당업계에서 일반적으로 "TATAAT"(프립나우(Pribnow) 박스) 공통 영역 및 "TTGACA" 공통 영역으로 지칭된다. 코어 프로모터(10 및 -35) 영역의 간격은 일반적으로 약 15 내지 20개의 개재 염기쌍(뉴클레오티드)에 의해 분리된다(또는 이격된다).
- [0034] 프로모터는 그 전체가 천연 유전자로부터 유래되거나, 자연에서 발견되는 상이한 프로모터로부터 유래된 상이한 요소로 구성되거나, 심지어 합성 핵산 절편을 포함할 수 있다. 상이한 프로모터가 상이한 세포 유형에서, 또는 상이한 발달 단계에서, 또는 상이한 환경 조건 또는 생리적 조건에 응답하여 유전자의 발현을 지시할 수 있음이 당업자에 의해 이해된다. 프로모터는 구성적 프로모터, 유도적 프로모터, 조정가능한 프로모터, 하이브리드 프로모터, 합성 프로모터, 탠덤 프로모터 등일 수 있다. 대부분의 세포 유형에서 대부분의 시점에 유전자가 발현 되도록 하는 프로모터를 일반적으로 "항시성 프로모터"라고 한다. 또한, 대부분의 경우 조절 서열의 정확한 경계가 완전히 정의되지 않았기 때문에, 길이가 다른 DNA 단편이 동일한 프로모터 활성을 가질 수 있는 것으로 인정된다. 특정 양태에서, 성숙 관심 단백질을 암호화하는 다운스트림(3') DNA 서열(ORF)에 작동가능하게 연결된 프로-영역 서열(PRO)을 암호화하는 다운스트림 DNA 서열에 작동가능하게 연결된 단백질 신호 서열(SS)을 암호화하는 다운스트림 DNA 서열에 작동가능하게 연결된 업스트림(5') 프로모터 서열(pro)은 5'-[pro]-[SS]-[PRO]-[ORF]-3'으로 개략적으로 표시될 수 있다.
- [0035] 본원에서 사용되는 바와 같이, "관심 유전자의 단백질 암호화 서열에 연결된 관심 유전자의 발현을 제어하는 기능적 프로모터 서열"은, 원하는 그람 양성 숙주 세포에서 암호화 서열의 전사 및 번역을 제어하는 프로모터 서열을 지칭한다. 예를 들어, 특정 구현예에서, 본 발명은 그람 양성 세포에서 기능적인 업스트림(5') 프로모터 (또는 5' 프로모터 영역 또는 탠덤 5' 프로모터 등)를 포함하는 폴리뉴클레오티드를 제공한다(기능적 프로모터 영역은 관심 단백질을 암호화하는 핵산 서열에 작동가능하게 연결됨).
- [0036] 본원에서 사용되는 바와 같이, 용어 "전구체 단백질"은 단백질의 비활성 형태를 지칭한다. 특정 양태에서, 전장

단백질은 프로-서열 및 성숙 단백질(약칭 "전단백질")의 형태로 전구체로서 합성된다. 다른 양태에서, 전장 단백질은 신호 펩티드 서열, 프로-서열, 및 성숙 단백질(약칭 "전-전단백질")의 형태로 전구체로서 합성된다. 예를 들어, 전-서열은 일반적으로 수송을 위한 신호 펩티드 역할을 하며, 프로-서열은 일반적으로 연관된(성숙) 단백질의 올바른 폴딩에 필수적이다.

[0037] 본원에서 사용되는 바와 같이, 용어 "성숙 단백질"은 비활성 전구체(전장) 단백질과 대조적으로 단백질의 활성 형태를 지칭한다.

[0038] 본원에서 사용되는 바와 같이, 용어 "신호 서열", "분비 신호", 및 "신호 펩티드"는 상호교환적으로 사용될 수 있으며 전구체 단백질의 분비 또는 직접 수송에 참여할 수 있는 아미노산 잔기의 서열을 지칭한다. 신호(pre) 서열은 일반적으로 전좌 중에 신호 펩티다제에 의해 전구체 단백질로부터 절단된다. 신호(pre) 서열은 일반적으로 성숙 단백질 서열의 N-말단에 위치하거나, 신호(pre) 서열과 프로-영역(pro) 서열이 작동가능한 조합으로 사용되고 성숙 POI 서열의 업스트림(5')에 있는 경우 프로-영역(pro) 서열의 N-말단에 위치한다.

[0039] 본원에서 사용되는 바와 같이, 용어 "프로 서열", "프로-서열", 및 "프로-영역 서열"은 상호교환적으로 사용될 수 있으며 "PRO" 서열, "Pro" 서열, "pro" 서열 등으로 약칭될 수 있다. 본원에서 사용되는 바와 같이, 용어 프로-서열은 당업계에서 이해되는 것과 동일한 의미를 갖는다. 예를 들어, *B. 서브틸리스* 알칼리성 세린 프로테아제 "서브틸리신"은 단백질 분비를 위한 신호(pre) 서열, 후속적으로 77개 아미노산 프로-영역(pro) 서열, 후속적으로 성숙한 서브틸리신(예를 들어 pre-pro-서브틸리신)을 암호화하는 아미노산 서열로 이루어진 pre-pro-서브틸리신으로서 먼저 생성된다. 프로-서열은 종종 연관된(성숙) 단백질의 올바른 폴딩에 필수적이며, 분자 내 사폐론(예를 들어 단백질 폴딩 반응을 직접 촉매)으로서 작용한다. 마찬가지로, 프로-서열은 성숙 관심 단백질의 폴딩과 세포 내 수송(또는 분비) 모두에 필요할 수 있으며, 이는 이러한 두 기능이 밀접하게 관련되어 있음을 시사한다. 특정 양태에서, 본 발명의 프로-영역 서열은 서열번호 15의 야생형(WT, 기준) *B. 켄투스* 프로-영역 서열로부터 유래된 아미노산 서열을 포함한다. 특정 구현예에서, 프로-영역 서열은 서열번호 15, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 31, 서열번호 32, 및/또는 서열번호 33에 대한 상동성을 포함하는 아미노산 서열로부터 유래된다. 따라서, 특정 양태에서, 본원에 기재된 하나 이상의 프로-영역 변이체의 아미노산 변형은 서열번호 15, 서열번호 9, 서열번호 11, 또는 서열번호 14의 프로-영역 아미노산 서열을 참조하여 넘버링된다.

[0040] 예를 들어, 본원에 기재된 하나 이상의 프로-영역 변이체의 아미노산 서열은 정렬 알고리즘을 사용하여 서열번호 15의 아미노산 서열과 정렬될 수 있으며, 서열번호 15의 아미노산 잔기와 정렬되는(바람직하게는 최적으로 정렬되는) 제공된 아미노산 서열의 각 아미노산 잔기는 편의상 해당 아미노산 잔기의 숫자 위치를 참조하여 넘버링된다. 서열 정렬 알고리즘은 쿼리 서열(때로는 "기준 서열"이라고도 함)과 비교할 때 대상 서열에서 삽입 또는 결실이 발생하는 위치 또는 위치들을 식별할 것이다. 다른 프로-영역 아미노산 서열과의 서열 정렬은 아미노산 정렬 알고리즘을 사용하여 확인할 수 있다.

[0041] 특정 양태에서, 프로-영역 아미노산 서열 "위치"에 대한 언급은 위치 번호가 뒤따라오는 단일 문자 아미노산(잔기)(예를 들어 서열번호 15, 위치 1의 알라닌은 "A1"로 표시, 위치 2의 글루탐산은 "E2"로 표시 등)로 표시될 수 있다. 관련 구현예에서, 변이체(돌연변이체) 프로-영역 아미노산 서열에 대한 언급은 단일 문자 아미노산(잔기)으로 표시될 수 있으며, 모체(기준) 서열의 아미노산 위치는 넘버링되고 동일한 위치의 돌연변이된 아미노산(잔기)이 뒤에 붙는다(예를 들어 서열번호 15, 글리신(G)으로 치환된 위치 1의 알라닌(A)은 "A1G"로 표시되고, 히스티딘(H)으로 치환된 위치 30의 글루탐산(E)은 "E30H"로 표시됨 등). 또한 다수의 아미노산 잔기가 프로-영역 아미노산 서열의 동일한 위치에서 치환될 수 있으며, 모체(기준) 서열의 아미노산 위치는 넘버링되고 앞슬래시로 구분된 동일한 위치의 돌연변이된 아미노산 잔기가 뒤에 붙는다(예를 들어 서열번호 15, 글리신(G), 글루탐산(E), 또는 라이신(K)으로 치환된 위치 1의 알라닌(A)은 "A1G/E/K"로 표시될 수 있음 등).

[0042] 본원에서 특정 범위는 용어 "약"이 선행되는 수치 값으로 제시된다. 본원에서 용어 "약"은 뒤에 오는 정확한 수 뿐만 아니라 이 용어 뒤에 오는 수에 근사하거나 대략적인 수를 문자 그대로 뒷받침하기 위하여 사용된다. 숫자가 구체적으로 언급된 숫자와 거의 동일한 숫자 또는 근사적인 숫자인지를 판단함에 있어서, 언급되지 않은 인근의 또는 근사치의 숫자는 그것이 제시된 문맥에서, 구체적으로 언급된 숫자와 실질적으로 등가인 숫자일 수 있다. 예를 들어, 수치 값과 관련하여, "약"이라는 용어는 이 용어가 문맥에서 달리 구체적으로 정의되지 않으면, 그 수치 값의 -10% 내지 +10%의 범위를 나타낸다.

[0043] 본원에서 사용되는 바와 같이, "PRO" 서열을 암호화하는 폴리뉴클레오티드, "프로-영역" 서열을 암호화하는 폴리뉴클레오티드, 및 "프로-서열"을 암호화하는 폴리뉴클레오티드는 상호교환적으로 사용될 수 있으며, 성숙 관

심 단백질(POI)을 암호화하는 다운스트림(3') 유전자 암호화 서열(CDS)에 작동가능하게 연결되고 바로 업스트림(5')에 위치한 DNA 서열을 지칭한다. 특정 양태에서, 프로-영역 서열을 암호화하는 폴리뉴클레오티드는 업스트림(5') 단백질 신호 서열(SS)과 성숙 POI를 암호화하는 다운스트림(3') 유전자 CDS 사이에 위치하고 작동가능하게 연결된 DNA 서열을 지칭한다. 예를 들어, 성숙 POI를 암호화하는 다운스트림(3') DNA 유전자 CDS에 작동가능하게 연결된 단백질 신호 서열을 암호화하는 업스트림(5') DNA 서열(SS)에 작동가능하게 연결된 프로-영역 DNA 서열(*PRO*)(예를 들어 5'-[SS]-[*PRO*]-[*ORF*]-3'; N-[pre]-[pro]-[단백질]-C).

- [0044] 본원에서 사용되는 바와 같이, "야생형(WT) *B. 켈투스* 프로-영역" 서열은 서열번호 15에 제시된 아미노산 서열을 포함한다. 특정 양태에서, WT(기준) 프로-영역 서열(서열번호 15)은 "E30 프로-서열" 또는 "E₃₀ 프로-서열"로 지칭될 수 있다.
- [0045] 본원에서 사용되는 바와 같이, "변이체 프로-영역 A"로 명명된 돌연변이체 프로-영역 서열은 서열번호 9에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 A 서열(서열번호 9)은 "H30 프로-서열" 또는 "H₃₀ 프로-서열"로 지칭될 수 있다.
- [0046] 본원에서 사용되는 바와 같이, "변이체 프로-영역 B"로 명명된 돌연변이체 프로-영역 서열은 서열번호 11에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 B 서열(서열번호 11)은 "G30 프로-서열" 또는 "G₃₀ 프로-서열"로 지칭될 수 있다.
- [0047] 본원에서 사용되는 바와 같이, "변이체 프로-영역 C"로 명명된 돌연변이체 프로-영역 서열은 서열번호 14에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 C 서열(서열번호 14)은 "GK 삽입 프로-서열" 또는 "G₂K₃ 삽입 프로-서열"로 지칭될 수 있다.
- [0048] 본원에서 사용되는 바와 같이, "변이체 프로-영역 D"로 명명된 돌연변이체 프로-영역 서열은 서열번호 29에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 D 서열(서열번호 29)은 "GKS 삽입 프로-서열" 또는 "G₂K₃S₄ 삽입 프로-서열"로 지칭될 수 있다.
- [0049] 본원에서 사용되는 바와 같이, "변이체 프로-영역 E"로 명명된 돌연변이체 프로-영역 서열은 서열번호 30에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 E 서열(서열번호 30)은 "GKA 삽입 프로-서열" 또는 "G₂K₃A₄ 삽입 프로-서열"로 지칭될 수 있다.
- [0050] 본원에서 사용되는 바와 같이, "변이체 프로-영역 F"로 명명된 돌연변이체 프로-영역 서열은 서열번호 31에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 F 서열(서열번호 31)은 "GKAA 삽입 프로-서열" 또는 "G₂K₃A₄A₅ 삽입 프로-서열"로 지칭될 수 있다.
- [0051] 본원에서 사용되는 바와 같이, "변이체 프로-영역 G"로 명명된 돌연변이체 프로-영역 서열은 서열번호 32에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 G 서열(서열번호 32)은 L68K/I72V 프로-서열로 지칭될 수 있다.
- [0052] 본원에서 사용되는 바와 같이, "변이체 프로-영역 H"로 명명된 돌연변이체 프로-영역 서열은 서열번호 32에 제시된 아미노산 서열을 포함한다. 특정 구현예에서, 변이체 프로-영역 H 서열(서열번호 33)은 L68K/E80I 프로-서열로 지칭될 수 있다.
- [0053] 도 3에 일반적으로 도시된 바와 같이, 아미노산(잔기) 위치 넘버링은 WT 프로-영역 아미노산 서열(E₃₀, 서열번호 15), 변이체 프로-영역 서열 A(H₃₀; 서열번호 9), 변이체 프로-영역 서열 B(G₃₀; 서열번호 11), 변이체 프로-영역 서열 C(GK 삽입, 서열번호 14), 변이체 프로-영역 서열 D(GKS 삽입, 서열번호 29), 변이체 프로-영역 서열 E(GKA 삽입, 서열번호 30), 변이체 프로-영역 서열 F(GKAA 삽입, 서열번호 31), 변이체 프로-영역 서열 G(L68K/I72V, 서열번호 32), 및/또는 변이체 프로-영역 서열 H(L68K/E80I, 서열번호 33)를 참조할 수 있다.
- [0054] 본원에서 사용되는 바와 같이, 문구 "전장 단백질을 암호화하는 폴리뉴클레오티드"는 "전구체" 단백질을 암호화하는 DNA 서열을 지칭하고, 문구 "성숙 단백질을 암호화하는 폴리뉴클레오티드"는 본원에 정의된 바와 같이, "성숙" 단백질을 암호화하는 DNA 서열을 지칭한다. 특정 양태에서, 전구체 단백질을 암호화하는 폴리뉴클레오티드는, 성숙 관심 단백질(POI)의 아미노산 서열을 암호화하는 다운스트림(3') DNA 서열(예를 들어 오픈 리딩 프레임, ORF)에 작동가능하게 연결된 프로-영역 아미노산 서열을 암호화하는 업스트림(5') DNA 서열을 적어도 포함한다. 다른 양태에서, 전구체 단백질을 암호화하는 폴리뉴클레오티드는, 성숙 POI의 아미노산 서열을 암호화

하는 다운스트림(3') ORF에 작동가능하게 연결된 프로-영역 아미노산 서열을 암호화하는 다운스트림(3') DNA 서열에 작동가능하게 연결된 단백질 신호 서열을 암호화하는 업스트림(5') DNA 서열을 적어도 포함한다.

- [0055] 본원에서 사용되는 바와 같이, 용어 "비번역 영역"은 "UTR"로 약칭될 수 있다.
- [0056] 본원에서 사용되는 바와 같이, 문구 "5 프라임(5') 비번역 영역", "5' 비번역 영역", 및/또는 "5' 전사 리더"는 상호교환적으로 사용되고 "5'-UTR"로 약칭될 수 있다. 당업계에서 일반적으로 이해되는 바와 같이, 5'-UTR은 개시 코돈으로부터 바로 업스트림(5')에 있는 메신저 RNA(mRNA)의 영역으로 알려져 있다.
- [0057] 핵산은 다른 핵산 서열과 기능적 관계에 있을 때 "작동가능하게 연결된다". 예를 들어, 분비 리더(즉, 신호 서열)를 암호화하는 DNA가 폴리펩티드의 분비에 참여하는 전단백질(pre-protein)로서 발현되는 경우에 해당 폴리펩티드를 암호화하는 DNA에 작동가능하게 연결되거나; 프로모터 또는 인핸서가 암호화 서열(CDS, ORF)의 전사에 영향을 미칠 경우 해당 서열에 작동가능하게 연결되거나; 리보솜 결합 부위가 번역을 촉진하도록 위치하는 경우 암호화 서열에 작동가능하게 연결된다. 일반적으로, "작동가능하게 연결된"은 연결되는 DNA 서열이 연속적이고, 분비 리더의 경우, 연속적이면서 리딩 단계에 있다는 것을 의미한다. 그러나, 인핸서는 연속적일 필요가 없다. 연결은 편리한 제한 부위에서 결합에 의해 달성된다. 이러한 부위가 존재하지 않는 경우, 합성 올리고뉴클레오타이드 어댑터 또는 링커가 통상적인 실시예에 따라 사용된다. 따라서, 용어 "작동가능하게 연결된"은 일반적으로 하나의 기능이 다른 기능에 의해 영향받도록 하는 단일 핵산 절편 상 핵산 서열의 결합(병치)을 나타낸다. 예를 들어, 프로모터(*pro*)는 유전자 CDS의 전사를 제어하는 경우 유전자 암호화 서열(*유전자 CDS*)에 작동가능하게 연결된다(예를 들어 5'-[*pro*]-[*유전자 CDS*]3').
- [0058] 본원에서 사용되는 바와 같이, 문구 "변이체 *rrnI*-P2 프로모터 및 5'-UTR 영역"(약어, "P2 프로모터/5'-UTR 영역")은 서열번호 1에 제시된 변이체 *B. 서브틸리스 rrnI*-P2 프로모터/5'-*aprE* UTR 영역 DNA 서열을 지칭한다.
- [0059] 본원에서 사용되는 바와 같이, "야생형 *B. 서브틸리스 aprE* 신호 펩티드 서열"을 암호화하는 DNA는 "*aprE* SS"로 약칭될 수 있으며, 서열번호 4의 뉴클레오타이드 서열을 포함한다.
- [0060] 본원에서 사용되는 바와 같이, "야생형 *B. 아밀로리케파시엔스 BPN* 종결자(*BPN* term)"는 "*term*"으로 약칭될 수 있으며, 서열번호 6의 뉴클레오타이드 서열을 포함한다.
- [0061] 본원에서 사용되는 바와 같이, 서열번호 3의 아미노산 서열을 포함하는 변이체 *바실러스 지브소니*(BG46) 서브틸리신은 "BG46_변이체 1"로 약칭될 수 있다.
- [0062] 본원에서 사용되는 바와 같이, 서열번호 8의 아미노산 서열을 포함하는 변이체 *바실러스 지브소니*(BG46) 서브틸리신은 "BG46_변이체 2"로 약칭될 수 있고, 이 BG46_변이체 2는 야생형 *B. 지브소니*(BG46) 서브틸리신 리포터 단백질에서 유래된 것이다(서열번호 2).
- [0063] 본원에서 사용되는 바와 같이, "적합한 조절 서열"은 암호화 서열의 업스트림(5' 비암호화 서열), 암호화 서열 내, 또는 암호화 서열의 다운스트림(3' 비암호화 서열)에 위치하며 관련 암호화 서열의 전사, RNA 프로세싱 또는 안정성, 또는 번역에 영향을 미치는 뉴클레오타이드 서열을 지칭한다. 조절 서열은 프로모터, 전사 리더 서열, RNA 프로세싱 부위, 이펙터 결합 부위, 및 스템-루프 구조를 포함할 수 있다.
- [0064] 본원에서 사용되는 바와 같이, "숙주 세포"는 새로 도입된 DNA 서열에 대한 숙주 또는 발현 비히클로서 작용하는 능력을 갖는 세포를 의미한다. 따라서, 본 발명의 특정 구현예에서, 숙주 세포는 그람 양성 세포(예를 들어, *바실러스* 종) 세포 및/또는 그람 음성 세포(예를 들어, *E. 콜라이*)이다.
- [0065] 본원에서 사용되는 바와 같이, "변형된 세포"는 변형된 세포가 유래된 모세포, 기준 세포, 또는 대조군 세포에 존재하지 않는 적어도 하나의 유전자 변형을 포함하는 재조합 세포를 의미한다.
- [0066] 본원에서 사용되는 바와 같이, "재조합(변형) 세포에서의 관심 단백질(POI)의 발현 및/또는 생산이 비변형(대조군) 세포에서의 동일한 POI의 발현 및/또는 생산과 비교될 때, 변형 세포와 비변형 세포가 동일한 조건(예를 들어, 배지, 온도, pH 등이 동일한 조건)하에 성장/배양/발효된다는 것은 이해될 것이다.
- [0067] 본원에서 사용되는 바와 같이, "재조합 세포는 비변형(대조군) 세포에 비해 관심 단백질의 '증가된 양을 발현/생산한다"와 같은 문구에서 사용될 때, "증가된 양"은 특히, 재조합 세포에서 발현/생산된 관심 단백질(POI)의 "증가된 양"을 나타내며, "증가된 양"은 항상, 동일한 POI를 발현/생산하는 비변형(대조군) 세포에 대해 상대적인 것이고, 여기서 변형된 세포와 비변형 세포는 동일한 조건하에 성장/배양/발효된다.
- [0068] 본원에서 사용되는 바와 같이, 단백질 생산의 "증가" 또는 "증가된" 단백질 생산은 증가된 양의 생산된 단백질

(예를 들어, 관심 단백질)을 의미한다. 단백질은 숙주 세포 내에서 생산되거나, 배양 배지 내로 분비(또는 수송)될 수 있다. 특정 구현예에서, 관심 단백질은 배양 배지 내로 생산(분비)된다. 예를 들어, 단백질 생산의 증가는 예를 들어 모체 숙주 세포와 비교하여 더 높은 최대 수준의 단백질 또는 효소 활성(예를 들어, 아밀라제 활성 등), 또는 생산된 총 세포의 단백질로서 검출될 수 있다.

[0069] 본원에서 사용되는 바와 같이, 용어 "변형" 및 "유전자 변형"은 상호교환적으로 사용되며, (a) 유전자(또는 이의 ORF)에서의 하나 이상의 뉴클레오티드의 도입, 치환, 또는 제거, 또는 유전자 또는 이의 ORF의 전사 또는 번역을 위해 요구되는 조절 요소에서의 하나 이상의 뉴클레오티드의 도입, 치환, 또는 제거, (b) 유전자 파괴, (c) 유전자 전환, (d) 유전자 결실, (e) 유전자의 하향조절, (f) 특이적 돌연변이유발, 및/또는 (g) 본원에 개시된 임의의 하나 이상의 유전자의 무작위 돌연변이유발을 포함한다.

[0070] 특정 양태에서, 유전자 변형은 특히 본 발명의 프로-영역(아미노산) 서열을 암호화하는 핵산(DNA) 서열에서 하나 이상의 뉴클레오티드의 도입, 치환, 또는 제거를 지칭한다. 예를 들어, 특정 구현예에서, 서열번호 15에 제시된 네이티브 프로-영역 아미노산 서열을 암호화하는 DNA 서열은 본원에 기재된 바와 같이 유전적으로 변형된다.

[0071] 본원에서 사용되는 바와 같이, "그람 양성 박테리아 세포 내로의 '유전자', '폴리뉴클레오티드', '오픈 리딩 프레임'(ORF), '유전자 암호화 서열', '벡터', '발현 카세트' 등의 도입"과 같은 문구에서 사용되는 바와 같은 용어 "도입"은 원형질체 융합, 천연 또는 인공 형질전환(예를 들어 염화칼슘, 전기천공), 형질도입, 형질감염, 접합 등을 포함하여(이에 한정되지 않음), 폴리뉴클레오티드(DNA)를 세포에 도입하기 위한 당업계에 알려진 방법을 포함한다.

[0072] 본원에서 사용되는 바와 같이, "형질전환"은 재조합 DNA 기술을 사용하여 세포가 형질전환되었음을 의미한다. 형질전환은 일반적으로 하나 이상의 뉴클레오티드 서열(예를 들어, 폴리뉴클레오티드, ORF, 또는 유전자)을 세포에 삽입함으로써 일어난다. 삽입된 뉴클레오티드 서열은 이중성 뉴클레오티드 서열(즉, 형질전환된 세포에서 자연적으로 발생하지 않는 서열)일 수 있다. 따라서, 형질전환은 일반적으로 외인성 DNA를 숙주 세포에 도입하여 DNA가 염색체 통합체 또는 자가 복제성 염색체의 벡터로서 유지되도록 하는 것을 의미한다.

[0073] 본원에서 사용되는 바와 같이, "형질전환 DNA", "형질전환 서열", 및 "DNA 구성체"는 서열을 숙주 세포 또는 유기체 내로 도입하는 데 사용되는 DNA를 지칭한다. 형질전환 DNA는 서열을 숙주 세포 또는 유기체 내로 도입하는 데 사용되는 DNA이다. 이러한 DNA는 PCR 또는 임의의 다른 적합한 기술에 의해 시험관내에서 생성될 수 있다. 일부 구현예에서, 형질전환 DNA는 유입 서열을 포함하는 반면, 다른 구현예에서는 상동성 박스에 의해 플랭킹된 유입 서열을 추가로 포함한다. 다른 추가 구현예에서, 형질전환 DNA는 말단에 추가된 다른 비상동성 서열(즉, 스테퍼 서열 또는 플랭크)을 포함한다. 말단은 형질전환 DNA가 예를 들어 벡터 내 삽입과 같이 닫힌 원을 형성하도록 닫힐 수 있다.

[0074] 본원에서 사용되는 바와 같이, "유전자의 파괴" 또는 "유전자 파괴"는 상호교환적으로 사용되며, 숙주 세포가 기능적 유전자 산물(예를 들어, 단백질)을 생산하는 것을 실질적으로 방지하는 임의의 유전자 변형을 광범위하게 지칭한다. 따라서, 본원에서 사용되는 바와 같이, 유전자 파괴는 프레임시프트 돌연변이, 조기 정지 코돈(즉, 기능적 단백질이 만들어지지 않음), 단백질의 활성을 제거하거나 감소시키는 치환, 내부 결실(즉, 기능적 단백질이 만들어지지 않음), 암호화 서열을 파괴하는 삽입, 전사에 필요한 천연 프로모터와 오픈 리딩 프레임 사이의 작동가능한 연결을 제거하는 돌연변이 등을 포함하나 이에 한정되지 않는다.

[0075] 본원에서 사용되는 바와 같이, "유입 서열"은 박테리아 세포 염색체 내로 도입되는 DNA 서열을 지칭한다. 일부 구현예에서, 유입 서열은 DNA 구성체의 일부이다. 다른 구현예에서, 유입 서열은 하나 이상의 관심 단백질을 암호화한다. 일부 구현예에서, 유입 서열은 형질전환된 세포의 게놈에 이미 존재할 수 있거나 존재하지 않을 수 있는 서열을 포함한다(즉, 상동성 서열이거나 이중성 서열일 수 있음). 일부 구현예에서, 유입 서열은 하나 이상의 관심 단백질, 유전자, 및/또는 돌연변이되었거나 변형된 유전자를 암호화한다. 대안적인 구현예에서, 유입 서열은 기능적 야생형 유전자 또는 오페론, 기능적 돌연변이 유전자 또는 오페론, 또는 비기능적 유전자 또는 오페론을 암호화한다. 일부 구현예에서, 유전자의 기능을 파괴하기 위해 비기능적 서열이 유전자에 삽입될 수 있다. 다른 구현예에서, 유입 서열은 선별 마커를 포함한다. 추가 구현예에서, 유입 서열은 2개의 상동성 박스를 포함한다.

[0076] 본원에서 사용되는 바와 같이, "상동성 박스"는 박테리아 세포 염색체 내의 서열에 대해 상동성인 핵산 서열을 지칭한다. 보다 구체적으로, 상동성 박스는 본 발명에 따라 결실, 파괴, 비활성화, 하향조절 등이 될 유전자 또

는 유전자의 일부의 인접 플랭킹 암호화 영역과 약 80% 내지 100%의 서열 동일성, 약 90% 내지 100%의 서열 동일성, 또는 약 95% 내지 100%의 서열 동일성을 갖는 업스트림 또는 다운스트림 영역이다. 이러한 서열은 DNA 구성체가 박테리아 세포 염색체에서 어느 위치에 통합되는지를 지시하고, 염색체의 어느 부분이 유입 서열에 의해 대체되는지를 지시한다. 본 발명을 한정하고자 하는 것은 아니지만, 상동성 박스는 약 1 염기쌍(bp) 내지 200 킬로염기(kb)를 포함할 수 있다. 바람직하게는, 상동성 박스는 약 1 bp 내지 10.0 kb; 1 bp 내지 5.0 kb; 1 bp 내지 2.5 kb; 1 bp 내지 1.0 kb, 및 0.25 kb 내지 2.5 kb를 포함한다. 상동성 박스는 또한 약 10.0 kb, 5.0 kb, 2.5 kb, 2.0 kb, 1.5 kb, 1.0 kb, 0.5 kb, 0.25 kb, 및 0.1 kb를 포함할 수 있다. 일부 구현예에서, 선별 마커의 5' 및 3' 말단은 상동성 박스에 의해 플랭킹되고, 상동성 박스는 유전자의 암호화 영역을 바로 플랭킹하는 핵산 서열을 포함한다.

[0077] 본원에서 사용되는 바와 같이, 숙주 세포 "계놈", 박테리아 (숙주) 세포 "계놈", 또는 *바실러스* 종 (숙주) 세포 "계놈"은 염색체 유전자 및 염색체외 유전자를 포함한다.

[0078] 본원에서 사용되는 바와 같이, 용어 "플라스미드", "벡터" 및 "카세트"는 염색체의 요소를 지칭하며, 이는 종종 전형적으로 세포의 중심 대사의 일부가 아니고 보통 원형 이중 가닥 DNA 분자의 형태인 유전자를 운반한다. 이러한 요소는 선택된 유전자 산물에 대한 DNA 서열 및 프로모터 단편을 적절한 3' 비번역 서열과 함께 세포 내로 도입할 수 있는 독특한 구성체로 다수의 뉴클레오타이드 서열이 연결되거나 재조합된, 임의의 공급원 유래의 단일 가닥 또는 이중가닥 DNA 또는 RNA의 선형 또는 원형, 자율 복제 서열, 계놈 통합 서열, 파지 또는 뉴클레오타이드 서열일 수 있다.

[0079] 본원에서 사용되는 바와 같이, 용어 "플라스미드"는 클로닝 벡터로서 사용되며 많은 박테리아 및 일부 진핵 생물에서 염색체의 자가 복제성 유전 요소를 형성하는 원형 이중가닥(ds) DNA 구성체를 지칭한다. 일부 구현예에서, 플라스미드는 숙주 세포의 계놈에 통합된다. 일부 구현예에서, 플라스미드는 모세포에 존재하고 딸세포에서 소실된다.

[0080] 본원에서 사용되는 바와 같이, "형질전환 카세트"는 유전자(또는 이의 ORF)를 포함하고 외래 유전자 이외에 특정 숙주 세포의 형질전환을 용이하게 하는 요소를 갖는 특정 벡터를 나타낸다.

[0081] 본원에서 사용되는 바와 같이, "벡터"라는 용어는 세포에서 복제(증식)될 수 있고 새로운 유전자 또는 DNA 절편을 세포 내로 운반할 수 있는 임의의 핵산을 지칭한다. 따라서, 이 용어는 상이한 숙주 세포들 간의 전달을 위해 설계된 핵산 구성체를 지칭한다. 벡터는 바이러스, 박테리오파지, 프로-바이러스(pro-virus), 플라스미드, 파지미드(phagemid), 트랜스포존(transposon), 및 인공 염색체, 예컨대 YAC(효모 인공 염색체), BAC(박테리아 인공 염색체), PLAC(식물 인공 염색체) 등을 포함하는데, 이들은 "에피솜"이다(즉, 자율적으로 복제되거나 숙주 생물의 염색체 내에 통합될 수 있음).

[0082] "발현 벡터"는 세포에서 이중성 DNA를 통합하고 발현시키는 능력을 갖는 벡터를 지칭한다. 많은 원핵 및 진핵 발현 벡터가 상업적으로 이용가능하고 당업자에게 알려져 있다. 적절한 발현 벡터를 선택하는 것은 당업자가 알고 있다.

[0083] 본원에서 사용되는 바와 같이, 용어 "발현 카세트" 및 "발현 벡터"는 표적 세포에서 특정 핵산의 전사를 가능케 하는 일련의 명시된 핵산 요소(즉, 전술한 바와 같은 벡터 또는 벡터 요소)를 사용해 재조합적으로 또는 합성에 의해 생성된 핵산 구성체를 지칭한다. 재조합 발현 카세트는 플라스미드, 염색체, 미토콘드리아 DNA, 색소체 DNA, 바이러스, 또는 핵산 단편에 통합될 수 있다. 일반적으로, 발현 벡터의 재조합 발현 카세트 부분은 서열들 중에서도 특히, 전사될 핵산 서열 및 프로모터를 포함한다. 일부 구현예에서, DNA 구성체는 표적 세포에서 특정 핵산의 전사를 가능하게 하는 일련의 명시된 핵산 요소를 또한 포함한다. 특정 구현예에서, 본 발명의 DNA 구성체는 본원에 정의된 바와 같은 선별 마커 및 비활성화 염색체 또는 유전자 또는 DNA 절편을 포함한다.

[0084] 본원에서 사용되는 바와 같이, "표적화 벡터"는 표적화 벡터가 형질전환되는 숙주 세포의 염색체 내 영역에 대해 상동성이고, 그 영역에서 상동성 재조합을 유도할 수 있는 폴리뉴클레오타이드 서열을 포함하는 벡터이다. 예를 들어, 표적화 벡터는 상동성 재조합을 통해 숙주 세포의 염색체에 돌연변이를 도입하는 데 사용된다. 일부 구현예에서, 표적화 벡터는 예를 들어 말단에 추가된 다른 비상동성 서열(즉, 스터퍼 서열 또는 플랭킹 서열)을 포함한다. 말단은 표적화 벡터가 예를 들어 벡터 내 삽입과 같이 단편 원을 형성하도록 단할 수 있다. 예를 들어, 특정 구현예에서, 모체 *B. 리철티포르미스* (숙주) 세포는 하나 이상의 "표적화 벡터"를 도입함으로써 변형(예를 들어, 형질전환)된다.

[0085] 본원에서 사용되는 바와 같이, 용어 "관심 단백질" 또는 "POI"는 변형(재조합) 그람 양성 숙주 세포에서 발현되

기를 원하는 관심 폴리펩티드를 지칭하며, POI는 바람직하게는 증가된(즉, "비변형" (모 또는 대조군) 세포에 비해 증가된) 수준으로 발현된다. 따라서, 본원에서 사용되는 바와 같이, POI는 효소, 기질-결합 단백질, 표면-활성 단백질, 구조 단백질, 수용체 단백질 등일 수 있다. 특정 구현예에서, 본 발명의 변형된 세포는 대조군 세포에 비해 증가된 양의 이중성 관심 단백질을 생산한다. 특정 구현예에서, 본 발명의 변형된 세포에 의해 생산된 관심 단백질의 증가된 양은 대조군 세포에 비해 0.5% 이상의 증가, 1.0% 이상의 증가, 5.0% 이상의 증가, 또는 5.0% 초과 증가이다.

[0086] 마찬가지로, 본원에서 정의되는 바와 같이, "관심 유전자" 또는 "GOI"는 POI를 암호화하는 핵산 서열(예를 들어, 폴리뉴클레오티드, 유전자, 또는 ORF)을 지칭한다. "관심 단백질"을 암호화하는 "관심 유전자"는 자연 발생적 유전자, 돌연변이 유전자, 또는 합성 유전자일 수 있다.

[0087] 본원에서 사용되는 바와 같이, 용어 "폴리펩티드" 및 "단백질"은 상호교환적으로 사용되며, 펩티드 결합에 의해 연결된 아미노산 잔기를 포함하는 임의의 길이의 중합체를 지칭한다. 아미노산 잔기에 대한 통상적인 1문자 또는 3문자 코드가 본원에서 사용된다. 폴리펩티드는 선형 또는 분지형일 수 있고, 변형된 아미노산을 포함할 수 있으며, 비아미노산이 개재될 수 있다. 폴리펩티드라는 용어는 자연적으로 변형되었거나, 이황화 결합 형성, 글리코실화, 지질화, 아세틸화, 인산화, 또는 임의의 다른 조작 또는 변형, 예컨대, 표지 성분과의 접합과 같은 개입에 의해 변형된 아미노산 중합체를 또한 포함한다. 예를 들어, 아미노산의 하나 이상의 유사체(예컨대, 비천연 아미노산 등을 포함함)뿐만 아니라 당업계에 알려진 다른 변형을 포함하는 폴리펩티드도 정의에 포함된다.

[0088] 특정 구현예에서, 본 발명의 유전자는 상업적으로 관련된 산업적 관심 단백질, 예컨대 효소(예를 들어, 아세틸 에스테라제, 아미노펩티다제, 아밀라제, 아라비나제, 아라비노푸라노시다제, 탄산무수화효소, 카복시펩티다제, 카탈라제, 셀룰라제, 키티나제, 키모신, 큐티나제, 데옥시리보뉴클레아제, 에피머라제, 에스테라제, α -갈락토시다제, β -갈락토시다제, α -글루카나제, 글루칸 라이아제, 엔도- β -글루카나제, 글루코아밀라제, 글루코스 옥시다제, α -글루코시다제, β -글루코시다제, 글루쿠로니다제, 글리코실 하이드롤라제, 헤미셀룰라제, 헥소스 옥시다제, 하이드롤라제, 인버타제, 이소머라제, 라카제, 리파제, 리아제, 만노시다제, 옥시다제, 옥시도리덕타제, 펙테이트 리아제, 펙틴 아세틸 에스테라제, 펙틴 데폴리머라제, 펙틴 메틸 에스테라제, 펙틴 분해 효소, 퍼하이드롤라제, 폴리올 옥시다제, 퍼옥시다제, 페놀옥시다제, 피타제, 폴리갈락투로나제, 프로테아제, 펩티다제, 람노-갈락투로나제, 리보뉴클레아제, 트랜스퍼라제, 수송 단백질, 트랜스글루타미나제, 자일라나제, 헥소스 옥시다제, 및 이들의 조합)를 암호화한다.

[0089] 본원에서 사용되는 바와 같이, "변이체" 폴리펩티드는, 일반적으로 재조합 DNA 기술에 의한 하나 이상의 아미노산의 치환, 추가, 또는 결실에 의해 모체(또는 기준) 폴리펩티드로부터 유래되는 폴리펩티드를 지칭한다. 변이체 폴리펩티드는 적은 수의 아미노산 잔기만큼 모체 폴리펩티드와 상이할 수 있으며, 모체(기준) 폴리펩티드와의 일차 아미노산 서열 상동성/동일성 수준에 의해 정의될 수 있다.

[0090] 하나 이상의 특정 구현예에서, 변이체 폴리펩티드는 모체(기준) 폴리펩티드 서열과 적어도 약 40% 내지 약 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 심지어 적어도 99%의 아미노산 서열 동일성을 갖는다.

[0091] 본원에서 사용되는 바와 같이, "변이체" 폴리뉴클레오티드는 모체(또는 기준) 폴리뉴클레오티드와 명시된 정도의 서열 상동성/동일성을 갖거나, 엄격한 혼성화 조건하에 모체 폴리뉴클레오티드(또는 이의 상보체)와 혼성화되는 폴리뉴클레오티드를 지칭한다. 하나 이상의 특정 구현예에서, 본 발명의 변이체 폴리뉴클레오티드는 모체(기준) 폴리뉴클레오티드 서열에 대해 적어도 약 40% 내지 적어도 약 100%의 뉴클레오티드 서열 동일성을 포함한다. 다른 특정 구현예에서, 변이체 폴리뉴클레오티드는 모체(기준) 폴리뉴클레오티드 서열에 대해 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 적어도 99%의 뉴클레오티드 서열 동일성을 포함한다.

[0092] 본원에서 사용되는 바와 같이, "돌연변이"는 핵산 서열에서의 임의의 변화 또는 변경을 나타낸다. 점 돌연변이, 결실 돌연변이, 침묵 돌연변이, 프레임 시프트 돌연변이, 스플라이싱 돌연변이 등을 포함하는 여러 가지 유형의 돌연변이가 존재한다. 돌연변이는 특이적으로(예를 들어, 부위 지정 돌연변이유발을 통해) 또는 무작위로(예를 들어, 화학작용제, 복구 결핍된 박테리아 균주를 통한 계대배양을 통해) 수행될 수 있다.

[0093] 본원에서 사용되는 바와 같이, 폴리펩티드 또는 이의 서열의 맥락에서, 용어 "치환"은 하나의 아미노산을 또 다른 아미노산으로 대체(즉, 치환)하는 것을 의미한다.

- [0094] 본원에서 사용되는 바와 같이, 용어 "상동성"은 상동 폴리뉴클레오타이드 또는 폴리펩티드와 관련된다. 둘 이상의 폴리뉴클레오타이드 또는 둘 이상의 폴리펩티드가 상동성일 경우, 이는 상동 폴리뉴클레오타이드 또는 폴리펩티드가 적어도 60%, 더 바람직하게는 적어도 70%, 한층 더 바람직하게는 적어도 85%, 훨씬 더 바람직하게는 적어도 90%, 더욱 바람직하게는 적어도 95%, 및 가장 바람직하게는 적어도 98%의 "동일성 정도"를 갖는 것을 의미한다. 2개의 폴리뉴클레오타이드 또는 폴리펩티드 서열이 본원에서 정의된 바와 같이 상동성이 되기에 충분히 높은 정도의 동일성을 갖는지 여부는 본 기술 분야에 공지된 컴퓨터 프로그램, 예를 들어 GCG 프로그램 패키지로 제공되는 "GAP"(Program Manual for the Wisconsin Package, 버전 8, 1994년 8월, [53711] 미국 위스콘신주 매디슨 사이언스 드라이브 575 소재의 제네틱스 컴퓨터 그룹(Genetics Computer Group))을 사용하여 두 서열을 정렬함으로써 적절히 조사할 수 있다(문헌[Needleman and Wunsch, (1970)]). DNA 서열 비교를 위하여 하기 설정을 갖는 GAP를 사용한다: 5.0의 GAP 생성 페널티 및 0.3의 GAP 연장 페널티.
- [0095] 본원에서 사용되는 바와 같이, 용어 "동일성 백분율(%)"은, 폴리펩티드를 암호화하는 핵산 서열들 또는 폴리펩티드의 아미노산 서열들을 서열 정렬 프로그램을 사용하여 정렬할 때, 이들 사이의 핵산 서열 또는 아미노산 서열의 동일성 수준을 지칭한다.
- [0096] 본원에서 사용되는 바와 같이, 용어 "비생산성(specific productivity)"은 주어진 기간에 걸쳐 시간당 세포당 생산된 단백질의 총량이다.
- [0097] 본원에서 사용되는 바와 같이, "정제된", "단리된" 또는 "풍부화된"이라는 용어는, 생체분자(예를 들어, 폴리펩티드 또는 폴리뉴클레오타이드)가 자연에서 이들과 결합되어 있는 자연 발생 구성 성분 중 일부 또는 전부로부터 분리됨으로써 이의 자연 상태에서부터 변경되는 것을 의미한다. 이러한 단리 또는 정제는 최종 조성물에서 바람직하지 않은 전체 세포, 세포 잔여물, 불순물, 외부 단백질, 또는 효소를 제거하기 위한 이온 교환 크로마토그래피, 친화성 크로마토그래피, 소수성 분리, 투석, 프로테아제 처리, 황산암모늄 침전 또는 기타 단백질 염 침전, 원심분리, 크기 배제 크로마토그래피, 여과, 미세여과, 겔 전기영동 또는 구배에 의한 분리와 같은 본 기술 분야에서 인정된 분리 기술에 의해 달성될 수 있다. 이후, 추가 이득을 제공하는 구성 성분, 예를 들어 활성화제, 항저해제, 바람직한 이온, pH를 제어하기 위한 화합물 또는 기타 효소 또는 화학물질을 정제되거나 단리된 생체분자 조성물에 첨가하는 것이 추가로 가능하다.
- [0098] **II. 단백질 생산 강화를 위한 돌연변이체 프로-서열**
- [0099] 본원에 일반적으로 기재되고 하기 실시예에 제시된 바와 같이, 출원인은 재조합 단백질의 생산성 강화를 위해 특정 유전자 변형(돌연변이)을 테스트/스크리닝하기 위한 부위 평가 라이브러리(SEL)를 설계 및 구축하였다. 특히, 실시예에 제시된 바와 같이, 재조합 단백질(효소)은 본 발명의 재조합 *바실러스* 균주에서 단백질 발현을 모니터링하기 위해 리포터로 사용되었다. 이종성(재조합) 성숙 단백질의 발현/생산에 사용하기에 적합한 특정 프로-영역 서열이 기재된 바 있다. 예를 들어, PCT 공개 W02008/112258호 및 PCT 공개 2010/123754호는 특정 돌연변이체 프로 서열은 성숙 프로테아제의 생산성을 향상시킨 *B. 클라우시* 알칼리성 전구체 프로테아제와 *B. 렌투스* 서브틸리신 전구체 프로테아제의 프로 서열을 기술하고 있다. PCT 공개 W01993/20214호는 *B. 서브틸리스*에서 그람 음성 리파제의 발현을 위한 특정 AprE 신호(pre) 및 프로-영역 서열을 기술하고 있다.
- [0100] 보다 구체적으로, 하기 실시예에 기재된 바와 같이, 유전자 변형(SEL)을 대표적인(서브틸리신) 리포터 단백질을 암호화하는 발현 구성체의 프로-영역에서 수행하였다. 특히, 실시예에 일반적으로 기재된 바와 같이(도 1~3 참조), SEL을 야생형(기준) 프로-서열(서열번호 15)로부터 유래된 프로-영역 서열에서 수행하여 성숙 서브틸리신 리포터 단백질의 생산성을 강화하는 돌연변이를 스크리닝하였다.
- [0101] 실시예 1에 기재된 바와 같이, 변이체 프로-영역 서열 A(30H, 서열번호 9; 도 1)를 암호화하는 아미노산 서열을 템플릿으로 사용하여 부위 평가 라이브러리(SEL)를 생성하고 4.4 kb 단편으로 개발하였으며, 발현 카세트의 선형 DNA를 사용하여 적격한 *바실러스* 세포를 형질전환시켰다. 실시예 1에 기재된 바와 같이, SEL에서 고유한 프로-영역 서열 A 변이체 및/또는 프로-영역 서열 B 변이체를 확인하기 위해 서열 분석을 수행하고, 변이체 프로-영역 서열 A(서열번호 9)의 84개 아미노산 위치 각각을 다른 19개 자연 발생 아미노산 잔기로 변경(치환)하였다. 특히, 표 1(실시예 1)은 리포터 단백질이 WT PRO 서열(서열번호 15)로 발현된 기준 구성체와 비교하여 72시간 후의 리포터 단백질 생산성(성능 지수(PI) 값)의 결과를 나타낸 것이다. 예를 들어, 표 1에 도시된 바와 같이, 프로-영역 서열 A(E30H, 서열번호 9)의 PI는 WT PRO 서열(30E; 서열번호 15)의 PI에 비해 약 50% 증가하였다. 마찬가지로, 표 1에 표시된 바와 같이, 위치 30과 조합된 아미노산 위치 1, 2, 3, 4, 6, 14, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 83, 또는 84 중 어느 하나에서의 돌연변이는 기준 프로-영역 서열(서열번호 15)에 비해 72시간에서 향상된 리포터 단백질 생산성을 입증하였다.

또한, 선택된 돌연변이체 조합의 약 80%는 추가 양 전하를 가졌다(표 1, D 전하 참조).

[0102] 실시예 3에서 추가로 기재된 바와 같이, 유전자 변형(SEL)을 대표적인(서브틸리신) 리포터 단백질(BG46_변이체 2)을 암호화하는 발현 구성체의 프로-영역에서 수행하였다. 보다 구체적으로, 변이체 프로-영역 서열 B(30G, 서열번호 11; 도 1)를 암호화하는 아미노산 서열은 본 발명의 하나 이상의 프로-영역을 포함하는 선형 DNA 발현 카세트를 개발하기 위한 템플릿으로 사용되었다. 특히, 78개 아미노산 야생형 AprE 프로펩티드 서열(서열번호 12)의 5' 서열을 사용하여 첫 번째 3개 아미노산 잔기(위치 1~3, "AGK"; 서열번호 12)를 알라닌(A)인 첫 번째 아미노산 잔기로 교환하여 변이체 프로-영역 서열 C(서열번호 14; GK 삽입, 도 2)를 생성하였다. 예를 들어, 표 1(실시예 1)에 도시된 바와 같이, 프로-영역 변이체 A 서열의 돌연변이체(서열번호 9; 표 1, 돌연변이체 "E30G")는 야생형 프로-영역(표 1, PI 1.0) 및 프로-영역 변이체 A 서열(서열번호 11; 표 1, 돌연변이체 "E30H", PI 1.5)과 비교하여 72시간의 성장 후 리포터 단백질의 높은 발현을 보였다(PI 2.3).

[0103] 본 발명의 실시예 4는 아미노산 글리신(G) 및 라이신(K)이 삽입된("GK") 기준 프로-영역 서열(서열번호 15)을 기반으로 하는 조합 라이브러리의 제조를 추가로 기술한다. 보다 구체적으로, 실시예 3에 기재된 바와 같이, 아미노산 잔기 GK를 야생형 프로-영역 서열(서열번호 15)의 N-말단에 삽입하여, 도 2 정렬에 도시된 86개 아미노산 위치를 포함하는, 서열번호 14에 제시된 프로-영역 변이체 C 아미노산 서열을 생성하였다. 예를 들어, 표 2(실시예 4)는 BG46_변이체 2 리포터 단백질의 증가된 생산성(PI)을 보인 프로-영역 돌연변이의 조합을 제시한 것으로, 여기서 전하(D 전하)의 변화는 일반적으로 +1과 +4 사이였다(표 2). 특정 양태에서, 출원인은 이러한 양 전하가 프로-영역 서열의 루프 영역에서 조합으로 적합하다는 것에 주목한다(예를 들어 기준 서열번호 15의 잔기 위치 66~73).

[0104] 표 2에 도시된 바와 같이, SEL 프로-영역 변이체 C 서열(서열번호 14로부터 유래됨)의 PI 값을 야생형(WT) 프로-영역 서열(서열번호 15)의 PI 값과 비교하고, 여기서 아미노산 위치 넘버링(표 2, 제1 열)을 WT 프로-영역 서열(즉, "AGK" 삽입 없음)과 비교한다. 따라서, 실시예 4에 기재된 바와 같이, 표 2에 제시된 프로-영역 변이체 C 서열은 야생형 프로-영역(표 2, PI 1.0) 및 프로-영역 변이체 B 서열(돌연변이체 "E30G", PI 1.7)과 비교하여 72시간의 성장 후 리포터 단백질의 높은 발현을 보였다(PI 값은 약 1.9~2.3 사이).

[0105] 실시예 5에 일반적으로 기재된 바와 같이, BG46_변이체 2 서브틸리신은 *B. 리첸니포르미스* 균주에서와 같이 발현을 모니터링하기 위한 리포터 단백질로 사용되었다. 보다 구체적으로, 본 발명의 변이체 프로-영역 서열을 암호화하는 DNA를 성숙 BG46_변이체 2 단백질을 암호화하는 DNA 서열에 작동가능하게 연결하고, 균주를 일반적으로 실시예 5에 기재된 바와 같이 구성하였다. 예를 들어, 변이체 프로-영역 서열을 포함하는 *B. 리첸니포르미스* 균주의 PI 값은 E30G 돌연변이를 포함하는 대조군 프로-영역 서열 B(서열번호 11)와 대비하여 표 3에 도시되어 있으며, 여기서 변이체 프로-영역 서열은 대조군 프로-영역 서열(E30G)과 비교하여 72시간의 성장 후 리포터 단백질의 높은 발현을 보였다.

[0106] 실시예 6에 추가로 기재된 바와 같이, 변이체 프로-영역 서열 C(서열번호 14)의 추가 N-말단 변형을 수행하였다. 예를 들어, 변이체 프로-영역 서열 C(86개 아미노산 위치를 포함)를 위치 3에 라이신(K) 뒤에 알라닌(A)이나 세린(S)과 같은 추가 아미노산을 도입(삽입)하기 위해 추가로 변형하였다. 특히, 실시예 6에 기재된 바와 같이(도 1~3에 도시됨), 변이체 프로-영역 서열 D(서열번호 29, GKS 삽입), E(서열번호 30, GKA 삽입), 및 F(GKAA 삽입; 서열번호 31)를 구성하였고, 프로-영역 변이체 서열 D, E, 및 F(표 4)의 성능 지수(PI) 값을 변이체 프로-영역 서열 C(서열번호 14, "GK" 삽입)와 비교한다. 표 4에 제시된 바와 같이, 3개의 N-말단 프로-영역 돌연변이체(GKA, GKS, GKAA)의 PI 지수는 프로-영역 서열 C(GK; 서열번호 14)와 비교하여 72시간 성장 후 리포터 단백질 BG46_변이체 2 서브틸리신(서열번호 8)의 더 높은 발현을 입증하였다.

[0107] 마찬가지로, 프로-영역 서열의 특정 C 말단 변형을 실시예 7에 기재된 바와 같이 수행하였다. 특히, 도 1~3에 도시된 바와 같이, 위치 68의 류신(L) 잔기를 라이신(K)으로 치환하고 위치 72의 이소류신(I)을 발린(V)으로 치환하여 변이체 프로-영역 서열 G(서열번호 32)를 생성하거나, 위치 68의 류신(L) 잔기를 라이신(K)으로 치환하고 위치 72(80)의 글루탐산(E)을 이소류신(I)으로 치환하여 변이체 프로-영역 서열 F(서열번호 33)를 생성하도록 변이체 프로-영역 서열 B(서열번호 11)를 추가로 조작하였다. 표 5에 제시된 바와 같이, 변이체 프로-영역 서열 G(서열번호 32) 및 H(서열번호 33)의 PI 값은 기준(대조군) 변이체 프로-영역 B 서열(서열번호 11)과 비교하여 증가하였다.

[0108] 따라서, 본 발명의 특정 구현에는 특히, 신규한 프로-영역 서열을 암호화하는 재조합 폴리뉴클레오티드, 성숙 관심 단백질을 암호화하는 DNA 서열에 작동가능하게 연결된 신규한 프로-영역 서열을 암호화하는 DNA 서열을 포함하는 발현 카세트, 신규한 프로-영역 서열을 포함하는 전구체 단백질을 암호화하는 폴리뉴클레오티드를 발현

하는 변형된 그람 양성 박테리아 세포/균주 등을 제공한다. 특정 양태에서, 본 발명의 신규한 변이체(돌연변이체) 프로-영역 서열은 서열번호 15의 야생형 *B. 켈투스* 프로-서열로부터 유래된 아미노산 서열을 포함한다.

[0109] 본원에서 사용되는 바와 같이, 문구 "변이체 프로-영역"은, 일반적으로 재조합 DNA 기술에 의한 하나 이상의 아미노산의 치환, 추가, 또는 결실에 의해 기준 프로-영역 서열로부터 유래된 폴리펩티드 서열을 지칭한다. 변이체 프로-영역(아미노산) 서열은 소수의 아미노산 잔기만큼 기준(모체) 프로-영역 서열과 다를 수 있으며, 기준(모체) 프로-영역 서열과의 주요 아미노산 서열 상동성/동일성 수준으로 정의될 수 있다.

[0110] 앞서 제시된 바와 같이, 두 폴리뉴클레오티드 또는 폴리펩티드 서열과 관련하여 "동일한"이라는 용어는, 하기에 기술되고 당업계에 알려진 서열 비교 또는 분석 알고리즘을 사용하여 측정된 바와 같이, 최대 대응성을 위해 정렬했을 때 두 서열의 뉴클레오티드 또는 아미노산이 동일하다는 것을 의미한다. 문구 "동일성 백분율(%)"(약어, "PID")은 폴리뉴클레오티드(핵산) 또는 폴리펩티드(아미노산) 서열 동일성을 지칭한다. 동일성 백분율은 당업계에 알려진 표준 기법을 이용해서 결정될 수 있다. 특정 양태에서, 관심 서열에 의해 공유되는 아미노산 동일성 백분율은 서열 정보를 직접 비교하기 위해 서열을 정렬하여 측정할 수 있다(예를 들어, BLAST, MUSCLE, 또는 CLUSTAL과 같은 정렬 프로그램/알고리즘을 사용). 예를 들어, BLAST 알고리즘은 문헌[Altschul *et al.* (1990) and Karlin *et al.* (1993)]에 기재되어 있다. 특히, 아미노산 서열 동일성 백분율(%)값은 최적/최대 정렬을 위해 프로그램에 의해 생성된 임의의 갭을 포함하는 "기준" 서열의 총 잔기수로 나눈 매칭하는 동일한 잔기의 수에 의해 결정된다. BLAST 알고리즘은 "기준" 서열을 "쿼리" 서열로 나타낸다.

[0111] 본원에서 사용되는 바와 같이, "상동성 프로-영역"은 일차, 이차, 및/또는 삼차 구조에서 구별되는 유사성을 갖는 프로-영역을 지칭한다. 단백질 상동성은 단백질이 정렬되는 경우, 선형 아미노산 서열에서의 유사성을 나타낼 수 있다. 상동성은, 예컨대 BLAST, MUSCLE, 또는 CLUSTAL과 같은 프로그램을 이용하여, 아미노산 서열 정렬에 의해 결정될 수 있다. 단백질 서열의 상동성 검색은 NCBI BLAST의 BLASTP 및 PSI-BLAST를 사용하여 역치(E-값 컷오프)가 0.001인 상태에서 수행할 수 있다(예를 들어 문헌[Altschul *et al.*, 1997] 참조).

[0112] BLAST 프로그램에서는 대부분 디폴트 값으로 설정된 몇 가지 검색 파라미터를 사용한다. NCBI BLAST 알고리즘은 생물학적 유사성의 관점에서 가장 관련된 서열을 찾지만, 20개 미만 잔기의 쿼리 서열에 대해서는 권장되지 않는다(Altschul *et al.*, 1997 and Schaffer *et al.*, 2001). 핵산 서열 검색을 위한 예시적인 디폴트 BLAST 파라미터에는 다음이 포함된다: 이웃 단어 역치 = 11; E-값 컷오프 = 10; 스코어링 매트릭스 = NUC.3.1(매치 = 1, 미스매치 = -3); 갭 개방 = 5; 및 갭 연장 = 2. 아미노산 서열 검색을 위한 예시적인 디폴트 BLAST 파라미터에는 다음이 포함된다: 단어 크기 = 3; E-값 컷오프 = 10; 스코어링 매트릭스 = BLOSUM62; 갭 개방 = 11; 및 갭 연장 = 1. 이 정보를 사용하면 단백질 서열의 그룹화 및/또는 이를 바탕으로 계통학적 트리를 구축할 수 있다. 아미노산 서열은 프로그램, 예컨대 Vector NTI Advance suite에 입력될 수 있고 GuideTree가 이웃 연결(Neighbor Joining, NJ) 방법(Saitou and Nei, 1987)을 이용하여 생성될 수 있다. 트리 구축은 서열 거리에 대한 Kimura 교정을 이용하고 갭을 포함하는 위치를 무시하여 계산될 수 있다. 프로그램, 예컨대 AlignX는 계통학적 트리 상에 디스플레이된 분자명 뒤 괄호에 계산된 거리값을 디스플레이할 수 있다.

[0113] 분자 간 상동성의 이해는 분자의 진화 역사뿐만 아니라 이의 기능에 대한 정보를 드러낼 수 있다; 새로 시퀀싱된 단백질이 이미 특성규명된 단백질과 상동성인 경우, 새로운 단백질의 생화학적 기능을 강력히 시사한다. 두 분자가 공통 원형으로부터 유래되었을 경우 이를 상동이라고 한다. 상동성 분자, 또는 상동체는 파라로그(paralog) 및 오르소로그(ortholog)의 2개 클래스로 구분될 수 있다. 파라로그는 하나의 종 내에 존재하는 상동체이다. 파라로그는 종종 이의 상세한 생화학적 기능이 상이하다. 오르소로그는 상이한 종 내에 존재하는 상동체이며, 매우 유사하거나 동일한 기능을 갖는다. 단백질 슈퍼패밀리는 공통 가계가 추정될 수 있는 단백질의 최대 그룹화(분기)이다. 보통 이 공통 가계는 서열 정렬 및 기계적 유사성에 기반한다.

[0114] CLUSTAL W 알고리즘은 서열 정렬 알고리즘의 또 다른 예이다(Thompson *et al.*, 1994). CLUSTAL W 알고리즘에 대한 기본 파라미터에는 다음이 포함된다: 갭 개방 페널티 = 10.0; 갭 연장 페널티 = 0.05; 단백질 가중치 매트릭스 = BLOSUM 시리즈; DNA 가중치 매트릭스 = IUB; Delay 확산 서열% = 40; 갭 분리 거리 = 8; DNA 변화 가중치 = 0.50; 친수성 잔기 목록 = GPSNDQEKR; 음성 매트릭스 이용 = OFF; Toggle 잔기 특정 페널티 = ON; Toggle 친수성 페널티 = ON; 및 Toggle 말단 갭 분리 페널티 = OFF. CLUSTAL 알고리즘에서는, 어느 한 말단에 생기는 결실이 포함된다. 예를 들어, 500개 아미노산 폴리펩티드의 어느 한 말단(또는 폴리펩티드 내)에서 5개 아미노산 결실을 갖는 변이체는 "참조" 폴리펩티드에 비해 99%(495개/500개의 동일한 잔기 $\times$ 100)의 서열 동일성 백분율을 가질 것이다. 이러한 변이체는 폴리펩티드와 "적어도 99%의 서열 동일성"을 갖는 변이체에 포괄될 것이다.

- [0115] 특정 구현예에서, 변이체 프로-영역 서열은 본 발명의 기준(모체) 프로-영역 서열과 적어도 약 40% 내지 약 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 심지어 적어도 99%의 아미노산 서열 동일성을 갖는다. 특정 구현예에서, 프로-영역 서열은 서열번호 15, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 31, 서열번호 32, 및/또는 서열번호 33에 대한 상동성을 포함하는 아미노산 서열로부터 유래된다. 따라서, 특정 양태에서, 본원에 기재된 하나 이상의 프로-영역 변이체의 아미노산 변형은 서열번호 15, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 31, 서열번호 32, 및/또는 서열번호 33의 프로-영역 아미노산 서열을 참조하여 넘버링된다.
- [0116] 특정 양태에서, 신규한(변이체) 프로-영역 서열은 서열번호 15의 아미노산 서열과 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체(기준) 프로-영역 서열로부터 유래된다. 특정 구현예에서, 변이체 프로-영역 서열은 서열번호 15 아미노산(위치) 넘버링에 따라, 위치 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로 이루어진 군으로부터 선택된 하나 이상의 아미노산 잔기 위치 및 아미노산(잔기) 위치 30에 돌연변이를 포함한다.
- [0117] 다른 구현예에서, 본 발명의 변이체 프로-영역 서열은 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 위치 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하며, 아미노산 위치는 서열번호 14에 따라 넘버링된다. 특정 구현예에서, 신규한(변이체) 프로-영역 서열은 서열번호 14의 아미노산 서열에 대해 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 포함하는 기준 프로-영역 서열로부터 유래된다.
- [0118] 다른 구현예에서, 본 발명의 변이체 프로-영역 서열은 위치 2에 글리신(G), 위치 3에 라이신(K), 및 위치 4에 세린(S)의 아미노산 삽입을 포함하며, 아미노산 위치는 서열번호 29에 따라 넘버링된다. 특정 구현예에서, 신규한(변이체) 프로-영역 서열은 서열번호 29의 아미노산 서열에 대해 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 포함하는 기준 프로-영역 서열로부터 유래된다.
- [0119] 다른 구현예에서, 본 발명의 변이체 프로-영역 서열은 위치 2에 글리신(G), 위치 3에 라이신(K), 및 위치 4에 알라닌(A)의 아미노산 삽입을 포함하며, 아미노산 위치는 서열번호 30에 따라 넘버링된다. 특정 구현예에서, 변이체 프로-영역 서열은 서열번호 30의 아미노산 서열에 대해 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 포함하는 기준 프로-영역 서열로부터 유래된다.
- [0120] 다른 구현예에서, 본 발명의 변이체 프로-영역 서열은 위치 2에 글리신(G), 위치 3에 라이신(K), 위치 4에 알라닌(A), 및 위치 5에 알라닌(A)의 아미노산 삽입을 포함하며, 아미노산 위치는 서열번호 31에 따라 넘버링된다. 특정 구현예에서, 변이체 프로-영역 서열은 서열번호 31의 아미노산 서열에 대해 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 포함하는 기준 프로-영역 서열로부터 유래된다.
- [0121] 다른 특정 양태에서, 변이체 프로-영역 서열은 서열번호 11의 아미노산 서열과 적어도 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체(기준) 프로-영역 서열로부터 유래된다(G30; 프로-영역 B 서열).
- [0122] 특정 구현예에서, 변이체 프로-영역 서열은 위치 68, 위치 72, 및 위치 80으로부터 선택된 위치에 적어도 하나의 아미노산 치환을 포함하며, 아미노산 위치는 서열번호 9, 서열번호 11, 서열번호 15, 서열번호 32, 또는 서열번호 33에 따라 넘버링된다.
- [0123] 다른 특정 구현예에서, 변이체 프로-영역 서열은, 서열번호 9, 서열번호 11, 서열번호 15, 서열번호 32, 또는 서열번호 33 아미노산(위치) 넘버링에 따라, 위치 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로 이루어진 군으로부터 선택된 하나 이상의 아미노산 잔기 위치, 및 아미노산(잔기) 위치 30에 돌연변이를 포함한다.
- [0124] 특정 구현예에서, 변이체 프로-영역 서열은 프로-영역 서열의 순 전하가 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 15, 서열번호 29, 서열번호 30, 서열번호 31, 서열번호 32, 및 서열번호 33의 기준(모체) 서열과

같은 기준(모체) 프로-영역 서열 대비 양 전하(예를 들어 +1 순 전하)가 되도록 설계, 조작, 구성된다. 특정 구현예에서, 본 발명의 변이체 프로-영역 서열은 프로-영역 서열의 순 전하가 적어도 양 전하 1(+1) 내지 약 양 전하 4(+4)가 되도록 설계, 조작, 구성된다. 하나 이상의 다른 구현예에서, 본 발명의 변이체 프로-영역 서열은 표 1, 표 2, 표 3, 표 4, 및/또는 표 5에 제시된 하나 이상의 아미노산 변형을 포함한다.

[0125] III. 재조합 폴리뉴클레오타이드 및 분자생물학

[0126] 일반적으로 전술된 바와 같이, 본 발명의 특정 구현예는 신규한 돌연변이체 프로-영역 서열에 관한 것이다. 관련 양태에서, 본 발명은 하나 이상의 돌연변이체 프로-영역 핵산(DNA) 서열을 포함하는 재조합 폴리뉴클레오타이드를 제공한다. 따라서, 특정 구현예는 재조합 폴리뉴클레오타이드(예를 들어 벡터, 플라스미드, 발현 카세트 등), 관심 단백질을 발현하는 재조합 그람 양성 박테리아 세포/균주 등에 관한 것이다. 특정 양태에서, 본 발명은 관심 단백질의 생산 강화를 위한 재조합 그람 양성 박테리아 세포(균주)에 도입하기에 적합한 폴리뉴클레오타이드 구성체를 제공한다. 특정 양태에서, 본 발명의 폴리뉴클레오타이드 구성체는 발현 카세트에 지칭되며, 이러한 카세트는, 5'에서 3' 방향으로 작동가능한 조합으로, 성숙 관심 단백질(POI)을 암호화하는 다운스트림(3') 유전자 CDS에 연결된 업스트림(5') 프로-영역 DNA 서열을 적어도 포함한다.

[0127] 예를 들어, 본원에 기재된 하나 이상의 핵산 서열은 임의의 적합한 합성, 조작, 및/또는 단리 기술, 또는 이들의 조합을 사용하여 생성될 수 있다. 예를 들어, 본원에 기재된 하나 이상의 폴리뉴클레오타이드는 당업자에게 잘 알려진 고체상 합성 기술과 같은 표준 핵산 합성 기술을 사용하여 생성될 수 있다. 이러한 기술에서, 최대 50개 이상의 뉴클레오타이드 염기의 단편은 일반적으로 합성된 다음, 결합되어(예를 들어 효소적 또는 화학적 결합 방법을 통해) 본질적으로 목적하는 연속 핵산 서열을 형성한다. 본원에 기재된 하나 이상의 폴리뉴클레오타이드의 합성은 또한 고전적 포스포라미다이트 방법(예를 들어 문헌[Beaucage and Caruthers, 1981])을 사용한 화학적 합성 또는 문헌[Matthes *et al.* (1984)]에 기재된 방법(일반적으로 자동화 합성 방법에서 실행됨)을 포함하지만 이에 한정되지 않는 당업계에 알려진 임의의 적합한 방법으로 촉진될 수 있다. 본원에 기재된 하나 이상의 폴리뉴클레오타이드는 또한 자동 DNA 합성기를 사용하여 생성될 수 있다. 맞춤형 핵산은 다양한 상업적 공급원(예를 들어 ATUM(DNA 2.0), Newark, CA, USA; Life Tech(GeneArt), Carlsbad, CA, USA; GenScript, Ontario, Canada; Base Clear B. V., Leiden, Netherlands; Integrated DNA Technologies, Skokie, IL, USA; Ginkgo Bioworks(Gen9), Boston, MA, USA; Twist Bioscience, San Francisco, CA, USA)에서 주문할 수 있다. 핵산을 합성하는 다른 기법 및 관련 원리는 당업계에 설명되어 있으며 알려져 있다.

[0128] 핵산 변형에 유용한 재조합 DNA 기법은 당업계에 잘 알려져 있으며, 예를 들어 제한 엔도뉴클레아제 분해, 라이게이션, 역전사 및 cDNA 생산, 및 폴리머라제 연쇄 반응(예를 들어 PCR) 등이 있다. 본원에 기재된 하나 이상의 폴리뉴클레오타이드는 또한 본원에 기재된 하나 이상의 변이체를 암호화하는 폴리뉴클레오타이드에 하이브리드화하거나 PCR 증폭할 수 있는 하나 이상의 올리고뉴클레오타이드 프로브를 사용하여 cDNA 라이브러리를 스크리닝함으로써 수득할 수 있다. cDNA 클론을 스크리닝하고 단리하는 절차 및 PCR 증폭 절차는 당업자에게 잘 알려져 있으며 당업자에게 알려진 표준 참고문헌에 기재되어 있다. 본원에 기재된 하나 이상의 폴리뉴클레오타이드는 예를 들어, 알려진 돌연변이유발 절차(예를 들어 부위 지정 돌연변이유발, 부위 포화 돌연변이유발, 및 시험관 내 재조합)를 통해 자연 발생 폴리뉴클레오타이드 백본(예를 들어 본원에 기재된 하나 이상의 변이체 프로-영역 서열을 암호화하는 것)을 변경하여 수득할 수 있다. 본원에 기재된 하나 이상의 변이체를 암호화하는 본원에 기재된 변형 폴리뉴클레오타이드를 생성하는 데 적합한 다양한 방법(예를 들어, 부위 포화 돌연변이유발, 스캐닝 돌연변이유발, 삽입 돌연변이유발, 결실 돌연변이유발, 무작위 돌연변이유발, 부위 지정 돌연변이유발, 및 지정 진화뿐만 아니라 다양한 다른 재조합적 접근 방식을 포함하지만 이에 한정되지 않음)이 당업계에 알려져 있다.

[0129] 일반적으로 전술되고 하기 실시예에 추가로 기재된 바와 같이, 본 발명의 특정 구현예는 이중성 관심 단백질의 증가된 양을 생산할 수 있는 재조합(변형) 그람 양성 세포에 관한 것이다. 따라서 특정 구현예는 증가된 단백질 생산 능력을 갖는 이러한 재조합 그람 양성 세포를 구성하는 방법에 관한 것이다. 특정 구현예에서, 관심 단백질을 암호화하는 하나 이상의 발현 카세트가 본 발명의 그람 양성 세포에 도입된다. 예시적 구현예에서, 카세트는 세포의 게놈에 통합된다. 따라서, 특정 구현예는 재조합(변형) 그람 양성 숙주 세포를 구성하는 데 사용하기에 적합한 핵산 분자, 폴리뉴클레오타이드(예를 들어 벡터, 플라스미드, 발현 카세트), 조절 요소 등에 관한 것이다.

[0130] 따라서, 실시예에 나타내고 본원에 일반적으로 기재된 바와 같이, 본 발명의 재조합 세포는 당업계에 잘 알려진 표준 및 통상적인 재조합 DNA 및 분자 클로닝 기법을 사용하는 기술 중 하나에 의해 구성될 수 있다. 유전자 변형 방법으로는 (a) 유전자 내의 하나 이상의 뉴클레오타이드의 도입, 치환, 또는 제거, 또는 유전자의 전사 또는

번역에 필요한 조절 요소 내의 하나 이상의 뉴클레오타이드의 도입, 치환, 또는 제거, (b) 유전자 파괴, (c) 유전자 전환, (d) 유전자 결실, (e) 유전자 하향조절, (f) 부위 특이적 돌연변이유발, 및/또는 (g) 무작위 돌연변이유발이 포함되지만, 이에 한정되지 않는다.

[0131] 특정 구현예에서, 본 발명의 변형 세포는 유전자의 발현을 당업계에 잘 알려진 방법(예를 들어, 삽입, 파괴, 대체, 또는 결실)을 사용하여 감소시키거나 제거함으로써 구성될 수 있다. 변형 또는 비활성화 대상인 유전자 부분은 예를 들어 암호화 영역이거나 암호화 영역의 발현에 필요한 조절 요소일 수 있다.

[0132] 이러한 조절 또는 제어 서열의 예는 프로모터 서열 또는 이의 기능적 부분(즉, 핵산 서열의 발현에 영향을 미치기에 충분한 부분)일 수 있다. 변형을 위한 다른 제어 서열은 리더 서열, 프로-펩티드 서열, 신호 서열, 전사 종결자, 전사 활성화자 등을 포함하나 이에 한정되지 않는다.

[0133] 다른 특정 구현예에서, 변형 세포는 유전자의 발현을 제거하거나 감소시키기 위한 유전자 결실에 의해 구성된다. 유전자 결실 기술은 유전자(들)를 부분적으로 또는 완전하게 제거하여, 이들의 발현이 제거되거나 비기능적(또는 활성이 감소된) 단백질 산물이 발현되도록 할 수 있다. 이러한 방법에서, 유전자(들)의 결실은 유전자를 플랭킹하는 5' 및 3' 영역을 인접하여 포함하도록 구성된 플라스미드를 사용하는 상동성 재조합에 의해 달성될 수 있다. 인접한 5' 및 3' 영역은 예를 들어 온도 감수성 플라스미드 상에서, 플라스미드가 세포에서 확립될 수 있게 하는 허용 온도에서 제2 선별 마커와 함께 세포에 도입될 수 있다. 그런 다음, 플라스미드를 상동성 플랭킹 영역 중 하나에서 염색체에 통합시키는 세포를 선택하기 위해 허용되지 않는 온도로 세포를 이동시킨다. 플라스미드의 통합에 대한 선택은 제2 선별 마커에 대한 선택에 의해 수행된다. 통합 후, 선택 없이 여러 세대 동안 허용 온도로 세포를 이동시킴으로써 제2 상동성 플랭킹 영역에서의 재조합 이벤트가 자극된다. 세포를 플레이팅하여 단일 콜로니를 수득하고, 두 선별 마커의 손실에 대해 콜로니를 검사한다. 따라서, 당업자는 전체 또는 부분 결실에 적합한 유전자의 암호화 서열 및/또는 유전자의 비암호화 서열의 뉴클레오타이드 영역을 쉽게 확인할 수 있다.

[0134] 다른 구현예에서, 변형 세포는 유전자 또는 이의 전사 또는 번역에 필요한 조절 요소에서의 하나 이상의 뉴클레오타이드의 도입, 치환, 또는 제거에 의해 구성된다. 예를 들어, 정지 코돈의 도입, 시작 코돈의 제거, 또는 오픈리딩 프레임의 프레임 시프트를 일으키도록 뉴클레오타이드를 삽입하거나 제거할 수 있다. 이러한 변형은 당업계에 알려진 방법에 따라 부위 지정 돌연변이유발 또는 PCR 생성 돌연변이유발에 의해 달성될 수 있다. 따라서, 특정 구현예에서, 본 발명의 유전자는 완전 또는 부분 결실에 의해 비활성화된다.

[0135] 다른 구현예에서, 변형된 세포는 유전자 전환의 프로세스에 의해 구성된다. 예를 들어, 유전자 전환 방법에서, 유전자(들)에 상응하는 핵산 서열을 시험관내 돌연변이화하여 결함 있는 핵산 서열을 생성한 후, 이를 모체 세포로 형질전환시켜 결함 있는 유전자를 생성한다. 상동성 재조합에 의해, 결함 있는 핵산 서열이 내인성 유전자를 대체한다. 결함 유전자 또는 유전자 단편이 결함 유전자를 함유하는 형질전환체의 선택에 사용될 수 있는 마커를 또한 암호화하는 것이 바람직할 수 있다. 예를 들어, 결함 있는 유전자는 선별 마커와 함께 비복제 또는 온도 감수성 플라스미드 상에 도입될 수 있다. 플라스미드의 통합에 대한 선택은 플라스미드 복제를 허용하지 않는 조건하에 마커에 대한 선택에 의해 수행된다. 유전자 대체로 이어지는 제2 재조합 이벤트에 대한 선택은 선별 마커의 손실 및 돌연변이된 유전자의 획득에 대한 콜로니 검사에 의해 수행된다. 대안적으로, 결함 있는 핵산 서열은 후술하는 바와 같이, 유전자의 하나 이상의 뉴클레오타이드의 삽입, 치환, 또는 결실을 포함할 수 있다.

[0136] 다른 구현예에서, 변형 세포는 유전자의 핵산 서열에 상보적인 뉴클레오타이드 서열을 사용하는 확립된 안티센스 기술에 의해 구성된다. 보다 구체적으로, 그람 양성 세포에 의한 유전자의 발현은 세포에서 전사될 수 있고 세포에서 생성된 mRNA에 혼성화될 수 있는, 유전자의 핵산 서열에 상보적인 뉴클레오타이드 서열을 도입함으로써 감소(하향조절)되거나 제거될 수 있다. 따라서, 상보적인 안티센스 뉴클레오타이드 서열이 mRNA에 혼성화될 수 있는 조건하에서, 번역된 단백질의 양은 감소되거나 제거된다. 이러한 안티센스 방법은 RNA 간섭(RNAi), 소간섭 RNA(siRNA), 마이크로RNA(miRNA), 안티센스 올리고뉴클레오타이드 등을 포함하나 이에 한정되지 않으며, 이들 모두는 당업자에게 잘 알려져 있다.

[0137] 다른 구현예에서, 변형 세포는 CRISPR-Cas9 편집을 통해 생산/구성된다. 예를 들어, 관심 단백질을 암호화하는 유전자는, DNA 상의 표적 서열에 엔도뉴클레아제를 보충하는 가이드 RNA(예를 들어 Cas9) 및 Cpf1 또는 가이드 DNA(예를 들어 NgAgo)에 결합함으로써 표적 DNA를 찾아내는 핵산 가이드된 엔도뉴클레아제에 의해 편집 또는 파괴(또는 결실 또는 하향조절)될 수 있고, 엔도뉴클레아제는 DNA에서 단일가닥 또는 이중가닥 절단을 생성할 수 있다. 이러한 표적화된 DNA 절단은 DNA 복구를 위한 기질이 되며, 제공된 편집 주형과 재조합하여 유전자를 파

괴하거나 결실시킬 수 있다. 예를 들어, 핵산 가이드된 엔도뉴클레아제를 암호화하는 유전자(이 목적의 경우, *S. 피오게네스*로부터의 Cas9) 또는 Cas9 뉴클레아제를 암호화하는 코돈 최적화된 유전자는 그람 양성 세포에서 활성이 있는 프로모터 및 그람 양성 세포에서 활성이 있는 종결자에 작동가능하게 연결됨으로써, 그람 양성 세포 Cas9 발현 카세트를 생성한다. 마찬가지로, 관심 유전자에 대해 고유한 하나 이상의 표적 부위가 당업자에 의해 쉽게 확인된다. 예를 들어, 관심 유전자 내의 표적 부위로 지시된 gRNA를 암호화하는 DNA 구성체를 구성하기 위해, 가변 표적화 도메인(VT)은 프로토-스페이서 인접 모티프(PAM)의 5'인 표적 부위의 뉴클레오티드(TGG)를 포함할 것이며, 이 뉴클레오티드는 *S. 피오게네스* Cas9에 대한 Cas9 엔도뉴클레아제 인식 도메인(CER)을 암호화하는 DNA에 융합된다. VT 도메인을 암호화하는 DNA와 CER 도메인을 암호화하는 DNA의 조합은 이에 따라 gRNA를 암호화하는 DNA를 생성한다. 따라서, gRNA를 암호화하는 DNA를 그람 양성 세포에서 활성이 있는 프로모터 및 그람 양성 세포에서 활성이 있는 종결자에 작동가능하게 연결함으로써 gRNA에 대한 그람 양성 발현 카세트가 생성된다.

[0138] 특정 구현예에서, 엔도뉴클레아제에 의해 유도된 DNA 절단은 유입 서열로 복구/대체된다. 예를 들어, 상기 Cas9 발현 카세트 및 gRNA 발현 카세트에 의해 생성된 DNA 절단을 정확하게 복구하기 위해, 세포의 DNA 복구 기구가 편집 주형을 이용할 수 있도록 뉴클레오티드 편집 주형이 제공된다. 예를 들어, 표적화된 유전자의 5'에 있는 약 500 bp가 표적화된 유전자의 3'에 있는 약 500 bp에 융합되어 편집 주형을 생성할 수 있고, 이 주형은 그람 양성 숙주의 기구에 의해 사용되어 RGEN에 의해 생성된 DNA 절단을 복구한다.

[0139] Cas9 발현 카세트, gRNA 발현 카세트, 및 편집 주형은 다양한 방법(예를 들어, 원형질체 융합, 전기천공, 자연적 능력, 또는 유도된 능력)을 사용하여 사상 진균 세포에 공동으로 전달될 수 있다. 형질전환된 세포는 정방향 및 역방향 프라이머로 유전자좌를 증폭시켜 표적 유전자좌를 PCR 증폭함으로써 스크리닝된다. 이러한 프라이머는 야생형 유전자좌 또는 RGEN에 의해 편집된 변형된 유전자좌를 증폭시킬 수 있다. 이후 시퀀싱 프라이머를 사용하여 이들 단편을 시퀀싱하여, 편집된 콜로니를 식별한다.

[0140] 또 다른 구현예에서, 변형 세포는 당업계에 잘 알려진 방법(화학적 돌연변이유발 및 전위를 포함하나 이에 한정되지 않음)을 사용하여 무작위 또는 특이적 돌연변이유발에 의해 구성된다. 유전자의 변형은 모세포에 대한 돌연변이유발 및 유전자의 발현이 감소 또는 제거된 돌연변이 세포에 대한 스크리닝에 의해 수행될 수 있다. 특이적 또는 무작위적일 수 있는 돌연변이유발은 예를 들어 적합한 물리적 또는 화학적 돌연변이유발제의 사용, 적합한 올리고뉴클레오티드의 사용, 또는 DNA 서열에 대한 PCR 생성 돌연변이유발에 의해 수행될 수 있다. 또한, 돌연변이유발은 이들 돌연변이유발 방법의 임의의 조합을 사용하여 수행될 수 있다.

[0141] 본 목적에 적합한 물리적 또는 화학적 돌연변이유발제의 예는 자외선(UV) 조사, 하이드록실아민, N-메틸-N'-니트로-N-니트로소구아니딘(MNNG), N-메틸-N'-니트로소구아니딘(NTG), O-메틸 하이드록실아민, 아질산, 에틸 메탄설포네이트(EMS), 아황산수소나트륨, 포름산, 및 뉴클레오티드 유사체를 포함한다. 이러한 제제가 사용되는 경우, 돌연변이유발은 일반적으로, 돌연변이 대상 모세포를 적합한 조건에서 선택 돌연변이유발제의 존재하에 인큐베이션하고, 유전자의 발현이 감소되거나 나타나지 않는 돌연변이 세포를 선택함으로써 수행된다.

[0142] PCT 공개 W02003/083125호는 *E. 콜라이*를 우회하는 PCR 융합을 사용한 바실러스 결실 균주 및 DNA 구성체의 생성과 같은 그람 양성 (*바실러스*) 세포의 변형 방법을 개시한다. PCT 공개 W02002/14490호는 (1) 통합 플라스미드(pComK)의 구성 및 형질전환, (2) 암호화 서열, 신호 서열, 및 프로펩티드 서열의 무작위 돌연변이유발, (3) 상동성 재조합, (4) 형질전환 DNA에 대한 비상동성 플랭크의 추가에 의한 형질전환 효율 증대, (5) 이중 교차(cross-over) 통합의 최적화, (6) 부위 지정 돌연변이유발, 및 (7) 마커가 없는(marker-less) 결실을 포함하는, *바실러스* 세포의 변형 방법을 개시한다.

[0143] 당업자는 폴리뉴클레오티드 서열을 박테리아 세포(예를 들어, 그람 음성 세포 및 그람 양성 세포)에 도입하기 위한 적합한 방법을 잘 알고 있다. 실제로, 원형질체 형질전환과 응축(congression)을 포함하는 형질전환, 형질도입, 및 원형질체 융합과 같은 방법이 알려져 있으며 본 발명에서 사용하기에 적합하다. 본 발명의 DNA 구성체를 숙주 세포에 도입하는 데 형질전환 방법이 특히 바람직하다.

[0144] 일반적으로 사용되는 방법에 추가적으로, 일부 구현예에서, 숙주 세포는 직접 형질전환된다(즉, 숙주 세포에 도입되기 전에 DNA 구성체의 증폭 또는 다른 처리를 위해 중간 세포가 사용되지 않음). DNA 구성체를 숙주 세포에 도입하는 것은 DNA를 플라스미드 또는 벡터 내 삽입 없이 숙주 세포에 도입하기 위한, 당업계에 알려진 물리적 및 화학적 방법을 포함한다. 이러한 방법은 염화칼슘 침전, 전기천공, 네이키드 DNA, 리포솜 등을 포함하나 이에 한정되지 않는다. 추가 구현예에서, DNA 구성체는 플라스미드 내 삽입 없이 플라스미드와 공동으로 형질전환된다. 다른 구현예에서, 변형된 *바실러스* 균주로부터 선별 마커가 당업계에 알려진 방법에 의해 결실되거나 실

질적으로 절단된다. 일부 구현예에서, 숙주 염색체로부터 벡터를 분해하면 고유 염색체 영역이 제거되면서 염색체에 플랜킹 영역을 남긴다.

[0145] 그람 양성 세포에서 유전자, 암호화 서열(CDS), 오픈 리딩 프레임(ORF), 및/또는 이의 변이체 서열의 발현에 사용하기 위한 프로모터 및 프로모터 서열 영역은 일반적으로 당업자에게 알려져 있다. 본 발명의 프로모터 서열은 그람 양성 세포에서 이들이 기능할 수 있도록 일반적으로 선택된다. 예를 들어, *바실러스* 세포에서 유전자 발현을 유도하는 데 유용한 프로모터는 *B. 서브틸리스* 알칼리성 프로테아제(*aprE*) 프로모터, *B. 서브틸리스*의 α -아밀라제 프로모터(*amyE*), *B. 리체니포르미스*의 α -아밀라제 프로모터(*amyL*), *B. 아밀로리케파시엔스*의 α -아밀라제 프로모터, *B. 서브틸리스*로부터의 중성 프로테아제(*nprE*) 프로모터, 돌연변이 *aprE* 프로모터, 또는 *B. 리체니포르미스* 또는 기타 관련 *바실러스*로부터의 임의의 다른 프로모터를 포함하나 이에 한정되지 않는다. *바실러스* 세포에서 다양한 활성(프로모터 강도)을 갖는 프로모터 라이브러리를 스크리닝하고 생성하는 방법은 공개 WO2002/14490호에 기재되어 있다.

[0146] IV. 단백질의 생산을 위한 그람 양성 세포의 발효

[0147] 일반적으로 전술한 바와 같이, 특정 구현예는 단백질 생산 표현형이 증가된 그람 양성 세포를 구성하고 수득하기 위한 조성물 및 방법에 관한 것이다. 따라서, 특정 구현예는 적합한 배지에서 세포를 발효시켜 그람 양성 세포에서 관심 단백질을 생산하는 방법에 관한 것이다. 당업계에 잘 알려진 발효 방법을 적용하여 본 발명의 그람 양성 세포를 발효시킬 수 있다.

[0148] 일부 구현예에서, 세포는 회분식 또는 연속식 발효 조건하에 배양된다. 전형적인 회분식 발효는 배지의 조성이 발효 시작시 설정되어 발효 중에 변경되지 않는 폐쇄 시스템이다. 발효 시작시, 배지에 목적하는 유기체(들)를 접종한다. 이 방법에서는, 시스템에 어떤 구성요소도 첨가하지 않고 발효가 일어날 수 있다. 일반적으로, 회분식 발효는 탄소원의 추가와 관련하여 "회분식"으로 간주되며, pH 및 산소 농도와 같은 인자를 제어하려는 시도가 종종 이루어진다. 회분식 시스템의 대사산물 및 바이오매스의 조성은 발효가 중단되는 시점까지 지속적으로 변화한다. 일반적인 회분식 배양물 내에서, 세포는 정적 지체기를 거쳐 고성장 대수기로 진행하고, 최종적으로 성장 속도가 감소하거나 성장이 멈추는 정지기로 진행할 수 있다. 정지기의 세포는 처치되지 않으면 결국 사멸한다. 일반적으로, 대수기의 세포가 대부분의 제품 생산을 담당한다.

[0149] 표준 회분식 시스템에 대한 적절한 변형은 "유가식(fed-batch)" 발효 시스템이다. 일반적인 회분식 시스템의 이러한 변형에서는, 발효가 진행됨에 따라 기질이 증분으로 추가된다. 유가식 시스템은 이화대사물 역제가 세포의 대사를 억제할 가능성이 있는 경우 및 배지에 제한된 양의 기질이 있는 것이 바람직한 경우에 유용하다. 유가식 시스템에서 실제 기질 농도는 측정이 어렵기 때문에, pH, 용존 산소량, 및 CO₂와 같은 폐가스의 분압과 같은 측정가능한 인자의 변화에 기초하여 추정된다. 회분식 및 유가식 발효는 일반적이며 당업계에 알려져 있다.

[0150] 연속식 발효는 규정 발효 배지가 생물 반응기에 연속적으로 첨가되고 동량의 조정된(conditioned) 배지가 프로세싱을 위해 동시에 제거되는 개방 시스템이다. 연속식 발효는 일반적으로, 세포가 주로 대수기 성장 중인 배양물을 일정한 고밀도로 유지한다. 연속식 발효는 세포 성장 및/또는 생성물 농도에 영향을 미치는 하나 이상의 인자의 조절이 가능하다. 예를 들어, 일 구현예에서, 탄소원 또는 질소원과 같은 제한 영양소가 고정 비율로 유지되고 다른 모든 파라미터는 조절될 수 있다. 다른 시스템에서, 배지 탁도에 의해 측정되는 세포 농도가 일정하게 유지되는 반면, 성장에 영향을 미치는 여러 인자는 계속 변경될 수 있다. 연속식 시스템은 정상 상태 성장 조건을 유지하려고 한다. 따라서, 배지의 배출로 인한 세포 손실은 발효에서의 세포 성장 속도와 균형을 이루어야 한다. 연속식 발효 공정을 위한 영양소 및 성장 인자를 조절하는 방법뿐만 아니라, 생성물 형성 속도를 최대화하는 기술은 산업 미생물학 분야에 잘 알려져 있다.

[0151] 특정 구현예에서, 본 발명의 그람 양성 세포에 의해 발현/생산되는 관심 단백질은 원심분리 또는 여과에 의해 배지로부터 숙주 세포를 분리하는 것, 또는 필요한 경우 세포를 파괴하는 것, 및 세포 분획 및 잔사로부터 상청액을 제거하는 것을 포함하는 통상적인 절차에 의해 배양 배지로부터 회수할 수 있다. 일반적으로, 청정 후, 상청액 또는 여과액의 단백질 성분은 염(예를 들어 황산암모늄)에 의해 침전된다. 침전된 단백질은 이후 가용화되고 다양한 크로마토그래피 절차, 예를 들어 이온 교환 크로마토그래피, 겔 여과에 의해 정제될 수 있다.

[0152] 일부 구현예에서, 세포는 회분식 또는 연속식 발효 조건하에 배양된다. 전형적인 회분식 발효는 배지의 조성이 발효 시작시 설정되어 발효 중에 변경되지 않는 폐쇄 시스템이다. 발효 시작시, 배지에 목적하는 유기체(들)를 접종한다. 이 방법에서는, 시스템에 어떤 구성요소도 첨가하지 않고 발효가 일어날 수 있다. 일반적으로, 회분식 발효는 탄소원의 추가와 관련하여 "회분식"으로 간주되며, pH 및 산소 농도와 같은 인자를 제어하려는 시도

가 종종 이루어진다. 회분식 시스템의 대사산물 및 바이오매스의 조성은 발효가 중단되는 시점까지 지속적으로 변화한다. 일반적인 회분식 배양물 내에서, 세포는 정적 지체기를 거쳐 고성장 대수기로 진행하고, 최종적으로 성장 속도가 감소하거나 성장이 멈추는 정지기로 진행할 수 있다. 정지기의 세포는 처치되지 않으면 결국 사멸한다. 일반적으로, 대수기의 세포가 대부분의 제품 생산을 담당한다.

[0153] 표준 회분식 시스템에 대한 적절한 변형은 "유가식(fed-batch)" 발효 시스템이다. 일반적인 회분식 시스템의 이러한 변형에서는, 발효가 진행됨에 따라 기질이 증분으로 추가된다. 유가식 시스템은 이화대사물 억제가 세포의 대사를 억제할 가능성이 있는 경우 및 배지에 제한된 양의 기질이 있는 것이 바람직한 경우에 유용하다. 유가식 시스템에서 실제 기질 농도는 측정이 어렵기 때문에, pH, 용존 산소량, 및 CO₂와 같은 폐가스의 분압과 같은 측정가능한 인자의 변화에 기초하여 추정된다. 회분식 및 유가식 발효는 일반적이며 당업계에 알려져 있다.

[0154] 연속식 발효는 규정 발효 배지가 생물 반응기에 연속적으로 첨가되고 동량의 조정된(conditioned) 배지가 프로세싱을 위해 동시에 제거되는 개방 시스템이다. 연속식 발효는 일반적으로, 세포가 주로 대수기 성장 중인 배양물을 일정한 고밀도로 유지한다. 연속식 발효는 세포 성장 및/또는 생성물 농도에 영향을 미치는 하나 이상의 인자의 조절이 가능하다. 예를 들어, 일 구현예에서, 탄소원 또는 질소원과 같은 제한 영양소가 고정 비율로 유지되고 다른 모든 파라미터는 조절될 수 있다. 다른 시스템에서, 배지 탁도에 의해 측정되는 세포 농도가 일정하게 유지되는 반면, 성장에 영향을 미치는 여러 인자는 계속 변경될 수 있다. 연속식 시스템은 정상 상태 성장 조건을 유지하려고 한다. 따라서, 배지의 배출로 인한 세포 손실은 발효에서의 세포 성장 속도와 균형을 이루어야 한다. 연속식 발효 공정을 위한 영양소 및 성장 인자를 조절하는 방법뿐만 아니라, 생성물 형성 속도를 최대화하는 기술은 산업 미생물학 분야에 잘 알려져 있다.

[0155] 특정 구현예에서, 본 발명의 그람 양성 세포에 의해 발현/생산되는 관심 단백질은 원심분리 또는 여과에 의해 배지로부터 숙주 세포를 분리하는 것, 또는 필요한 경우 세포를 파괴하는 것, 및 세포 분획 및 잔사로부터 상청액을 제거하는 것을 포함하는 통상적인 절차에 의해 배양 배지로부터 회수할 수 있다. 일반적으로, 청정 후, 상청액 또는 여과액의 단백질 성분은 염(예를 들어 황산암모늄)에 의해 침전된다. 침전된 단백질은 이후 가용화되고 다양한 크로마토그래피 절차, 예를 들어 이온 교환 크로마토그래피, 겔 여과에 의해 정제될 수 있다.

[0156] V. 관심 단백질

[0157] 본 발명의 관심 단백질(POI)은 임의의 내인성 또는 이종성 단백질일 수 있으며, 이는 이러한 POI의 변이체일 수 있다. 단백질은 하나 이상의 이황화 브리지를 함유할 수 있거나, 기능적 형태가 단량체 또는 다량체인 단백질이다. 즉, 단백질은 4차 구조를 가지며, 복수의 동일한(상동성) 또는 동일하지 않은(이종성) 서브유닛으로 구성되고, POI 또는 이의 변이체 POI는 바람직하게는 관심 특성을 갖는 것이다.

[0158] 예를 들어, 특정 구현예에서, 본 발명의 변형된 그람 양성 세포는 이의 비변형(기준 또는 대조군) 세포에 비해 적어도 약 0.1%, 적어도 약 0.5%, 적어도 약 1%, 적어도 약 5%, 적어도 약 6%, 적어도 약 7%, 적어도 약 8%, 적어도 약 9%, 또는 적어도 약 10% 이상 증가된 POI를 생산한다.

[0159] 특정 구현예에서, 본 발명의 변형 그람 양성 세포는 대조군 세포에 비해 증가된 POI 비생산성(Qp)을 나타낸다. 예를 들어, 비생산성(Qp)의 검출은 단백질 생산을 평가하는 데 적합한 방법이다. 비생산성(Qp)은 다음 식을 사용하여 결정될 수 있다:

$$[0160] \text{"Qp"} = \text{gP/gDCW} \cdot \text{hr}''$$

[0161] 여기서, "gP"는 탱크에서 생산된 단백질의 그램이고, "gDCW"는 탱크 내 건조 세포 중량(DCW)의 그램이고, "hr"은 집중 시점으로부터의 발효 시간(h)으로서, 생산 시간과 성장 시간이 포함된다.

[0162] 따라서, 다른 특정 구현예에서, 본 발명의 변형 그람 양성 세포는 비변형 (모) 세포에 비해 적어도 약 0.1%, 적어도 약 1%, 적어도 약 5%, 적어도 약 6%, 적어도 약 7%, 적어도 약 8%, 적어도 약 9%, 또는 적어도 약 10% 이상의 비생산성(Qp) 증가를 포함한다.

[0163] 특정 구현예에서, POI 또는 이의 변이체 POI는 아세틸 에스테라제, 아미노펩티다제, 아밀라제, 아라비나제, 아라비노푸라노시다제, 탄산무수화효소, 카복시펩티다제, 카탈라제, 셀룰라제, 키티나제, 키모신, 큐티나제, 테옥시리보뉴클레아제, 에피머라제, 에스테라제, α-갈락토시다제, β-갈락토시다제, α-글루카나제, 글루칸 라이아제, 엔도-β-글루카나제, 글루코아밀라제, 글루코스 옥시다제, α-글루코시다제, β-글루코시다제, 글루루로니다제, 글리코실 하이드롤라제, 헤미셀룰라제, 헥소스 옥시다제, 하이드롤라제, 인버타제, 이소머라제, 라카제, 리가제, 리파제, 리아제, 만노시다제, 옥시다제, 옥시도리덕타제, 펙테이트 리아제, 펙틴 아세틸 에스테라제,

펙틴 해중합효소, 펙틴 메틸 에스테라제, 펙틴분해 효소, 퍼하이드롤라제, 폴리올 옥시다제, 퍼옥시다제, 페놀 옥시다제, 피타제, 폴리갈락투로나제, 프로테아제, 펩티다제, 람노-갈락투로나제, 리보뉴클레아제, 트랜스퍼라제, 수송 단백질, 트랜스글루타미나제, 자일라나제, 헥소스 옥시다제, 및 이들의 조합으로 이루어진 군으로부터 선택된다.

[0164] 따라서, 특정 구현예에서, POI 또는 이의 변이체 POI는 효소 위원회(EC) 번호 EC 1, EC 2, EC 3, EC 4, EC 5, 또는 EC 6으로부터 선택된 효소이다.

[0165] 세포내 및 세포외 발현 단백질의 활성을 검출하고 측정하기 위한 다양한 분석법이 당업자에 알려져 있다.

[0166] VI. 예시적 구현예

[0167] 본원에 개시된 조성물 및 방법의 비제한적인 구현예는 다음과 같다:

[0168] 1. 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링되는, 변이체 프로-영역 서열.

[0169] 2. 구현예 1에 있어서, 서열번호 15의 위치 1~84와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

[0170] 3. 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하는 변이체 프로-영역 서열로서, 상기 변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링되는, 변이체 프로-영역 서열.

[0171] 4. 구현예 3에 있어서, 서열번호 14의 위치 1~86과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

[0172] 5. 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 위치 4에 알라닌(A)의 아미노산 삽입을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 30에 따라 넘버링되는, 변이체 프로-영역 서열.

[0173] 6. 구현예 5에 있어서, 서열번호 30의 위치 1~87과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

[0174] 7. 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 위치 4에 세린(S)의 아미노산 삽입을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 29에 따라 넘버링되는, 변이체 프로-영역 서열.

[0175] 8. 구현예 7에 있어서, 서열번호 29의 위치 1~87과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

[0176] 9. 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 위치 4에 알라닌(A)의 아미노산 삽입, 및 위치 5에 알라닌(A)의 아미노산 삽입을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 31에 따라 넘버링되는, 변이체 프로-영역 서열.

[0177] 10. 구현예 9에 있어서, 서열번호 31의 위치 1~88과 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

[0178] 11. 위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 73에 이소류신(I)에서 발린(V)으로의 치환(I72V)을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 32에 따라 넘버링되는, 변이체 프로-영역 서열.

[0179] 12. 구현예 11에 있어서, 서열번호 15 또는 서열번호 32의 위치 1~84와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖

는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.

- [0180] 13. 위치 30에 글루탐산(E)에서 글리신(G)으로의 치환(E30G), 위치 68에 류신(L)에서 라이신(K)으로의 치환(L68K), 및 위치 80에 글루탐산(E)에서 이소류신(I)으로의 치환(E80I)을 포함하는 변이체 프로-영역 서열로서, 변이체 프로-영역의 아미노산 위치는 서열번호 33에 따라 넘버링되는, 변이체 프로-영역 서열.
- [0181] 14. 구현예 13에 있어서, 서열번호 15 또는 서열번호 33의 위치 1~84와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.
- [0182] 15. 구현예 1 내지 14 중 어느 하나의 변이체 프로-영역을 암호화하는 핵산.
- [0183] 16. 구현예 15의 변이체 프로-영역 핵산을 포함하는 폴리뉴클레오티드.
- [0184] 17. 관심 단백질(POI)을 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된 구현예 1 내지 14 중 어느 하나의 변이체 프로-영역을 암호화하는 업스트림(5') 핵산을 포함하는 폴리뉴클레오티드.
- [0185] 18. 관심 단백질(POI)을 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 구현예 1 내지 14 중 어느 하나의 변이체 프로-영역을 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된 단백질 신호(분비) 서열을 암호화하는 업스트림(5') 핵산을 포함하는 폴리뉴클레오티드.
- [0186] 19. 구현예 16 내지 18 중 어느 하나의 다운스트림(3') 폴리뉴클레오티드에 작동가능하게 연결된 업스트림(5') 프로모터 영역 서열을 포함하는 발현 카세트.
- [0187] 20. 구현예 16 내지 18 중 어느 하나의 도입된 폴리뉴클레오티드 또는 구현예 19의 도입된 카세트를 포함하는 그람 양성 숙주 세포.
- [0188] 21. 구현예 1에 있어서, 하나 이상의 아미노산 치환은 서열번호 15의 기준 프로-영역에 대비 변이체 프로-영역 서열의 순 전하를 증가시키는, 변이체 프로-영역.
- [0189] 22. 구현예 3에 있어서, 하나 이상의 아미노산 치환은 서열번호 14의 기준 프로-영역에 대비 변이체 프로-영역 서열의 순 전하를 증가시키는, 변이체 프로-영역.
- [0190] 23. 구현예 21 또는 구현예 22에 있어서, 순 전하의 증가는 최소한 1(+1) 내지 4(+4)인, 변이체 프로-영역.
- [0191] 24. 구현예 1에 있어서, 표 1에 제시된 아미노산 변형을 포함하는 변이체 프로-영역.
- [0192] 25. 구현예 1에 있어서, 위치 30의 아미노산 치환은 글리신(G), 히스티딘(H), 또는 아스파라진(N)인, 변이체 프로-영역.
- [0193] 26. 구현예 1에 있어서, 위치 1의 아미노산 치환은 트레오닌(T)인, 변이체 프로-영역.
- [0194] 27. 구현예 1에 있어서, 위치 2의 아미노산 치환은 알라닌(A) 또는 프롤린(P)인, 변이체 프로-영역.
- [0195] 28. 구현예 1에 있어서, 위치 3의 아미노산 치환은 알라닌(A), 시스틴(C), 글리신(G), 아르기닌(R), 세린(S), 또는 발린(V)인, 변이체 프로-영역.
- [0196] 29. 구현예 1에 있어서, 위치 4의 아미노산 치환은 글루타민(Q)인, 변이체 프로-영역.
- [0197] 30. 구현예 1에 있어서, 위치 6의 아미노산 치환은 아르기닌(R)인, 변이체 프로-영역.
- [0198] 31. 구현예 1에 있어서, 위치 14의 아미노산 치환은 트레오닌(T)인, 변이체 프로-영역.
- [0199] 32. 구현예 1에 있어서, 위치 19의 아미노산 치환은 아르기닌(R)인, 변이체 프로-영역.
- [0200] 33. 구현예 1에 있어서, 위치 20의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0201] 34. 구현예 1에 있어서, 위치 23의 아미노산 치환은 라이신(K), 아스파라진(N), 또는 글루타민(Q)인, 변이체 프로-영역.
- [0202] 35. 구현예 1에 있어서, 위치 36의 아미노산 치환은 글리신(G)인, 변이체 프로-영역.
- [0203] 36. 구현예 1에 있어서, 위치 37의 아미노산 치환은 트레오닌(T)인, 변이체 프로-영역.
- [0204] 37. 구현예 1에 있어서, 위치 38의 아미노산 치환은 히스티딘(H)인, 변이체 프로-영역.

- [0205] 38. 구현예 1에 있어서, 위치 39의 아미노산 치환은 라이신(K) 또는 글루타민(Q)인, 변이체 프로-영역.
- [0206] 39. 구현예 1에 있어서, 위치 44의 아미노산 치환은 이소류신(I)인, 변이체 프로-영역.
- [0207] 40. 구현예 1에 있어서, 위치 49의 아미노산 치환은 트레오닌(T) 또는 글루타민(Q)인, 변이체 프로-영역.
- [0208] 41. 구현예 1에 있어서, 위치 50의 아미노산 치환은 아르기닌(R)인, 변이체 프로-영역.
- [0209] 42. 구현예 1에 있어서, 위치 64의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0210] 43. 구현예 1에 있어서, 위치 65의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0211] 44. 구현예 1에 있어서, 위치 67의 아미노산 치환은 세린(S)인, 변이체 프로-영역.
- [0212] 45. 구현예 1에 있어서, 위치 68의 아미노산 치환은 라이신(K) 또는 메티오닌(M)인, 변이체 프로-영역.
- [0213] 46. 구현예 1에 있어서, 위치 79의 아미노산 치환은 글루타민(Q)인, 변이체 프로-영역.
- [0214] 47. 구현예 1에 있어서, 위치 83의 아미노산 치환은 이소류신(I)인, 변이체 프로-영역.
- [0215] 48. 구현예 1에 있어서, 위치 84의 아미노산 치환은 류신(L)인, 변이체 프로-영역.
- [0216] 49. 구현예 3에 있어서, 표 2, 표 3, 표 4, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 31, 및 이들의 조합에 제시된 아미노산 변형을 포함하는 변이체 프로-영역.
- [0217] 50. 구현예 3에 있어서, 위치 1의 아미노산 치환은 트레오닌(T)인, 변이체 프로-영역.
- [0218] 51. 구현예 3에 있어서, 위치 32의 아미노산 치환은 글리신(G)인, 변이체 프로-영역.
- [0219] 52. 구현예 3에 있어서, 위치 38의 아미노산 치환은 글리신(G)인, 변이체 프로-영역.
- [0220] 53. 구현예 3에 있어서, 위치 46의 아미노산 치환은 이소류신(I)인, 변이체 프로-영역.
- [0221] 54. 구현예 3에 있어서, 위치 66의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0222] 55. 구현예 3에 있어서, 위치 67의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0223] 56. 구현예 3에 있어서, 위치 70의 아미노산 치환은 라이신(K)인, 변이체 프로-영역.
- [0224] 57. 구현예 3에 있어서, 위치 73의 아미노산 치환은 아르기닌(R)인, 변이체 프로-영역.
- [0225] 58. 표 1~5, 도 1, 도 2, 도 3, 서열번호 9, 서열번호 11, 서열번호 14, 서열번호 29, 서열번호 30, 서열번호 32, 서열번호 33, 및 이들의 조합 중 어느 하나에 제시된 아미노산 변형을 포함하는 변이체 프로-영역.
- [0226] 59. 구현예 58에 있어서, 서열번호 15의 위치 1~84와 적어도 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%의 아미노산 서열 동일성을 갖는 모체 또는 기준 폴리펩티드로부터 유래된 변이체 프로-영역.
- [0227] 60. 그람 양성 박테리아 세포에서 이중성 관심 단백질(POI)을 생산하는 방법으로서, (a) POI를 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된, 1, 2, 3, 4, 6, 14, 16, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 79, 83, 및 84로부터 선택된 하나 이상의 위치 및 위치 30에 아미노산 치환을 포함하는 변이체 프로-영역 서열을 암호화하는 다운스트림(3') 핵산에 작동가능하게 연결된 업스트림(5') 프로모터를 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(변이체 프로-영역의 아미노산 위치는 서열번호 15에 따라 넘버링됨), 및 (b) 변형된 세포를 POI의 생산에 적합한 조건하에 성장/배양/발효시키는 단계를 포함하는 방법.
- [0228] 61. 그람 양성 박테리아 세포에서 이중성 관심 단백질(POI)을 생산하는 방법으로서, (a) POI를 암호화하는 다운스트림(3') 핵산 서열에 작동가능하게 연결된, 위치 2에 글리신(G)의 아미노산 삽입, 위치 3에 라이신(K)의 아미노산 삽입, 및 1, 32, 38, 46, 66, 67, 70, 및 73으로부터 선택된 하나 이상의 위치에 아미노산 치환을 포함하는 변이체 프로-영역 서열을 암호화하는 다운스트림(3') 핵산에 작동가능하게 연결된 업스트림(5') 프로모터 서열을 포함하는 발현 카세트를 그람 양성 세포에 도입하는 단계(변이체 프로-영역의 아미노산 위치는 서열번호 14에 따라 넘버링됨), 및 (b) 변형된 세포를 POI의 생산에 적합한 조건하에 성장/배양/발효시키는 단계를 포함하는 방법.
- [0229] 62. 그람 양성 박테리아 세포에서 이중성 관심 단백질(POI)을 생산하는 방법으로서, (a) 구현예 19의 발현

카세트를 그람 양성 세포에 도입하는 단계, 및 (b) 변형된 세포를 POI의 생산에 적합한 조건하에 성장/배양/발효시키는 단계를 포함하는 방법.

- [0230] 63. 구현예 60 내지 62 중 어느 하나에 있어서, 카세트는, 프로모터와 변이체 프로-영역 서열 사이에 위치하고 작동가능하게 연결된 전단백질 신호(분비) 서열을 암호화하는 핵산을 추가로 포함하는, 방법.
- [0231] 64. 구현예 60 내지 62 중 어느 하나에 있어서, 변형된 세포는, 동일한 POI를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 서열번호 15의 프로-영역 서열을 암호화하는 다운스트림(3') 핵산에 작동가능하게 연결된 동일한 업스트림(5') 프로모터 서열을 포함하는 발현 카세트가 도입된 대조군 그람 양성 세포에 비해 증가된 양의 POI를 생산하는, 방법.
- [0232] 65. 구현예 64에 있어서, POI의 증가된 양은 POI가 대조군 세포에 비해 적어도 약 0.1%, 적어도 약 0.5%, 적어도 약 1%, 적어도 약 5%, 적어도 약 6%, 적어도 약 7%, 적어도 약 8%, 적어도 약 9%, 또는 적어도 약 10% 이상 증가된 것인, 방법.
- [0233] 66. 구현예 60 내지 63 중 어느 하나에 있어서, 변형된 세포의 성능 지수(PI) 값은, 변형된 세포와 동일한 POI를 암호화하는 다운스트림 핵산 서열에 작동가능하게 연결된 서열번호 15의 프로-영역 서열을 암호화하는 다운스트림 핵산에 작동가능하게 연결된 동일한 업스트림 프로모터 서열을 포함하는 발현 카세트가 도입된 대조군 그람 양성 세포 대비 적어도 약 1.1이고, 변형된 세포와 대조군 세포는 동일한 조건하에 배양되는, 방법.
- [0234] 67. 구현예 60 내지 63 중 어느 하나에 있어서, 도입된 카세트는 세포의 게놈에 통합되는, 방법.
- [0235] 68. 구현예 60 내지 63 중 어느 하나에 있어서, 적어도 2개의 도입된 카세트를 포함하는 방법.
- [0236] 69. 구현예 60 내지 63 중 어느 하나에 있어서, POI는 효소, 항체, 수용체, 및 조절 단백질로 이루어진 군으로부터 선택되는, 방법.
- [0237] 70. 구현예 60 내지 63 중 어느 하나에 있어서, POI는 아세틸 에스테라제, 아미노펩티다제, 아밀라제, 아라비나제, 아라비노푸라노시다제, 탄산무수화효소, 카복시펩티다제, 카탈라제, 셀룰라제, 키티나제, 키모신, 큐티나제, 데옥시리보뉴클레아제, 에피머라제, 에스테라제, α -갈락토시다제, β -갈락토시다제, α -글루카나제, 글루칸 라이아제, 엔도- β -글루카나제, 글루코아밀라제, 글루코스 옥시다제, α -글루코시다제, β -글루코시다제, 글루쿠로니다제, 글리코실 하이드롤라제, 헤미셀룰라제, 헥소스 옥시다제, 하이드롤라제, 인버타제, 이소머라제, 라카제, 리가제, 리파제, 리아제, 만노시다제, 옥시다제, 옥시도리덕타제, 펙테이트 리아제, 펙틴 아세틸 에스테라제, 펙틴 해중합효소, 펙틴 메틸 에스테라제, 펙틴분해 효소, 퍼하이드롤라제, 폴리올 옥시다제, 퍼옥시다제, 페놀옥시다제, 피타제, 폴리갈락투로나제, 프로테아제, 펩티다제, 람노-갈락투로나제, 리보뉴클레아제, 트랜스퍼라제, 수송 단백질, 트랜스글루타미나제, 자일라나제, 헥소스 옥시다제, 및 이들의 조합으로 이루어진 군으로부터 선택되는, 방법.
- [0238] 71. 구현예 70에 있어서, 프로테아제는 천연 또는 변이체 서브틸리신인, 방법.
- [0239] 72. 구현예 60 내지 71 중 어느 하나에 있어서, 그람 양성 박테리아 세포는 *바실러스* 종 세포이고, 임의로 *바실러스* 종 세포는 *B. 서브틸리스*, *B. 리체니포르미스*, *B. 렌투스*, *B. 브레비스*, *B. 스테아로써모필루스*, *B. 알칼로필루스*, *B. 아밀로리케파시엔스*, *B. 클라우시*, *B. 할로두란스*, *B. 메가테리움*, *B. 코아굴란스*, *B. 서쿨란스*, *B. 라우투스*, *B. 투링기엔시스*, 및 *지오바실러스 스테아로써모필루스*로 이루어진 군으로부터 선택되는, 방법.
- [0240] **실시예**
- [0241] 본 발명의 특정 양태는 하기 실시예를 고려하여 추가로 이해될 수 있으며, 이는 제한적인 것으로 해석되어서는 안 된다. 재료 및 방법의 변형은 당업자에게 명백할 것이다. 본원에서 사용된 표준 재조합 DNA 및 분자 클로닝 기술은 당업계에 잘 알려져 있다(Ausubel *et al.*, 1987; Sambrook *et al.*, 1989).
- [0242] **실시예 1**
- [0243] **부위 평가 라이브러리에 의한 GG36 PRO 서열의 변형**
- [0244] 본 실시예에서, 변이체 서브틸리신 단백질(BG46_변이체 1; 서열번호 2)을 본원에 기재된 바와 같이 단백질 발현을 모니터링하기 위한 리포터로 사용하였다. 보다 구체적으로, 업스트림(5') *aprE* 유전자 플랜킹 영역을 포함하는 DNA 단편을 *B. 서브틸리스 rrnI*-P2 프로모터/5'-*aprE* UTR 영역 DNA 서열의 변이체를 포함하는 폴리뉴클레오

티드 구성체(예를 들어 발현 카세트)(카나마이신(*kan*) 유전자 발현 카세트(서열번호 7)를 포함하는 다운스트림 (3') *aprE* 유전자 플랜킹 영역 서열에 작동가능하게 연결됨)에 작동가능하게 연결하여, *B. 아밀로리퀘파시엔스* BPN' 종결자 DNA 서열(서열번호 6)에 작동가능하게 연결된 성숙 BG46_변이체 1(서열번호 2)을 암호화하는 DNA 서열에 작동가능하게 연결된 변이체 프로-영역 서열 A(서열번호 9)를 암호화하는 DNA 서열(서열번호 5)에 작동가능하게 연결된 *AprE* 신호 서열을 암호화하는 DNA 서열(서열번호 4)에 작동가능하게 연결된 서열번호 1을 생성하였다. 보다 구체적으로, 이 DNA 단편을 표준 분자 생물학 기법을 사용하여 조립하고, 본원에 기재된 하나 이상의 프로-영역 변형(돌연변이)을 포함하는 선형 DNA 발현 카세트를 개발하기 위한 템플릿으로 사용하였다.

[0245] **프로-영역 서열 A의 부위 평가 라이브러리(SEL)**

[0246] 변이체 프로-영역 서열 A(30H, 서열번호 9)를 암호화하는 아미노산 서열을 부위 평가 라이브러리(SEL)를 생성하기 위한 템플릿으로 사용하고 Twist Bioscience HQ(South San Francisco)에서 4.4kb 단편으로 개발하였다. 발현 카세트의 선형 DNA를 적격한 *B. 서브틸리스* 세포를 형질전환하는 데 사용하고, 형질전환 혼합물을 1.8 ppm 카나마이신을 함유한 LA 플레이트에 플레이팅하여 37°C에서 하룻밤 동안 인큐베이션하였다. 단일 콜로니를 골라 항생제 선택 하에 37°C의 Luria 배양액에서 성장시켰다.

[0247] 96웰 마이크로타이터 플레이트(MTP)의 선별된 고유한 프로-영역 서열 A 변이체를 확인하기 위해 서열 분석을 수행하였다. 예를 들어, 변이체 프로-영역 서열 A(서열번호 9)의 첫 번째(제1) 아미노산 위치는 알라닌("Ala" 또는 "A")이며, 이는 다른 19개의 자연 발생 아미노산 잔기로 변경(치환)될 수 있고, 변이체 프로-영역 서열 A(서열번호 9)의 두 번째(제2) 아미노산 위치는 글루탐산("Glu" 또는 "E")이며, 이는 다른 19개의 자연 발생 아미노산 잔기로 변경(치환)될 수 있으며, 프로-영역 서열 A(서열번호 9)의 모든 84개 위치가 평가될 때까지 계속된다(예를 들어 도 1 참조).

[0248] 리포터 단백질 발현 실험을 위해, 형질전환된 세포를 96웰 MTP에서 배양 배지(MOP 버퍼를 기반으로 한 농축 반정의 배지) 내 32°C, 300 rpm, 습도 80%의 진탕 배양기에서 3일간 성장시킨 후 원심분리 및 여과하였다. 정화된 배양 상층액을 사용하여 리포터 프로테아제 활성을 측정(분석)하여 생산성 수준을 확인하였으며, 샘플은 72시간 후에 채취하였다. 리포터 프로테아제 활성 분석은 하기 실시예 2에서 추가로 기재되어 있다.

[0249] **생산성에 영향을 미치는 위치**

[0250] 아래에 제시된 바와 같이, 표 1은, 리포터 유전자가 네이티브 GG36 프로-영역 서열(서열번호 15)로 발현된 기준 구성체와 비교한 상대적 리포터 생산성의 결과를 나타낸 것으로. 여기서 발현은 아래 실시예 2에 기재된 활성 분석으로 측정하였다. 예를 들어, 표 1에 도시된 바와 같이, 72시간 후에 채취한 샘플에 대한 성능 지수(PI) 값이 제공되고(실시예 2에 기재된 바와 같이 계산된 PI), 여기서 위치 30과 조합된 생산성에 영향을 미치는 아미노산(잔기) 위치는 1, 2, 3, 4, 6, 14, 19, 20, 23, 36, 37, 38, 39, 42, 43, 44, 49, 50, 64, 65, 67, 68, 71, 83, 84이다. 5' 아미노산 변화와 3' 아미노산 변형은 생산성에 영향을 미치며, 선택된 조합의 80%는 추가 양 전하를 가졌다.

표 1

WT PRO 대비 프로-영역 변이체 A 서열 성능 지수

변이체 vs WT	PI	D 전하
없음	1.0	0
E30H	1.5	1
E30G	2.3	1
D5R-E30H	2.5	3
S19R-E30H	2.3	2
E30H-E39Q	2.4	2
E29Q-E30H	2.3	2
E3G-E30H	2.7	2
E30H-L44I	2.2	1
E30H-T83I	2.0	1
E30H-E38K	2.5	2
E23N-E30H	2.1	2
E3A-E30H	2.5	2
E30H-E67S	2.1	2
E30H-E38K	2.7	3
E30H-A79Q	2.2	1
E3V-E30H	2.5	2
E3C-E30H	2.4	2
A1T-E30H	2.5	1
A4Q-E30H	2.6	1
E30H-T50R	2.3	2
E30H-L68K	1.5	2
E30H-E37T	2.4	2
E30H-E36G	2.1	2
E2K-E30H	2.2	3
E14T-E30H	2.1	2
E2P-E30H	2.4	2
E2A-E30H	2.1	2
E3R-E30H	2.7	3
E30H-E49T	2.3	2
E3S-E30H	2.2	2
E30H-E49Q	2.3	2
E30H-M84L	2.2	2
E30H-A65K	2.3	2
E2Q-E30H	2.2	3
E30H-D64K	2.3	3
E30H-L68M	1.5	1
E30H-E43D	2.2	1
E16S-E30H	2.4	2
E30H-I42V	2.4	1
E30H-A71R	2.5	2
A4N-E30H	2.2	1
E30H-D64A	2.0	2
E3P-E30H	2.4	2
D5Q-E30H	2.4	2

[0251]

[0252] 실시예 2

[0253] 프로테아제 활성 분석

[0254] BG46_변이체 1 서브틸리신의 프로테아제 활성을 합성 suc-AAPF-pNA 펩티드 기질의 가수분해를 측정하여 확인하였다. AAPF 분석의 경우, 사용된 시약 용액은 100 mM Tris pH 8.6, 10 mM CaCl₂, 0.005% Tween®-80(Tris/Ca 버퍼), 및 DMSO 중 160 mM suc-AAPF-pNA(suc-AAPF-pNA 스톱 용액; Sigma: S-7388)였다. 작업 용액을 준비하기 위해 1 mL suc-AAPF-pNA 스톱 용액을 100 mL Tris/Ca 버퍼에 첨가하고 혼합하였다. 효소 샘플을 1 mg/mL suc-AAPF-pNA 작업 용액이 함유된 마이크로타이터 플레이트(MTP)에 첨가하고 실온의 동역학 모드에서 SpectraMax 플레이트 리더를 사용하여 3~5분 동안 405 nm에서 활성을 분석하였다. 프로테아제 활성을 mOD/분으로 표시하였다. 구성된 각 변이체의 활성을 측정하여 동일한 플레이트에서 성장시킨 기준 구성체와 비교하였다. 기준 샘플의 값을 변이체 샘플의 값으로 나누어 성능 지수(PI)를 구하였다.

[0255] 실시예 3

[0256] 5' GG36 PRO 서열의 변형

[0257] 본 실시예에서, BG46_변이체 2 서브틸리신(서열번호 8)을 본원에 기재된 바와 같이 발현을 모니터링하기 위한 리포터 단백질로 사용하였다. 보다 구체적으로, 업스트림(5') *aprE* 유전자 플랭킹 영역을 포함하는 DNA 단편을

(5') *B. 서브틸리스* *rrnI*-P2 프로모터/5'-UTR *aprE* 영역 DNA 서열의 변이체를 포함하는 폴리뉴클레오티드 구성체(카나마이신(*kan*) 유전자 발현 카세트(서열번호 7)를 포함하는 다운스트림(3') *aprE* 유전자 플랜킹 영역 서열에 작동가능하게 연결됨)에 작동가능하게 연결하여, *B. 아밀로리케파시엔스* BPN' 종결자 DNA 서열(서열번호 6)에 작동가능하게 연결된 성숙 BG46_변이체 2 서브틸리신(서열번호 8)을 암호화하는 DNA 서열에 작동가능하게 연결된 변이체 프로-영역 서열 B(서열번호 9, 30H)를 암호화하는 DNA 서열(서열번호 10)에 작동가능하게 연결된 *AprE* 신호 펩티드 서열을 암호화하는 DNA 서열(서열번호 4)에 작동가능하게 연결된 서열번호 1을 생성하였다. 보다 구체적으로, 이 DNA 단편을 표준 분자 생물학 기법을 사용하여 조립하고, 본원에 기재된 하나 이상의 프로모터 영역 변형을 포함하는 선형 DNA 발현 카세트를 개발하기 위한 템플릿으로 사용하였다.

[0258] 5' 프로-영역 변형의 도입

[0259] 78개 아미노산 *AprE* 프로-영역 서열(서열번호 12)의 5'(N-말단) 서열을 사용하여 첫 번째 3개 아미노산 잔기를 변이체 프로-영역 서열 C(서열번호 14)의 첫 번째 아미노산 잔기(A)로 교환하여 서열번호 13을 생성하였다. 위의 표 1에 표시된 바와 같이, 프로-영역 변이체 A 서열의 돌연변이체(즉, 표 1, "E030G"; 서열번호 11)는 리포터의 높은 발현을 보였다. 표준 분자 생물학 기법을 사용하여 발현 구성체를 준비하고 기준 프로-영역과 비교하였다. 변이체(돌연변이체)를 4.4 kb 선형 DNA 단편으로서 개발하여 적절한 *B. 서브틸리스* 세포를 형질전환하는 데 사용하고, 형질전환 혼합물을 1.8 ppm 카나마이신을 함유한 LA 플레이트에 플레이팅하여 37°C에서 하룻밤 동안 인큐베이션하였다. 단일 콜로니를 골라 항생제 선택 하에 37°C의 *Luria* 배양액에서 성장시켰다.

[0260] 리포터 단백질 발현 실험을 위해, 형질전환된 세포를 96웰 MTP에서 배양 배지(MOP 버퍼를 기반으로 한 농축 반정의 배지) 내 32°C, 300 rpm, 습도 80%의 진탕 배양기에서 3일간 성장시킨 후 원심분리 및 여과하였다. 정화된 배양 상층액을 사용하여 리포터 프로테아제 활성을 측정(분석)하여 생산성 수준을 확인하였으며, 샘플은 72시간 후에 채취하였다. 리포터 프로테아제 활성 분석을 위의 실시예 2에 기재한 바와 같이 수행하였으며, 변이체 프로-영역 C의 단백질 생산성은 E030G-GG36-Pro(서열번호 11)보다 약 1.1배 높았다.

[0261] 실시예 4

[0262] GG36 PRO 서열의 추가 변형

[0263] 본 실시예에서, BG46_변이체 2 서브틸리신(서열번호 8)을 본원에 기재된 바와 같이 발현을 모니터링하기 위한 리포터 단백질로 사용하였다. 구성은 위의 실시예 3에 제시된 설명에 따라 수행하였다. 보다 구체적으로, 아미노산 글리신(G)과 라이신(K)이 5'(N-말단) 서열에 삽입된("GK") 기준 프로-영역 서열(서열번호 15) 상에서 조합 라이브리리를 준비하였다(예를 들어 도 2, 서열번호 14 참조).

[0264] 변이체 프로-영역 C의 조합 라이브러리

[0265] 프로-영역 돌연변이를 조합하기 위해 표 1 데이터에서 선택하였으며, 표준 분자 생물학 기법을 사용하여 발현 구성체를 준비하고 이를 기준 프로-영역과 비교하였다. 선형 DNA 발현 카세트를 적절한 *B. 서브틸리스* 세포를 형질전환하는 데 사용하고, 형질전환 혼합물을 1.8 ppm 카나마이신을 함유한 LA 플레이트에 플레이팅하여 37°C에서 하룻밤 동안 인큐베이션하였다. 단일 콜로니를 골라 항생제 선택 하에 37°C의 *Luria* 배양액에서 성장시켰다. 리포터 단백질 발현 실험을 위해, 형질전환된 세포를 96웰 MTP에서 배양 배지(MOP 버퍼를 기반으로 한 농축 반정의 배지) 내 32°C, 300 rpm, 습도 80%의 진탕 배양기에서 3일간 성장시킨 후 원심분리 및 여과하였다. 정화된 배양 상층액을 사용하여 리포터 프로테아제 활성을 측정(분석)하여 생산성 수준을 확인하였으며, 샘플은 72시간 후에 채취하였다. 리포터 프로테아제 활성 분석은 위의 실시예 2에 기재한 바와 같이 수행하였다.

[0266] 생산성을 개선하는 프로-영역의 조합

[0267] 아래에 제시된 바와 같이, 표 2는 발현된 리포터(서열번호 8)의 생산성이 증가된 프로-영역 돌연변이 조합의 서열 결과를 나타낸 것으로, 델타(D) 전하는 +1과 +4 사이에서 변동한다. 보다 구체적으로, 루프 영역의 조합으로 서의 양 전하(즉, 잔기 위치 66~73). 72시간 후에 채취한 샘플의 성능 지수(PI) 값은 위의 실시예 2에 기재한 바와 같이 계산된다. 예를 들어, 프로-영역 변이체의 PI 값을 WT 프로-영역 서열(표 2, 서열번호 15)과 비교하고, 여기서 아미노산 위치 넘버링(표 2, 제1 열)은 "GK" 삽입 없이 서열번호 14와 비교된다. 도 3에 도시된 바와 같이, 변이체 프로-영역 서열은 기준 변이체 프로-영역 서열 C(서열번호 14; 86개 아미노산 위치 포함) 또는 야생형(기준) 프로-영역 서열(서열번호 15; 84개 아미노산 위치 포함)의 위치 넘버링에 따라 넘버링될 수 있다.

표 2

GG 삽입 없는 WT PRO(서열번호 15) 대비
프로-영역 변이체 C(서열번호 14) 서열 성능 지수

변이체 vs WT	PI	D 전하
없음	1.0	0
E32G	1.7	1
L46I-D66K-A73R	1.9	3
A67K-L70K-A73R	1.9	3
L70K-A73R	2.0	2
E38G-A67K-L70K	2.0	3
D66K-A67K-L70K	2.0	4
E32G-L70K	2.0	2
A1T-L46I-A73R	2.0	1
L70K-A73R	2.0	2
E38G-L46I-D66K-A67K	2.0	4
E38G-D66K-A67K	2.0	4
L46I-D66K-A67K-L70K	2.0	4
E38G-L46I-A67K	2.0	2
L46I-L70K	2.1	1
L46I-L70K-A73R	2.1	2
E38G-L46I-A67K	2.1	2
L46I-D66K-A67K-L70K	2.2	4
L46I-A67K-L70K	2.3	2

[0268]

[0269]

실시예 5

[0270]

바실러스 리체니포르미스에서 5' GG36 PRO 서열의 변형

[0271]

본 실시예에서, BG46_변이체 2 서브틸리신(서열번호 8)을 본원에 기재된 바와 같이 발현을 모니터링하기 위한 리포터 단백질로 사용하였다. 구성은 일반적으로 다음과 같이 수행되었다. 바실러스 리체니포르미스 *serA* 선별 마커 발현 카세트(서열번호 16)를 포함하는 (5') *serA* 유전자 플랭킹 영역(5' *serA* 유전자 FR)을 포함하는 첫 번째 DNA 단편은 (3') *serA* 유전자 플랭킹 영역(3' *serA* 유전자 FR)(서열번호 20)에 작동가능하게 연결된 *B. 리체니포르미스 amyL* 종결자(서열번호 19)에 작동가능하게 연결된 성숙 BG46_변이체 2 서브틸리신(서열번호 8)을 암호화하는 DNA 서열에 작동가능하게 연결된 변이체 프로-영역 C 서열(서열번호 14)을 암호화하는 DNA에 작동가능하게 연결된 *B. 서브틸리스* AprE 신호 서열(서열번호 4)을 암호화하는 DNA에 작동가능하게 연결된 *B. 서브틸리스 aprE* 5' 비번역 영역(5'-UTR; 서열번호 18)의 DNA 서열에 작동가능하게 연결된 업스트림(5') *B. 서브틸리스 rrnI*-p3 프로모터 영역 DNA 서열(서열번호 17)을 포함하는 폴리뉴클레오티드 구성체에 작동가능하게 연결되었다.

[0272]

B. 리체니포르미스 lysA 선별 마커 발현 카세트(서열번호 21)를 포함하는 (5') *lysA* 유전자 플랭킹 영역(5' *lysA* 유전자 FR)을 포함하는 두 번째 DNA 단편은 (3') *lysA* 유전자 플랭킹 영역(3' *lysA* 유전자 FR, 서열번호 22)에 작동가능하게 연결된 *B. 리체니포르미스 amyL* 종결자(서열번호 10)에 작동가능하게 연결된 성숙 BG46_변이체 2 서브틸리신(서열번호 8)을 암호화하는 DNA 서열에 작동가능하게 연결된 변이체 프로-영역 C 서열(서열번호 14)을 암호화하는 DNA에 작동가능하게 연결된 *B. 서브틸리스* AprE 신호 서열(서열번호 4)을 암호화하는 DNA에 작동가능하게 연결된 *B. 서브틸리스 aprE* 5'-UTR(서열번호 18)의 DNA 서열에 작동가능하게 연결된 업스트림(5') *B. 서브틸리스 rrnI*-p3 프로모터 영역 DNA 서열(서열번호 17)을 포함하는 폴리뉴클레오티드 구성체에 작동가능하게 연결되었다.

[0273]

보다 구체적으로, 이 DNA 단편을 표준 분자 생물학 기법을 사용하여 조립하고, 본원에 기재된 하나 이상의 프로-영역 서열 변형을 포함하는 선형 DNA 발현 카세트를 개발하기 위한 템플릿으로 사용하였다. 예를 들어, 변이체 프로-영역 서열을 포함하는 *B. 리체니포르미스* 균주는 게놈에서 첫 번째 및 두 번째 DNA 단편(위에서 설명)을 통합하여 구성되었으며, 여기서 첫 번째 및 두 번째 단편은 아래 표 3에 제시된 프로-영역 서열 변이체를 포함한다. 표 3에 도시된 바와 같이, 리포터 단백질의 생산은 표준 방법(예를 들어 W02019/055261 참조)을 사용하여 이전에 기재된 바와 같이 확인되었으며 E30G 변이체를 포함하는 대조군 프로-영역 서열로 정규화되었다.

표 3

대조군 프로-영역 서열 B(E32G; 서열번호 11) 대비
B. 리케니포르미스의 프로-영역 변이체 C(GK, 서열번호 14) 서열 성능 지수

변이체 vs WT	PI	D 전하
E32G	1.0	1
E32G-L70K	1.1	2
E32G-A73R	1.3	2
E32G-L70K-A73R	1.2	3

[0274]

[0275]

실시예 6

[0276]

5' GG36 PRO 서열의 추가 변형

[0277]

본 실시예에서, BG46_변이체 2 서브틸리신(서열번호 8)을 본원에 기재된 바와 같이 발현을 모니터링하기 위한 리포터 단백질로 사용하였다. 특히, 86개 아미노산으로 이루어진 돌연변이된 GG36 PRO 서열(서열번호 14; 변이체 프로-영역 서열 C)의 5'(N-말단) 서열을 위치 3의 라이신 뒤에 알라닌(A) 또는 세린(S)과 같은 추가 아미노산을 도입하기 위해 추가로 변형하였다(예를 들어 성숙 프로테아제가 배지로 방출되는 것을 잠재적으로 용이하게 하기 위해). 예를 들어, 이러한 N-말단 변형으로 도 1~3에 제시된 바와 같이 3개의 변이체 프로-영역 서열 D-F(서열번호 29~31)이 생성되었다.

[0278]

이러한 N-말단 변형 GG36 PRO 서열(서열번호 29, 서열번호 30, 및 서열번호 31)을 (5') B. 서브틸리시 *rrnI*-P2 프로모터/5'-UTR *aprE* 영역 DNA 서열의 변이체를 포함하는 폴리뉴클레오티드 구성체(다운스트림 카나마이신(*kan*) 유전자 발현 카세트(서열번호 7)를 포함하는 다운스트림(3') *aprE* 유전자 플랭킹 영역(3' *aprE* 유전자 FR) 서열에 작동가능하게 연결됨)에 작동가능하게 연결된 업스트림(5') *aprE* 유전자 플랭킹 영역을 포함하는 발현 카세트에 조립하여 B. 아밀로리케파시엔스 BPN' 종결자 DNA 서열(서열번호 6)에 작동가능하게 연결된 성숙 BG46_변이체 2 서브틸리신(서열번호 8)을 암호화하는 DNA 서열에 작동가능하게 연결된 변이체 프로-영역 서열을 암호화하는 DNA 서열에 작동가능하게 연결된 *AprE* 신호 펩티드 서열을 암호화하는 DNA 서열(서열번호 4)에 작동가능하게 연결된 서열번호 1을 생성하였다. 돌연변이/변이체 PRO 서열을 포함하는 카세트를 표준 분자 생물학 기법을 사용하여 바실러스 서브틸리스 세포로 형질전환시켰다.

[0279]

형질전환된 세포를 96웰 MTP에서 배양 배지(MOP 버퍼를 기반으로 한 농축 반정의 배지) 내 32℃, 300 rpm, 습도 80%의 진탕 배양기에서 3일간 성장시킨 후 원심분리 및 여과하였다. 정화된 배양 상층액을 사용하여 리포터 프로테아제 활성을 측정(분석)하여 생산성 수준을 확인하였으며, 샘플은 72시간 후에 채취하였다. 특히, 표 4는 리포터 단백질이 WT PRO 서열(서열번호 14)로 발현된 기준 구성체와 비교하여 72시간 후의 리포터 단백질 생산성(성능 지수(PI) 값)의 결과를 나타낸 것이다.

표 4

변이체 C 프로-영역 서열(서열번호 14) 대비
프로-영역 변이체 D-F(서열번호 29~31) 성능 지수

명칭	변이체	PI	서열번호
변이체 C	GK	1.00	14
변이체 E	GKA	1.3	30
변이체 D	GKS	1.2	29
변이체 F	GKAA	1.4	31

[0280]

[0281]

위의 표 4에 도시된 바와 같이, 프로-영역 변이체의 PI 값을 프로-영역 서열 변이체 프로-영역 서열 C(서열번호 14, "G₂K₃")와 비교하였고, 여기서 아미노산 위치 넘버링을 서열번호 14(즉, "GK" 삽입 포함)와 비교하였다. 3개의 N-말단 프로-영역 돌연변이체(AGKA, AGKS, AGKAA)의 PI 지수는 변이체 C(AGK; 서열번호 14) 대비 72시간 성장 후 리포터 단백질 BG46_변이체 2 서브틸리신(서열번호 8)의 높은 발현을 보였다.

[0282]

실시예 7

[0283] GG36 PRO 서열의 C-말단 변형

[0284] 본 실시예에서 위치 68의 류신(L)을 라이신(K)으로 치환하고 위치 72의 이소류신(I)을 발린(V)으로 치환하거나 위치 80의 글루탐산(E)을 이소류신(I)으로 치환하도록 GG36 변이체 프로-영역 서열 B(서열번호 11)를 추가로 조작하여 2개의 조작된 프로-영역 서열 변이체 서열 G(서열번호 32) 및 H(서열번호 33)를 생성하였다(예를 들어 도 1~3 참조). 특히, 형질전환된 세포는 전술한 바와 같이 성장시켰고, 정화된 배양 상층액을 사용하여 리포터 프로테아제 활성을 측정(분석)하여 생산성 수준을 확인하였으며, 샘플은 72시간 후에 채취하였다.

[0285] 예를 들어, 표 5는 리포터 단백질이 변이체 B 프로-영역 서열(서열번호 11)로 발현된 기준 구성체와 비교하여 72시간 후의 리포터 단백질 생산성(PI 값)의 결과를 나타낸 것이다. 보다 구체적으로, 표 5에 제시된 바와 같이, 변이체 프로-영역 서열 G 및 H의 PI 값은 기준/대조군 변이체 B 프로-영역 서열 대비 증가했으며, 여기서 변이체 프로-영역 서열 G(L68K 및 I72V 돌연변이 포함)는 프로-영역 서열 B(E30G) 대비 1.08의 PI를 나타냈고, 변이체 프로-영역 서열 H(L68K 및 E80I 돌연변이 포함)는 프로-영역 서열 B(E30G) 대비 1.18의 PI를 나타냈다.

표 5**프로-영역 변이체 C 대비 프로-영역 변이체 G 및 H 성능 지수**

명칭	변이체	PI	D 전하 vs WT
변이체 B	E30G	1.0	1
변이체 G	E30G L68K I72V	1.4	2
변이체 F	E30G L68K E80I	1.5	3

[0286]**[0287] 참조문헌****[0288]** PCT 공개 W01993/20214호**[0289]** PCT 공개 W02002/14490호**[0290]** PCT 공개 W02003/083125호**[0291]** PCT 공개 W02008/112258호**[0292]** PCT 공개 W02010/123754호**[0293]** PCT 공개 W02016/205710호**[0294]** Altschul *et al.*, “Basic local alignment search tool”, *J. Mol. Biol.*, 215(13):403-410, 1990.**[0295]** Altschul *et al.*, “Gapped BLAST and PSI BLAST a new generation of protein database search programs”, *Nucleic Acids Res, Set 1*;25(17):3389-402, 1997.**[0296]** Karlin *et al.*, “Applications and statistics for multiple high-scoring segments in molecular sequences”, *PNAS USA*, 90(12):5873-5787, 1993.**[0297]** Saitou and Nei, “The neighbor-joining method: a new method for reconstructing phylogenetic trees”, *Mol. Biol. Evol.*, Vol. 4, issue 4, pages 406-425, 1987.**[0298]** Schaffer *et al.*, “Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements”, *Nucleic Acids Res*, 29(14):2994-3005, 2001.**[0299]** Thompson *et al.*, “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”, *Nucleic Acids Res*, 22(22):4673-4680, 1994.**[0300]** Beaucage and Caruthers, “Deoxynucleoside phosphoramidites - A new class of key intermediates for deoxypolynucleotide synthesis”, *Tetrahedron Lett.*, 22,1859-1862, 1981.**[0301]** Matthes *et al.*, “Simultaneous rapid chemical synthesis of over one hundred oligonucleotides on a microscale”, *EMBO J.*, 3:801-805, 1984.

도면

도면1

야생형 및 변이체 프로-영역 서열

야생형 B. 렌투스 프로-영역 서열 (서열번호 15)

AEEAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 A (서열번호 9)

AEEAKEKYLIGFNEQEAVSEFVEQVEANDH~~V~~AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 B (서열번호 11)

AEEAKEKYLIGFNEQEAVSEFVEQVEANDG~~V~~AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 C (서열번호 14)

AG~~K~~EAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 D (서열번호 29)

AG~~K~~SEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 E (서열번호 30)

AG~~K~~AAEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 F (서열번호 31)

AG~~K~~AAEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 G (서열번호 32)

AEEAKEKYLIGFNEQEAVSEFVEQVEANDG~~V~~AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

변이체 프로-영역 서열 H (서열번호 33)

AEEAKEKYLIGFNEQEAVSEFVEQVEANDG~~V~~AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA~~EV~~TTM

도면2

프로-영역 서열 정렬

A--EEAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 15)
A--EEAKEKYLIGFNEQEAVSEFVEQVEANDH ~~V~~ AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 09)
A--EEAKEKYLIGFNEQEAVSEFVEQVEANDG ~~V~~ AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 11)
AG ~~K~~ EEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 14)
AG ~~K~~ SEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 29)
AG ~~K~~ AAEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 30)
AG ~~K~~ AAEAAKEKYLIGFNEQEAVSEFVEQVEANDEVAILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 31)
A--EEAKEKYLIGFNEQEAVSEFVEQVEANDG ~~V~~ AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 32)
A--EEAKEKYLIGFNEQEAVSEFVEQVEANDG ~~V~~ AILSEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDA ~~EV~~ TTM	(SEQ 33)

도면3

프로-영역 아미노산 위치 넘버링

야생형(기준) B. *렌투스* 프로-영역 서열 (서열번호 15, E30)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29**E30**V31A32I33L34S35E36E37E38E39V40
E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67L68D69P70A71I72S73Y74I75E76E77D78
A79E80V81T82T83M84

변이체(기준) 프로-영역 A 서열 (서열번호 9, H30)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29**H30**V31A32I33L34S35E36E37E38E39V40
E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67L68D69P70A71I72S73Y74I75E76E77D78
A79E80V81T82T83M84

변이체(기준) 프로-영역 B 서열 (서열번호 11, G30)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29**G30**V31A32I33L34S35E36E37E38E39V40
E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67L68D69P70A71I72S73Y74I75E76E77D78
A79E80V81T82T83M84

변이체(기준) 프로-영역 C 서열 (서열번호 14, GK 삽입)

A1**G2**K3E4E5A6K7E8K9Y10L11I12G13F14N15E16Q17E18A19V20S21E22F23V24E25Q26V27E28A29N30D31E32V33A34I35L36S37E38E39E40
E41V42E43I44E45L46L47H48E49F50E51T52I53P54V55L56S57V58E59L60S61P62E63D64V65D66A67L68E69L70D71P72A73I74S75Y76I77E78
E79D80A81E82V83T84T85M86

변이체(기준) 프로-영역 D 서열 (서열번호 29, GKS 삽입)

A1**G2K3S4**E5E6A7K8E9K10Y11L12I13G14F15N16E17Q18E19A20V21S22E23F24V25E26Q27V28E29A30N31D32E33V34A35I36L37S38E39E40E41
E42V43E44I45E46L47L48H49E50F51E52T53I54P55V56L57S58V59E60L61S62P63E64D65V66D67A68L69E70L71D72P73A74I75S76Y77I78E79
E80D81A82E83V84T85T86M87

변이체(기준) 프로-영역 E 서열 (서열번호 30, GKA 삽입)

A1**G2K3A4**E5E6A7K8E9K10Y11L12I13G14F15N16E17Q18E19A20V21S22E23F24V25E26Q27V28E29A30N31D32E33V34A35I36L37S38E39E40E41
E42V43E44I45E46L47L48H49E50F51E52T53I54P55V56L57S58V59E60L61S62P63E64D65V66D67A68L69E70L71D72P73A74I75S76Y77I78E79
E80D81A82E83V84T85T86M87

변이체(기준) 프로-영역 F 서열 (서열번호 31, GKAA 삽입)

A1**G2K3A4A5**E6E7A8K9E10K11Y12L13I14G15F16N17E18Q19E20A21V22S23E24F25V26E27Q28V29E30A31N32D33E34V35A36I37L38S39E40E41
E42E43V44E45I46E47L48L49H50E51F52E53T54I55P56V57L58S59V60E61L62S63P64E65D66V67D68A69L70E71L72D73P74A75I76S77Y78I79
E80E81D82A83E84V85T86T87M88

변이체(기준) 프로-영역 G 서열 (서열번호 32, L68K + I72V 치환)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29**G30**V31A32I33L34S35E36E37E38E39V40
E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67**K68**D69P70A71**V72**S73Y74I75E76E77D78
A79E80V81T82T83M84

변이체(기준) 프로-영역 H 서열 (서열번호 33, L68K + E80I 치환)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29**G30**V31A32I33L34S35E36E37E38E39V40
E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67**K68**D69P70A71I72S73Y74I75E76E77D78
A79**I80**V81T82T83M84

서 열 목 록 (첨부)



아이콘을 클릭하시면 서열목록 파일이 열립니다.

본 공보 PDF는 첨부파일을 가지고 있습니다. Acrobat Reader PDF뷰어를 제공하지 않는 브라우저(크롬, 파이어폭스, 사파리 등)의 경우 첨부파일 열기가 제한되어 있으므로 Acrobat Reader PDF뷰어 설치 후 공보 PDF를 다운로드 받아 해당 뷰어에서 조회해주시기 바랍니다.