



등록특허 10-2593668



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년10월24일
(11) 등록번호 10-2593668
(24) 등록일자 2023년10월20일

- (51) 국제특허분류(Int. Cl.)
C12N 15/80 (2006.01) *C07K 14/37* (2006.01)
C12N 1/14 (2018.01) *C12N 9/42* (2006.01)
C12P 21/02 (2006.01)
- (52) CPC특허분류
C12N 15/80 (2013.01)
C07K 14/37 (2013.01)
- (21) 출원번호 10-2019-7010929
- (22) 출원일자(국제) 2017년10월03일
심사청구일자 2020년09월29일
- (85) 번역문제출일자 2019년04월16일
- (65) 공개번호 10-2019-0057088
- (43) 공개일자 2019년05월27일
- (86) 국제출원번호 PCT/US2017/054983
- (87) 국제공개번호 WO 2018/067599
국제공개일자 2018년04월12일
- (30) 우선권주장
62/403,787 2016년10월04일 미국(US)

- (56) 선행기술조사문헌
WO2011151512 A2*
Microbiology (2012), 158, pp. 58-68.*
hypothetical protein M419DRAFT_98455
[Trichoderma reesei RUT C-30], GenBank:
ETS02557.1, 2015.03.23.*

*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 28 항

심사관 : 김정희

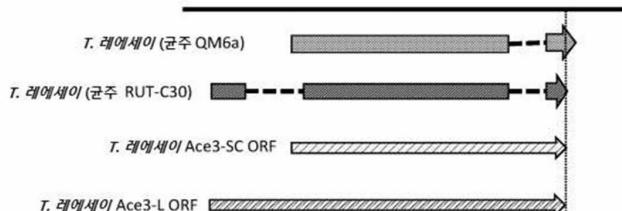
(54) 발명의 명칭 유도 기질 부재 하의 사상 진균 세포에서의 단백질 생산

(57) 요 약

본 발명의 일부 구현예는 변이체 사상 진균 세포, 이의 조성을 및 이의 1종 이상의 관심 단백질의 증가된 생산을 위한 방법을 대상으로 한다. 더욱 구체적으로는, 일부 구현예에서, 본 발명은 모 사상 진균 세포로부터 유래된 변이체 사상 진균 (숙주) 세포를 대상으로 하며, 이때, 변이체 숙주 세포는 유도 기질의 부재 하에 관심 단백질 (뒷면에 계속)

대 표 도 - 도1a

T. 레에세이/균주 QM6a 및 RUT-C30으로부터의 Ace3 단백질 암호화 영역의 개략도



(POI)의 발현/생산을 가능하게 하는 유전자 변형을 포함한다. 일부 구현예에서, 본 발명의 변이체 진균 속주 세포는 본 설명에서 Ace3-L로 지칭된 Ace3 단백질을 암호화하는 변이체 셀룰라아제 발현 활성화 인자 3(activator of cellulase expression 3, ace3) 유전자의 발현을 증가시키는 유전자 변형을 포함한다.

(52) CPC특허분류

C12N 1/14 (2021.05)

C12N 9/2437 (2013.01)

C12P 21/02 (2013.01)

(72) 발명자

벤데주 펠리페 오세아스

미국 19317 펜실베이니아주 채즈 포드 존슨 팜 레
인 7400 아파트먼트 115

발꼬넨 마리

핀랜드 04350 나흐겔라 니뚜멘띠에 7

살로헤이모 마르꾸

핀랜드 00390 헬싱키 퀸페엔띠에 18 비

아로 니나

핀랜드 06450 뾰르보 까이반또띠에 36

뻬끌라 띠나

핀랜드 02240 에스뽀 폐흐띠깔리온띠에 16 에이

명세서

청구범위

청구항 1

모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포로서, 변이체 세포는 서열 번호 6 의 Ace3 단백질을 암호화하는 도입된 폴리뉴클레오티드를 포함하고,

이때, 변이체 세포는 모 세포에 비해 유도 기질의 부존재 하에 증가된 양의 관심 단백질 (POI) 을 생산하며, 상기 관심 단백질은 셀룰라아제 또는 헤미셀룰라아제이고,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하고,

상기 사상 진균 세포는 트리코더마 종 진균 세포인, 변이체 사상 진균 세포.

청구항 2

제1항에 있어서, 변이체 세포는 모 세포에 비해 유도 기질의 존재 하에서 증가된 양의 관심 단백질 (POI) 을 더 생산하며,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하는 것인, 변이체 사상 진균 세포.

청구항 3

제1항에 있어서, POI는 내인성 POI 또는 이종성 POI인 변이체 사상 진균 세포.

청구항 4

제1항에 있어서, 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드를 포함하는 변이체 사상 진균 세포.

청구항 5

삭제

청구항 6

삭제

청구항 7

제1항에 있어서 셀룰라아제는 *cbh1*, *cbh2*, *eg11*, *eg12*, 및 *eg13* 로 구성된 군으로부터 선택되는 것인, 변이체 사상 진균 세포.

청구항 8

삭제

청구항 9

제1항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈 내로 통합되는 것인, 변이체 사상 진균 세포.

청구항 10

제9항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 텔로미어 부위 내로 통합되는 것인, 변이체 사상 진균 세포.

청구항 11

제9항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 글루코아밀라아제 (*gla1*) 유전자좌 내로 통합되는 것인,

변이체 사상 진균 세포.

청구항 12

제1항에 있어서, 폴리뉴클레오티드는 서열 번호 4 에 대해 90% 이상의 서열 동일성을 포함하는 뉴클레오티드 서열을 포함하는 것인, 변이체 사상 진균 세포.

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 내인성 셀룰라아제 또는 헤미셀룰라아제의 생산 방법으로,

(i) 5'에서 3' 방향으로 (a) 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드를 진균 세포 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 6 의 Ace3 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하는 것인, 단계, 및

(ii) 진균 세포 성장 및 단백질 생산을 위한 조건 하에서 단계 (i)의 세포를 배양하는 단계로서, 이때, 그러한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함하고,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하는 것인, 방법.

청구항 18

제17항에 있어서, 폴리뉴클레오티드는 제3 핵산 서열을 더 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이때, 제3 핵산 서열은 천연 ace3 종결자 서열을 포함하는 것인, 방법.

청구항 19

제17항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈 내로 통합되는 것인, 방법.

청구항 20

제17항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 텔로미어 부위 내로 통합되는 것인, 방법.

청구항 21

제17항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 글루코아밀라아제 (*gla1*) 유전자와 내로 통합되는 것인, 방법.

청구항 22

제17항에 있어서, 폴리뉴클레오티드는 서열 번호 4 에 대해 90% 이상의 서열 동일성을 포함하는 뉴클레오티드 서열을 포함하는 것인, 방법.

청구항 23

삭제

청구항 24

제17항에 있어서, 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택되는 것인, 방법.

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 이종성 관심 단백질의 생산 방법으로,

(i) 5'에서 3' 방향으로 (a) 구성성 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드를 진균 세포 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 6 의 *Ace3* 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하는 것인, 단계, 및

(ii) 진균 세포 성장 및 단백질 생산을 위한 조건 하에서 단계 (i)의 세포를 배양하는 단계로서, 그러한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함하고,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하고,

상기 이종성 관심 단백질은 셀룰라아제 또는 헤미셀룰라아제인, 방법.

청구항 34

제33항에 있어서, 진균 세포는 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드를 포함하는데, 이때, 폴리뉴클레오티드는 단계 (i) 이전에, 단계 (i) 중에, 또는 단계 (i) 이후에 진균 세포 내로 도입되는 것인, 방법.

청구항 35

제33항에 있어서, 폴리뉴클레오티드는 제3 핵산 서열을 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이때, 제3 핵산 서열은 천연 *ace3* 종결자 서열을 포함하는 것인, 방법.

청구항 36

제33항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈 내로 통합되는 것인, 방법.

청구항 37

제33항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 텔로미어 부위 내로 통합되는 것인, 방법.

청구항 38

제33항에 있어서, 폴리뉴클레오티드는 진균 세포 게놈의 글루코아밀라아제 (*gla1*) 유전자와 내로 통합되는 것인, 방법.

청구항 39

제33항에 있어서, 폴리뉴클레오티드는 서열 번호 4에 대해 90% 이상의 서열 동일성을 포함하는 뉴클레오티드 서열을 포함하는 것인, 방법.

청구항 40

삭제

청구항 41

삭제

청구항 42

제33항에 있어서, 프로모터는 *rev3* 프로모터(서열 번호 15), *b11* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택되는 것인, 방법.

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

유도 기질의 부존재 하에서의 내인성 셀룰라아제 또는 헤미셀룰라아제의 증가된 생산을 위하여 트리코더마 레에세이 주를 유전자 변형하는 방법으로,

(i) 서열 번호 3의 Ace3-S 단백질, 서열 번호 8의 Ace3-SC 단백질 또는 서열 번호 10의 Ace3-LC 단백질을 암호화하는 *ace3* 유전자를 포함하는 T. 레에세이 주를 스크리닝하고 확인하는 단계로서, 이때, 확인된 T. 레에세이 주는 서열 번호 6의 Ace3-L 단백질, 서열 번호 14의 Ace3-LN 단백질 또는 서열 번호 12의 Ace3-EL 단백질을 암호화하는 *ace3* 유전자를 포함하지 않는 것인, 단계,

(ii) 5'에서 3' 방향으로 (a) 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드를 단계 (i)에서 확인된 T. 레에세이 주 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 6의 Ace3 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하거나, 제2 핵산 서열은 서열 번호 12의 Ace3 단백질을 암호화하는 것인, 단계, 및

(iii) 진균 세포 성장 및 단백질 생산을 위한 조건 하에서 단계 (ii)의 세포를 배양하는 단계로서, 이때, 그러한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함하고,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하는 것인, 방법.

청구항 55

모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포로서, 변이체 세포는 이종성 프로모터를 갖는 서열 번호 6의 Ace3-L 단백질, 서열 번호 14의 Ace3-LN 단백질 또는 서열 번호 12의 Ace3-EL 단백질을 암호화하는 *ace3* 유전자를 포함하고, 이때, 변이체 세포는 모 세포에 비해 유도 기질의 부존재 하에 증가된 양의 셀룰라아제 또는 헤미셀룰라아제를 생산하며,

상기 유도 기질은 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하고,

상기 사상 진균 세포는 트리코더마 종 진균 세포인, 변이체 사상 진균 세포.

청구항 56

제55항에 있어서, 이종성 프로모터는 트리코더마 레에세이 프로모터인, 변이체 사상 진균 세포.

청구항 57

제56항에 있어서, 이종성 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번

호 27)로 구성된 군으로부터 선택된 프로모터인, 변이체 사상 진균 세포.

발명의 설명

기술 분야

[0001] [관련 출원에 대한 상호 참조]

[0002] 본 출원은 2016년 10월 04일자로 출원된 미국 가출원 제62/403,787호의 이익을 청구하며, 본 설명에 그 전체가 참조로 포함된다.

[0003] [서열 목록]

[0004] EFS를 통하여 함께 제출된 서열 목록 텍스트 파일은 2017년 10월 02일에 생성되고, 크기가 157 킬로바이트인 파일 "NB41159WOPCT_SEQLISTING.txt"를 포함한다. 본 서열 목록은 37 C.F.R. § 1.52(e)를 준수하며, 본 설명에 그 전체가 참조로 포함된다.

[0005] [기술분야]

[0006] 본 발명은 일반적으로 문자생물학, 생화학, 단백질 생산 및 사상 진균 분야에 관한 것이다. 본 발명의 일부 구현에는 변이체 사상 진균 세포, 이의 조성을 및 이의 1종 이상의 관심 단백질의 증가된 생산을 위한 방법을 대상으로 한다. 더욱 구체적으로는, 일부 구현예에서, 본 발명은 모 사상 진균 세포로부터 유래된 변이체 사상 진균 (숙주) 세포를 대상으로 하며, 이때, 변이체 숙주 세포는 유도 기질의 부재 하에 1종 이상의 관심 단백질 (POI)의 발현/생산을 가능하게 하는 유전자 변형을 포함한다.

배경 기술

[0007] 리그노셀룰로오스 식물성 물질의 구성요소인 셀룰로오스는 자연계에서 발견되는 가장 풍부한 다당류이다. 마찬가지로, 사상 진균은 식물성 바이오매스의 효율적인 분해자임이 당해 분야에 공지되어 있으며, 실제 산업상 관련 있는 리그노셀룰로오스 분해 효소(이하, 집합적으로 "셀룰라아제" 효소라 지칭됨)의 주된 공급원이다. 예를 들어, 사상 진균은 셀룰로오스의 β -(1,4)-연결된 글리코시드 결합을 가수분해하여 글루코오스를 생산하는(즉, 그에 의해 성장을 위해 셀룰로오스를 활용하는 이러한 사상 진균의 능력을 부여하는) 세포 외 셀룰라아제 효소(예컨대, 셀로비오하이드롤라아제, 엔도글루카나아제, β -글루코시다아제)를 생산한다고 알려져 있다.

[0008] 특히, 사상 진균 트리코더마 레에세이(*Trichoderma reesei*, *T. reesei*; 진균 하이포크레아 제코리나(*Hypocrea jecorina*)의 무성생식형)는 셀룰라아제 효소의 효율적인 생산자라고 알려져 있다(예컨대, PCT 국제 출원 제WO1998/15619호, 제WO2005/028636호, 제WO2006/074005호, 제WO1992/06221호, 제WO1992/06209호, 제WO1992/06183호, 제WO2002/12465호 등 참조). 따라서, *T. 레에세이*와 같은 사상 진균은 셀룰로오스 유래 에탄올, 섬유 및 의류, 세제, 섬유질, 식품 및 사료 첨가제와 같은 이러한 상품의 생산 및 기타 산업적 용도에 있어서 가치 있는 효소를 생산하는 그것들의 능력을 위하여 활용되어 왔다.

[0009] 트리코더마에서 이러한 산업적으로 관련 있는 효소의 발현(및 생산)은 성장에 이용할 수 있는 탄소 공급원에 의존한다고 알려져 있다. 더욱 상세하게는, 사상 진균에 의한 셀룰라아제 효소의 생산은 에너지를 소비하는 과정이며, 따라서, 이들 효소의 효율적인 생산을 확실히 하기 위하여 유도 및 억제 메커니즘이 사상 진균에서 진화하였다. 예를 들어, 식물 세포벽 물질의 분해에 필요한 효소(즉, 셀룰라아제/헤미셀룰라아제)를 암호화하는 다양한 유전자는 "유도" 기질의 존재 하에 "활성화되고", "탄소 이화대사산물 억제"(이하, "CCR")라고 알려져 있는 메커니즘을 통해 식물성 바이오매스에 비해 선호되는 용이하게 대사된 탄소 공급원(예컨대, D-글루코오스)의 존재 하에 "억제된다". 따라서, 셀룰라아제 유전자는 글루코오스에 의해 엄중하게 억제되고, 셀룰로오스 및 일부 이당류(예컨대, 소포로오스, 락토오스, 젠티오비오스)에 의해 수천 배 유도된다. 예를 들어, 주요 셀로비오하이드롤라아제 1(*cbh1*)의 발현 수준은 글루코오스 함유 배지와 비교하여 셀룰로오스 또는 소포로오스와 같은 유도성 탄소 공급원을 함유하는 배지 상에서 수천 배 "상향 조절된다"(Ilmen et al., 1997). 나아가, "유도된" *T. 레에세이* 배양물에 대한 "억제성" 탄소 공급원의 첨가는 (셀룰로오스 또는 소포로오스) 유도를 무력화하여, 셀룰라아제 유전자 발현의 하향 조절을 초래하는 것으로 나타났다(el-Gogary et al., 1989; Ilmen et al., 1997). 따라서, 셀룰라아제 시스템(효소)을 포함하는 유전자의 발현은 전사 수준에서 적어도 조직화되고 조절되는데, 이때, 이 시스템의 유전자 구성원은 상승적으로 작용하고, 위에서 언급된 바와 같이, 셀룰로오스의 가용성 올리고당으로의 가수분해에 필수적이다.

[0010]

더욱 구체적으로, 전 유전체 분석은 *T. 레에세이*에는 적어도 10종의 셀룰로오스 분해성 및 16종의 자일란 분해성 효소 암호화 유전자가 있음을 보여주었다(Martinez *et al.*, 2008). 특히, 가장 많이 분비된 효소는 두 가지 주요 셀로비오하이드롤라아제(EC. 3.2.1.91)인 *cbh1*(셀로비오하이드롤라아제 1) 및 *cbh2*(셀로비오하이드롤라아제 2), 그리고 두 가지 주요 특이적 엔도- β -1,4-자일라나아제(EC 3.3.1.8)인 *xyn1*(엔도-1,4-베타-자일라나아제 1) 및 *xyn2*(엔도-1,4-베타-자일라나아제 2)이며, 본 설명에서 "주요 산업적으로 관련 있는 헤미셀룰라아제 및 셀룰라아제(major industrially relevant hemicellulases and cellulases)" 또는 "MIHC"라 지칭된다. 이들 MIHC는 추가적인 효소와 함께 작용하여 셀룰로오스와 자일란을 분해하며, 이는 가용성 올리고당 및 단당류, 예컨대, 셀로비오스, D-글루코오스, 자일로비오스 및 D-자일로오스의 생성을 초래한다. 또한, 소포로오스는 이들 효소 중 일부의 글리코실전이반응 활성의 생성물이다(Vaheri *et al.*, 1979). 더욱 상세하게는, 이들 가용성 올리고당 및 단당류(즉, 셀로비오스, D-글루코오스, 자일로비오스, D-자일로오스 및 소포로오스)는 문헌에서 *T. 레에세이*에서 MICH의 발현에 영향을 미친다고 보고된 바 있다. 예를 들어, D-글루코오스의 존재는 CCR을 유발하며, 이는 소량의 MIHC의 분비를 초래한다. 소포로오스는 *cbh1* 및 *cbh2*의 발현을 위한 가장 강력한 유도인자인 것으로 여겨진다. D-자일로오스는 농도 의존적인 방식으로 *xyn1* 및 *xyn2* 발현을 조절한다.

[0011]

일반적으로, 트리코더마와 같은 사상 진균에 의한 효소/폴리펩티드의 상업적 규모의 생산은 전형적으로, 배치식, 유가식 및 연속 순환식 공정을 포함한 고체 또는 액침 배양에 의해 이루어진다. 예를 들어, 트리코더마에서의 산업적 셀룰라아제 생산의 가장 문제가 많고 값비싼 측면 중 하나는 트리코더마 숙주 세포에 적절한 유도인자(즉, 유도 기질)를 제공하는 것이다. 예를 들어, 실험실 규모의 실험의 경우와 같이, 상업적 규모의 셀룰라아제(효소) 생산은 고체 셀룰로오스(즉, 유도 기질) 상에 진균 세포를 성장시키거나 "락토오스"(즉, 유도 기질)와 같은 이당류 유도인자의 존재 하에서 세포를 배양함으로써 "유도된다".

[0012]

불행히도, 산업적 규모에서는 "유도"의 두 가지 방법은 셀룰라아제 생산과 관련된 높은 비용을 초래하는 결점이 있다. 예를 들어, 전술한 바와 같이, 셀룰라아제 합성은 셀룰로오스 유도 및 글루코오스 억제 둘 다를 거치게 된다. 따라서, 유도성 프로모터의 제어 하에 셀룰라아제 효소의 수율에 영향을 미치는 중요한 인자는 셀룰로오스 기질과 글루코오스 농도 사이의 적절한 균형을 유지하는 것이다(즉, 그것은 조절된 유전자 생성물의 합리적인 상업적 수율을 얻는 데 중요하다). 셀룰로오스는 효과적이고 저렴한 유도인자이지만, 사상 진균 세포를 고체 셀룰로오스 상에서 성장시킬 때 글루코오스 농도를 조절하는 것은 문제가 될 수 있다. 낮은 농도의 셀룰로오스에서, 글루코오스 생산은 활성 세포 성장 및 기능의 대사적 요구를 충족시키기에는 너무 느릴 수 있다. 반면, 셀룰라아제 합성은 글루코오스 생성이 그것의 소비보다 빠를 때 글루코오스 억제에 의해 중단될 수 있다. 따라서, 느린 기질 첨가 및 글루코오스 농도 모니터링을 제공하기 위해서는 값비싼 공정 제어 계획이 필요하다(Ju and Afolabi, 1999). 또한, 기질의 느린 연속적 전달은 셀룰로오스 물질의 고체 성질의 결과로서 달성하기가 어려울 수 있다.

[0013]

"유도 기질"로서의 셀룰로오스의 사용과 관련된 문제 중 일부는 "락토오스", "소포로오스" 또는 "젠티오비오스"와 같은 가용성 "유도 기질"의 사용을 통해 극복될 수 있다. 예를 들어, "유도 기질"로서 락토오스를 사용할 때, 락토오스는 유도인자 및 탄소 공급원으로서 기능하기 위하여 고농도로 제공되어야 한다(예컨대, Seiboth, *et. al.*, 2002 참조). 소포로오스는 셀룰로오스보다 더 강력한 유도인자이지만, 소포로오스는 값이 비싸고, 제조가 어렵다. 예를 들어, 글루코오스, 소포로오스 및 (즉, 글루코오스의 효소적 전환을 통해 생성된) 다른 이당류의 혼합물은 셀룰라아제의 효율적인 생산에 이용될 수 있는데, 이는 글루코오스를 단독으로 사용하는 것보다 더 많은 (생산) 비용을 초래한다. 따라서, 고체 셀룰로오스보다 다루기 쉽고 제어하기 쉽지만, 그럼에도 유도 기질로서 소포로오스를 사용하는 것은 셀룰라아제 생산 비용을 엄두도 못 낼 정도로 비싸게 만들 수 있다.

[0014]

전술한 바에 기초하여, 그러한 생산을 위하여 값비싼 유도 기질(예컨대, 소포로오스, 락토오스 등)을 제공할 필요성 또는 요구 없이, 사상 진균에 의한 효소/폴리펩티드의 비용 효과적인 상업적 규모의 생산을 위한 계속 진행 중인, 그리고 충족되지 않은 필요성이 당업계에 남아 있음을 명백하다. 더욱 상세하게는, 사상 진균 숙주 세포에 의한 1종 이상의 내인성 리그노셀룰로오스 분해 효소의 상업적 규모의 생산에 대한 필요성이 당업계에 남아 있으며, 이때, 그러한 사상 진균 세포는 유도 기질의 부존재 하에서 이들 유전자 중 1종 이상을 발현할 수 있다. 또한, 그러한 사상 진균 숙주 세포에서 발현되고 생산된 1종 이상의 이종성 단백질 생성물의 비용 효과적인 생산을 위한 당업계의 충족되지 않은 요구가 있으며, 이때, 이러한 단백질을 암호화하는 이종 유전자는 진균 숙주 세포 내로 도입되며, 이 진균 숙주 세포는 유도 기질의 부존재 하에 그러한 이종 유전자를 발현할 수 있다.

발명의 내용

[0015]

본 발명의 일부 구현예는 그러한 생산을 위하여 값비싼 유도 기질(예컨대, 소포로오스, 락토오스 등)을 제공할 필요성 또는 요구가 없는, 사상 진균에 의한 효소/폴리펩티드의 상업적 규모의 생산에 관한 것이다. 따라서, 일부 다른 구현예는 변이체 사상 진균 세포, 이의 조성물 및 이의 1종 이상의 관심 단백질의 증가된 생산을 위한 방법에 관한 것이다. 예를 들어, 본 발명의 일부 구현예는 모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포에 관한 것으로, 변이체 세포는 서열 번호 6의 Ace3 단백질에 대해 약 90%의 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 도입된 폴리뉴클레오티드 구성체를 포함하고, 이때, 변이체 세포는 모 세포에 비해 유도 기질의 부존재 하에 증가된 양의 관심 단백질(POI)을 생산하며, 이때, 변이체 세포와 모 세포는 유사한 조건 하에서 배양된다. 일부 다른 구현예에서, 변이체 세포는 모 세포에 비해 유도 기질의 존재 하에서 증가된 양의 POI를 생산하는데, 이때, 변이체 세포와 모 세포는 유사한 조건 하에서 배양된다.

[0016]

변이체 세포의 또 다른 구현예에서, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 암호화된 Ace3 단백질은 마지막 4개의 C 말단 아미노산으로 "Lys-Ala-Ser-Asp"를 포함한다. 또 다른 구현예에서, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 암호화된 Ace3 단백질은 서열 번호 6에 작동 가능하게 연결되며 이에 앞서는 서열 번호 98의 N 말단 아미노산 단편을 포함한다. 변이체 세포의 일부 다른 구현예에서, Ace-3 단백질은 서열 번호 12에 대해 약 90% 서열 동일성을 포함한다. 또 다른 구현예에서, Ace3 단백질을 암호화하는 도입된 폴리뉴클레오티드는 서열 번호 5에 대하여 약 90% 동일성을 포함하는 오픈 리딩 프레임(ORF) 서열을 포함한다.

[0017]

변이체 세포의 또 다른 구현예에서, POI는 내인성 POI 또는 이종성 POI이다. 따라서, 일부 구현예에서, 변이체 세포는 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드 구성체를 포함한다. 또 다른 구현예에서, 이종성 POI를 암호화하는 폴리뉴클레오티드 구성체는 셀룰로오스 유도성 유전자 프로모터의 제어 하에 발현된다. 일부 다른 구현예에서, 내인성 POI는 리그노셀룰로오스 분해 효소이다. 따라서, 다른 구현예에서, 리그노셀룰로오스 분해 효소는 셀룰라아제 효소, 헤미셀룰라아제 효소, 또는 이의 조합으로 구성된 군으로부터 선택된다. 일부 다른 구현예에서, 리그노셀룰로오스 분해 효소는 *cbh1*, *cbh2*, *eg11*, *eg12*, *eg13*, *eg14*, *eg15*, *eg16*, *bg11*, *bg12*, *xyn1*, *xyn2*, *xyn3*, *bx11*, *abf1*, *abf2*, *abf3*, *axe1*, *axe2*, *axe3*, *man1*, *ag11*, *ag12*, *ag13*, *glr1*, *sw01*, *cip1* 및 *cip2*로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 이종성 POI는 α -아밀라아제, 알칼리성 α -아밀라아제, β -아밀라아제, 셀룰라아제, 텍스트라나아제, α -글루코시다아제, α -갈락토시다아제, 글루코아밀라아제, 헤미셀룰라아제, 펜토사나아제, 자일라나아제, 인버타아제, 락타아제, 나린지나아제, 펩티나아제, 풀루라나아제, 산성 프로테아제, 알칼리성 프로테아제, 브로멜라인, 중성 프로테아제, 파파인, 펩신, 펩티다아제, 레닛, 레닌, 키모신, 서브틸리신, 서몰리신, 아스파르트산 프로테이나아제, 트립신, 리파아제, 에스테라아제, 포스포리파아제, 포스파타아제, 피타아제, 아미다아제, 이미노아실라아제, 글루타미나아제, 리소자임, 페니실린 아실라아제; 이소메라제, 옥시도리덕타아제, 카탈라아제, 클로로페록시다아제, 글루코오스 옥시다아제, 하이드록시스테로이드 데하이드로게나아제, 페록시다아제, 리아제, 아스파르트산 β -데카르복실라아제, 푸마라아제, 히스타다아제, 트랜스페라아제, 리가아제, 이미노펩티다아제, 카르복시펩티다아제, 키티나아제, 큐티나아제, 테옥시리보뉴클레아제, α -갈락토시다아제, β -갈락토시다아제, 락카아제, 만노시다아제, 뮤타나아제, 폴리페놀 옥시다아제, 리보뉴클레아제 및 트랜스글루타미나아제로 구성된 군으로부터 선택된다.

[0018]

변이체 세포의 또 다른 구현예에서, 폴리뉴클레오티드 구성체는 프로모터 서열 5'을 포함하고, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 폴리뉴클레오티드 서열에 작동 가능하게 연결된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 천연 *ace3* 종결자 서열 3'을 더 포함하고, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 폴리뉴클레오티드 서열에 작동 가능하게 연결된다. 일부 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 계놈 내로 통합된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 계놈의 텔로미어 부위 내로 통합된다. 일부 다른 구현예에서, 폴리뉴클레오티드 구성체는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 다른 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 95% 서열 동일성을 포함하는 아미노산 서열을 포함한다. 따라서, 다른 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14의 아미노산 서열을 포함한다.

[0019]

다른 구현예에서, 변이체 세포는 서열 번호 48의 야생형 자일라나아제 조절인자 1(Xyr1) 단백질 또는 서열 번호 46의 변이체 자일라나아제 조절인자 1(Xyr1) 단백질을 암호화하는 폴리뉴클레오티드를 발현하는 유전자 변형을 포함한다. 일부 다른 구현예에서, 변이체 세포는 내인성 탄소 이화대사산물 억제인자 1(Cre1) 단백질 또는 Ace1 단백질을 암호화하는 유전자의 발현을 감소시키거나 방지하는 유전자 변형을 포함한다. 또 다른 구현예에서, 변이체 세포는 Ace2 단백질 발현을 포함하는 유전자 변형을 포함한다. 다른 구현예에서, 사상 진균 세포는 자낭균

(Ascomycota) 아문의 폐지조마이코티나(*Pezizomycotina*) 세포이다. 일부 다른 구현예에서, 사상 진균 세포는 트리코더마 종 세포이다.

[0020] 다른 구현예에서, 본 발명은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 폴리뉴클레오티드 ORF에 관한 것이다.

[0021] 다른 구현예에서, 본 발명은 본 발명의 변이체 세포에 의해 생산된 리그노셀룰로오스 분해 효소에 관한 것이다. 다른 구현예에서, 본 발명은 본 발명의 세포의 변이체 세포에 의해 생산된 이종성 POI에 관한 것이다.

[0022] 다른 구현예에서, 본 발명은 유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 내인성 관심 단백질의 생산 방법에 관한 것으로, 방법은 (i) 5'에서 3' 방향으로 (a) 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드 구성체를 진균 세포 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하는 것인, 단계, 및 (ii) 진균 세포 성장 및 단백질 생산에 적합한 조건 하에서 단계 (i)의 세포를 발효시키는 단계로서, 이때, 그러한 적합한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함한다. 방법의 일부 구현예에서, 폴리뉴클레오티드 구성체는 제3 핵산 서열 3'을 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이때, 제3 핵산 서열은 천연 *ace3* 종 결자 서열을 포함한다. 또 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈 내로 통합된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 텔로미어 부위 내로 통합된다. 일부 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 글루코아밀라아제(*gla1*) 유전자와 내로 통합된다. 방법의 다른 구현예에서, 폴리뉴클레오티드 구성체는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 다른 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 95% 서열 동일성을 포함하는 아미노산 서열을 포함한다. 또 다른 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14의 아미노산 서열을 포함한다.

[0023] 방법의 일부 구현예에서, 단계 (i) 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택된다.

[0024] 방법의 관련된 구현예에서, 내인성 POI는 리그노셀룰로오스 분해 효소이다. 일부 구현예에서, 리그노셀룰로오스 분해 효소는 셀룰라아제 효소, 헤미셀룰라아제 효소, 또는 이의 조합으로 구성된 군으로부터 선택된다. 일부 구현예에서, 셀룰라아제 효소는 *cbh1*, *cbh2*, *eg11*, *eg12*, *eg13*, *eg14*, *eg15*, *eg16*, *bg11*, *bg12*, *swo1*, *cip1* 및 *cip2*로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 헤미셀룰라아제 효소는 *xyn1*, *xyn2*, *xyn3*, *xyn4*, *bx11*, *abf1*, *abf2*, *abf3*, *axe1*, *axe2*, *axe3*, *man1*, *ag11*, *ag12*, *ag13* 및 *g1r1*로 구성된 군으로부터 선택된다.

[0025] 방법의 다른 구현예에서, 단계 (i)은 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드 구성체를 더 포함한다. 일부 구현예에서, 이종성 POI를 암호화하는 폴리뉴클레오티드 구성체는 셀룰로오스 유도성 유전자 프로모터의 제어 하에 발현된다. 일부 구현예에서, 셀룰로오스 유도성 유전자 프로모터는 *cbh1*, *cbh2*, *eg11*, *eg12*, *xyn2* 또는 *stp1*로부터 선택된다.

[0026] 다른 구현예에서, 방법은 서열 번호 48의 야생형 자일라나아제 조절인자 1(*Xyr1*) 단백질 또는 서열 번호 46의 변이체 자일라나아제 조절인자 1(*Xyr1*) 단백질을 암호화하는 폴리뉴클레오티드를 발현하는 유전자 변형을 더 포함한다. 다른 구현예에서, 방법은 내인성 탄소 이화대사산물 억제인자 1(*Cre1*) 단백질 또는 Ace1 단백질을 암호화하는 유전자의 발현을 감소시키거나 방지하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 Ace2 단백질 발현을 포함하는 유전자 변형을 더 포함한다.

[0027] 다른 구현예에서, 본 발명은 유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 내인성 관심 단백질의 생산 방법에 관한 것으로, 방법은 (i) 5'에서 3' 방향으로 (a) 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드 구성체를 진균 세포 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 12에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 것인, 단계, 및 (ii) 진균 세포 성장 및 단백질 생산에 적합한 조건 하에서 단계 (i)의 세포를 발효시키는 단계로서, 이때, 그러한 적합한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함한다. 다른 구

현예에서, 폴리뉴클레오티드 구성체는 제3 핵산 서열 3'을 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이 때, 제3 핵산 서열은 천연 *ace3* 종결자 서열을 포함한다. 방법의 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈 내로 통합된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 텔로미어 부위 내로 통합된다. 일부 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 글루코아밀라아제(*gla1*) 유전자와 내로 통합된다. 방법의 또 다른 구현예에서, 폴리뉴클레오티드 구성체는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 일부 구현예에서, 암호화된 *Ace3* 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 95% 서열 동일성을 포함하는 아미노산 서열을 포함한다. 일부 다른 구현예에서, 암호화된 *Ace3* 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14의 아미노산 서열을 포함한다.

[0028] 따라서, 방법 8의 다른 구현예에서, 내인성 POI는 리그노셀룰로오스 분해 효소이다. 특정 구현예에서, 리그노셀룰로오스 분해 효소는 셀룰라아제 효소, 헤미셀룰라아제 효소, 또는 이의 조합으로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 셀룰라아제 효소는 *cbh1*, *cbh2*, *eg11*, *eg12*, *eg13*, *eg14*, *eg15*, *eg16*, *bg11*, *bg12*, *swo1*, *cip1* 및 *cip2*로 구성된 군으로부터 선택된다. 다른 구현예에서, 헤미셀룰라아제 효소는 *xyn1*, *xyn2*, *xyn3*, *xyn4*, *bx11*, *abf1*, *abf2*, *abf3*, *axe1*, *axe2*, *axe3*, *man1*, *ag11*, *ag12*, *ag13* 및 *g1r1*로 구성된 군으로부터 선택된다.

[0029] 방법의 또 다른 구현예에서, 단계 (i)은 내인성 POI를 암호화하는 도입된 폴리뉴클레오티드 구성체를 더 포함한다. 일부 구현예에서, 이종성 POI를 암호화하는 폴리뉴클레오티드 구성체는 셀룰로오스 유도성 유전자 프로모터의 제어 하에 발현된다. 방법의 일부 다른 구현예에서, 단계 (i) 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *d1d1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pk1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 방법은 서열 번호 48의 야생형 자일라나아제 조절인자 1(*Xyr1*) 단백질 또는 서열 번호 46의 변이체 자일라나아제 조절인자 1(*Xyr1*) 단백질을 암호화하는 폴리뉴클레오티드를 발현하는 유전자 변형을 더 포함한다. 일부 다른 구현예에서, 방법은 내인성 탄소 이화대사산물 억제인자 1(*Cre1*) 단백질 또는 *Ace1* 단백질을 암호화하는 유전자의 발현을 감소시키거나 방지하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 *Ace2* 단백질 발현을 포함하는 유전자 변형을 더 포함한다.

[0030] 일부 다른 구현예에서, 본 발명은 유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 이종성 관심 단백질의 생산 방법을 대상으로 하는데, 방법은 (i) 5'에서 3' 방향으로 (a) 구성성 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드 구성체를 진균 세포 내로 도입하는 단계로서, 이 때, 제2 핵산 서열은 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 *Ace3* 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하는 것인, 단계, 및 (ii) 진균 세포 성장 및 단백질 생산에 적합한 조건 하에서 단계 (i)의 세포를 발효시키는 단계로서, 이 때, 그러한 적합한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함한다. 따라서, 방법의 일부 구현예에서, 진균 세포는 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드 구성체를 포함하는데, 이 때, 구성체는 단계 (i) 이전에, 단계 (i) 중에, 또는 단계 (i) 이후에 진균 세포 내로 도입된다. 또 다른 구현예에서, 폴리뉴클레오티드 구성체는 제3 핵산 서열 3'을 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이 때, 제3 핵산 서열은 천연 *ace3* 종결자 서열을 포함한다. 방법의 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈 내로 통합된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 텔로미어 부위 내로 통합된다. 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 게놈의 글루코아밀라아제(*gla1*) 유전자와 내로 통합된다.

[0031] 방법의 일부 다른 구현예에서, 폴리뉴클레오티드 구성체는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 또 다른 구현예에서, 암호화된 *Ace3* 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 95% 서열 동일성을 포함하는 아미노산 서열을 포함한다. 다른 구현예에서, 암호화된 *Ace3* 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14의 아미노산 서열을 포함한다. 다른 구현예에서, 이종성 POI를 암호화하는 폴리뉴클레오티드 구성체는 셀룰로오스 유도성 유전자 프로모터의 제어 하에 발현된다. 일부 다른 구현예에서, 이종성 POI는 α -아밀라아제, 알칼리성 α -아밀라아제, β -아밀라아제, 셀룰라아제, 텍스트라나아제, α -글루코시다아제, α -갈락토시다아제, 글루코아밀라아제, 헤미셀룰라아제, 펜토사나아제, 자일라나아제, 인버타아제, 락타아제, 나린지나아제, 펩티나아제, 폴루라나아제, 산성 프로테아제, 알칼리성 프로테아제, 브로멜라인, 중성 프로테아제, 파파인, 펩신, 펩티다아제, 레닌, 레닌,

키모신, 서브틸리신, 서몰리신, 아스파르트산 프로테이나제, 트립신, 리파아제, 에스테라아제, 포스포리파아제, 포스파타아제, 피타아제, 아미다아제, 이미노아실라아제, 글루타미나아제, 리소자임, 페니실린 아실라아제; 이소메라제, 옥시도리덕타아제, 카탈라아제, 클로로페록시다아제, 글루코오스 옥시다아제, 하이드록시스테로이드하이드로게나아제, 페록시다아제, 리아제, 아스파르트산 β -데카르복실라아제, 푸마라아제, 히스타다아제, 트랜스페라아제, 리가아제, 이미노펩티다아제, 카르복시펩티다아제, 키티나아제, 큐티나아제, 테옥시리보뉴클레아제, α -갈락토시다아제, β -갈락토시다아제, β -글루코시다아제, 락카아제, 만노시다아제, 뮤타나아제, 폴리페놀 옥시다아제, 리보뉴클레아제 및 트랜스글루타미나아제로 구성된 군으로부터 선택된다.

[0032] 방법의 다른 구현예에서, 단계 (i) 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 방법은 서열 번호 48의 야생형 자일라나아제 조절인자 1(Xyr1) 단백질 또는 서열 번호 46의 변이체 자일라나아제 조절인자 1(Xyr1) 단백질을 암호화하는 폴리뉴클레오티드를 발현하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 내인성 탄소 이화 대사산물 억제인자 1(Cre1) 단백질 또는 Ace1 단백질을 암호화하는 유전자의 발현을 감소시키거나 방지하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 Ace2 단백질 발현을 포함하는 유전자 변형을 더 포함한다.

[0033] 또 다른 구현예에서, 본 발명은 유도 기질의 부존재 하의 트리코더마 종 진균 세포 내에서 이종성 관심 단백질의 생산 방법에 관한 것으로, 방법은 (i) 5'에서 3' 방향으로 (a) 구성성 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드 구성체를 진균 세포 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 12에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 것인, 단계, 및 (ii) 진균 세포 성장 및 단백질 생산에 적합한 조건 하에서 단계 (i)의 세포를 발효시키는 단계로서, 이때, 그러한 적합한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함한다. 특정 구현예에서, 진균 세포는 이종성 POI를 암호화하는 도입된 폴리뉴클레오티드 구성체를 포함하는데, 이때, 구성체는 단계 (i) 이전에, 단계 (i) 중에, 또는 단계 (i) 이후에 진균 세포 내로 도입된다. 다른 구현예에서, 폴리뉴클레오티드 구성체는 제3 핵산 서열 3'을 포함하고, 제2 핵산에 작동 가능하게 연결되는데, 이때, 제3 핵산 서열은 천연 *ace3* 종결자 서열을 포함한다. 방법의 다른 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 계놈 내로 통합된다. 일부 구현예에서, 폴리뉴클레오티드 구성체는 진균 세포 계놈의 글루코아밀라아제(*gla1*) 유전자와 내로 통합된다. 또 다른 구현예에서, 폴리뉴클레오티드 구성체는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 일부 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14에 대해 95% 서열 동일성을 포함하는 아미노산 서열을 포함한다. 또 다른 구현예에서, 암호화된 Ace3 단백질은 서열 번호 6, 서열 번호 12 또는 서열 번호 14의 아미노산 서열을 포함한다. 방법의 또 다른 구현예에서, 이종성 POI를 암호화하는 폴리뉴클레오티드 구성체는 셀룰로오스 유도성 유전자 프로모터의 제어 하에 발현된다. 일부 구현예에서, 이종성 POI는 α -아밀라아제, 알칼리성 α -아밀라아제, β -아밀라아제, 셀룰라아제, 텍스트라나아제, α -글루코시다아제, α -갈락토시다아제, 글루코아밀라아제, 헤미셀룰라아제, 펜토사나아제, 자일라나아제, 인버타아제, 락타아제, 나린지나아제, 펩티나아제, 폴루라나아제, 산성 프로테아제, 알칼리성 프로테아제, 브로멜라인, 중성 프로테아제, 파파인, 펩신, 펩티다아제, 레닛, 레닌, 키모신, 서브틸리신, 서몰리신, 아스파르트산 프로테이나제, 트립신, 리파아제, 에스테라아제, 포스포리파아제, 포스파타아제, 피타아제, 아미다아제, 이미노아실라아제, 글루타미나아제, 리소자임, 페니실린 아실라아제; 이소메라제, 옥시도리덕타아제, 카탈라아제, 클로로페록시다아제, 글루코오스 옥시다아제, 하이드록시스테로이드하이드로게나아제, 페록시다아제, 리아제, 아스파르트산 β -데카르복실라아제, 푸마라아제, 히스타다아제, 트랜스페라아제, 리가아제, 이미노펩티다아제, 카르복시펩티다아제, 키티나아제, 큐티나아제, 테옥시리보뉴클레아제, α -갈락토시다아제, β -갈락토시다아제, β -글루코시다아제, 락카아제, 만노시다아제, 뮤타나아제, 폴리페놀 옥시다아제, 리보뉴클레아제 및 트랜스글루타미나아제로 구성된 군으로부터 선택된다.

[0034] 방법의 다른 구현예에서, 단계 (i) 프로모터는 *rev3* 프로모터(서열 번호 15), *bx1* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택된다. 또 다른 구현예에서, 방법은 서열 번호 48의 야생형 자일라

나아제 조절인자 1(Xyr1) 단백질 또는 서열 번호 46의 변이체 자일라나아제 조절인자 1(Xyr1) 단백질을 암호화하는 폴리뉴클레오티드를 발현하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 내인성 탄소 이화 대사산물 억제인자 1(Cre1) 단백질 또는 Ace1 단백질을 암호화하는 유전자의 발현을 감소시키거나 방지하는 유전자 변형을 더 포함한다. 또 다른 구현예에서, 방법은 Ace2 단백질 발현을 포함하는 유전자 변형을 더 포함한다.

[0035] 일부 다른 구현예에서, 본 발명은 유도 기질의 부존재 하에서의 내인성 단백질의 증가된 생산을 위하여 트리코더마 레에세이 주를 유전자 변형하는 방법에 관한 것으로, 방법은 (i) 서열 번호 3의 Ace3-S 단백질, 서열 번호 8의 Ace3-SC 단백질 또는 서열 번호 10의 Ace3-LC 단백질을 암호화하는 *ace3* 유전자의 게놈 카페를 포함하는 T. 레에세이 주를 스크리닝하고 확인하는 단계로서, 이때, 확인된 T. 레에세이 주는 서열 번호 6의 Ace3-L 단백질, 서열 번호 14의 Ace3-LN 단백질 또는 서열 번호 12의 Ace3-EL 단백질을 암호화하는 *ace3* 유전자의 게놈 카페를 포함하지 않는 것인, 단계, (ii) 5'에서 3' 방향으로 (a) 프로모터를 포함하는 제1 핵산 서열 및 (b) 제1 핵산 서열에 작동 가능하게 연결된 제2 핵산 서열을 포함하는 폴리뉴클레오티드 구성체를 단계 (i)에서 확인된 T. 레에세이 주 내로 도입하는 단계로서, 이때, 제2 핵산 서열은 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하고, 마지막 네 개의 C 말단 아미노산으로서 "Lys-Ala-Ser-Asp"를 포함하거나 제2 핵산 서열은 서열 번호 12에 대해 약 90% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 것인, 단계, 및 (iii) 진균 세포 성장 및 단백질 생산에 적합한 조건 하에서 단계 (ii)의 세포를 발효시키는 단계로서, 이때, 그러한 적합한 성장 조건은 유도 기질을 포함하지 않는 것인, 단계를 포함한다.

[0036] 일부 다른 구현예에서, 본 발명은 모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포에 관한 것으로, 변이체 세포는 대안적인 프로모터에 의해 교체된 천연 *ace3* 유전자 프로모터를 포함하고, 이때, 변이체 세포는 모 세포에 비해 유도 기질의 부존재 하에 증가된 양의 관심 단백질(POI)을 생산하며, 이때, 변이체 세포와 모 세포는 유사한 조건 하에서 배양된다. 특정 구현예에서, 대안적인 프로모터는 트리코더마 레에세이 프로모터이다. 또 다른 구현예에서, 대안적인 프로모터는 *rev3* 프로모터(서열 번호 15), *bxi* 프로모터(서열 번호 16), *tk11* 프로모터(서열 번호 17), *PID104295* 프로모터(서열 번호 18), *dld1* 프로모터(서열 번호 19), *xyn4* 프로모터(서열 번호 20), *PID72526* 프로모터(서열 번호 21), *axe1* 프로모터(서열 번호 22), *hxk1* 프로모터(서열 번호 23), *dic1* 프로모터(서열 번호 24), *opt* 프로모터(서열 번호 25), *gut1* 프로모터(서열 번호 26) 및 *pki1* 프로모터(서열 번호 27)로 구성된 군으로부터 선택된 프로모터이다.

도면의 간단한 설명

[0037] 도 1은 Ace3 단백질 암호화 영역에 대한 개략도를 보여 준다. 더욱 상세하게는, 도 1a는 게놈의 동일한 DNA 서열에 대해 정렬시킨 T. 레에세이 주 QM6a 및 RUT-C30(도 1a)의 주석을 기반으로 한 Ace3 단백질 암호화 영역에 대한 개략도를 보여 준다. 예상되는 액손과 인트론을 각각 화살표와 점선으로 나타내었다. 점선으로 표시된 수직선은 RUT-C30 게놈의 네센스 돌연변이를 나타낸다. 서열 번호 2의 클로닝된 *ace3*-S (짧은) 오픈 리딩 프레임 및 서열 번호 5의 클로닝된 *ace3*-L (긴) 오픈 리딩 프레임을 본 출원에 기재된 바와 같이 스크리닝하고 시험하였다. 도 1b에는 T. 레에세이 주 RUT-C30으로부터의 Ace3-L 단백질(서열 번호 6), T. 레에세이 주 QM6a로부터의 Ace3-S 단백질(서열 번호 3), Ace3-SC 단백질(서열 번호 8), Ace3-LN 단백질(서열 번호 14), Ace3-LC 단백질(서열 번호 10) 및 Ace3-EL 단백질(서열 번호 12)의 아미노산 정렬을 기재하였다.

도 2는 Ace3-발현 벡터 pYL1(도 2a), pYL2(도 2b), pYL3(도 2c) 및 pYL4(도 2d)의 개략도를 보여 준다.

도 3은 T. 레에세이 모 세포 및 변이체 세포 상청액의 폴리아크릴아미드 젤 전기영동(SDS-PAGE)을 보여 주며, 이때, 모 세포와 변이체 세포는 20% 락토오스(lac) 또는 20% 글루코오스(glu)를 함유하는 srMTP 중의 합성 배지에서 성장시켰다. 동일한 부피의 배양 상청액을 각 레인에 로딩하였다. M은 분자량 마커이고, 모 T. 레에세이 균주는 대조군 균주로 기능하였다.

도 4는 T. 레에세이 모 세포 및 변이체 세포 상청액의 SDS-PAGE를 보여 주며, 이때, 모 세포와 변이체 세포는 진탕 플라스크에서 1.5% 글루코오스/소포로오스(sop) 또는 1.5% 글루코오스(glu)가 첨가된 합성 배지 중에서 성장시켰다. 동일한 부피의 배양 상청액을 각 레인에 로딩하였다. 배양 상청액의 총 단백질 농도는 각각의 해당하는 레인 하단에 열거하였다. M은 분자량 마커이고, KD는 킬로달톤이다.

도 5는 2L 발효기에서 탄소 공급원으로 글루코오스/소포로오스(sop) 또는 글루코오스(glu)가 첨가된 합성 배지 중에서 성장시킨 T. 레에세이 모 세포 및 변이체 세포의 SDS-PAGE를 보여 준다. M은 분자량 마커이고, 영(0)은 종배양(seed culture) 상청액이다.

도 6은 *ace3* 유전자좌에 천연 프로모터의 5' 영역 상류를 포함하는 DNA 단편, *loxP*가 측면에 있는 하이그로마이신 B-저항성(선택 가능한 마커) 카세트 및 *ace3* ORF의 5' 말단에 작동 가능하게 융합된(연결된) 관심 프로모터를 포함하는 DNA 단편을 융합시켜 제조된 프로모터 교체 구성체의 개략도를 보여 준다.

도 7은 *dic1* 프로모터를 포함하는 Ace3-발현 벡터 pYL8의 개략도를 보여 준다.

도 8은 *T. 레에세이* 모 세포 및 그것의 변형된 (딸) 세포 상청액의 SDS-PAGE를 보여 주며, 이때, 모 균주와 변형된 균주는 진탕 플라스크에서 2.5% 글루코오스/소포로오스("Sop", 유도 조건) 또는 2.5% 글루코오스("Glu", 비유도 조건)가 첨가된 합성 배지에서 성장시켰다. 동일한 부피의 배양 상청액을 각 레인에 로딩하였다. M은 분자량 마커이고, KD는 킬로달톤이다.

도 9는 소규모(2 L) 발효의 단백질 생산을 보여 준다. *T. 레에세이* 모 균주 및 딸 균주 "LT83"은 탄소 공급원으로 글루코오스/소포로오스(Sop, 유도 조건) 또는 글루코오스(Glu, 비유도 조건)가 첨가된 합성 배지에서 성장시켰다. 발효 종료 시 글루코오스/소포로오스 상의 모 균주에 의해 생산된 총 단백질을 임의로 100으로 설정하였고, 각 시점에서 각각의 균주에 의해 생산된 상대적인 단백질량을 그래프로 나타내었다.

도 10은 *T. 레에세이* 모 세포 및 그것의 변형된 (딸) 세포 상청액의 SDS-PAGE를 보여 주며, 이때, 모 균주와 변형된 균주는 진탕 플라스크에서 2.5% 글루코오스/소포로오스("Sop", 유도 조건) 또는 2.5% 글루코오스("Glu", 비유도 조건)가 첨가된 합성 배지에서 성장시켰다. 동일한 부피의 배양 상청액을 각 레인에 로딩하였다. M은 분자량 마커이고, KD는 킬로달톤이다.

도 11은 *ace3* 유전자좌에 대한 개략적인 이미지를 보여 준다. *ace3* 유전자좌의 5' 말단(N 말단)의 화살표는 cDNA 서열(화살표 1), RutC-30 주석형(화살표 2) 및 QM6a 주석형(화살표 3)이 제시한 형태에서의 상이한 전사 개시 부위를 나타낸다. *ace3* 유전자좌의 3' 말단(C 말단)의 화살표는 RutC-30 주석형(화살표 4) 및 QM6a 주석형(화살표 5)에서의 상이한 정지 코돈을 나타낸다.

도 12는 클로닝된 상이한 *ace3* 형태에 대한 개략적인 이미지를 보여 준다. 따라서, 도 12에 제시되고 실시예 6에 기술된 바와 같이, 다음의 *ace3* 형태가 클로닝되고 시험되었다: 1,713bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4를 포함하는 서열 번호 1의 "***ace3-S***", 1,713bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함하는 서열 번호 7의 "***ace3-SC***", 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함하는 서열 번호 4의 "***ace3-L***", 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4를 포함하는 서열 번호 9의 "***ace3-LC***", 61 bp 엑손 1, 142 bp 인트론 1, 332 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함하는 서열 번호 11의 "***ace3-EL***", 및 258 bp 엑손 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함하는 서열 번호 13의 "***ace3-LN***".

도 13은 ***ace3-SC*** 유전자 형태(서열 번호 7; 1,713bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4 포함)의 핵산 서열 및 암호화된 Ace3-SC 단백질 서열(서열 번호 8)을 보여 준다. *ace3-SC* 유전자 형태에 대해 도 13에 제시된 바와 같이, **굵은 검은색 글자체**로 표시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 14는 ***ace3-S*** 유전자 형태(서열 번호 1; 1,713bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4 포함)의 핵산 서열 및 암호화된 Ace3-S 단백질 서열(서열 번호 3)을 보여 준다. *ace3-S* 유전자 형태에 대해 도 14에 제시된 바와 같이, **굵은 검은색 글자체**로 표시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 15는 ***ace3-L*** 유전자 형태(서열 번호 4; 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4 포함)의 핵산 서열 및 암호화된 Ace3-L 단백질 서열(서열 번호 6)을 보여 준다. *ace3-L* 유전자 형태에 대해 도 15에 제시된 바와 같이, **굵은 검은색 글자체**로 표시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 16은 ***ace3-LC*** 유전자 형태(서열 번호 9; 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4 포함)의 핵산 서열 및 암호화된 Ace3-LC 단백질 서열(서열 번호 10)을 보여 준다. *ace3-LC* 유전자 형태에 대해 도 16에 제시된 바와 같이, **굵은 검은색 글자체**로 표시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 17은 ***ace3-EL*** 유전자 형태(서열 번호 11; 61 bp 엑손 1, 142 bp 인트론 1, 332 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4 포함)의 핵산 서열 및 암호화된 Ace3-EL 단백질 서열(서열 번호 12)을 보여 준다. *ace3-EL* 유전자 형태에 대해 도 17에 제시된 바와 같이, **굵은 검은색 글자체**로 표

시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 18은 *ace3-LN* 유전자 형태(서열 번호 13; 258 bp 엑손 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4 포함)의 핵산 서열 및 암호화된 *Ace3-LN* 단백질 서열(서열 번호 14)을 보여 준다. *ace3-LN* 유전자 형태에 대해 도 18에 제시된 바와 같이, **굵은 검은색 글자체**로 표시된 뉴클레오티드는 인트론 서열을 나타낸다.

도 19는 *gla1* 유전자좌에서 *ace3* 형태들의 통합을 위해 설계된 DNA 단편의 배열에 대한 개략도를 보여 준다.

도 20은 벡터 pMCM3282의 개략도를 보여 준다.

도 21은 벡터 pCHL760의 개략도를 보여 준다.

도 22는 벡터 pCHL761의 개략도를 보여 준다.

도 23은 유도("Glu/Sop") 및 비유도("Glu") 배양 조건 하에서의 *T. 레에세이* 모 세포(도 23, 세포 ID 1275.8.1) 및 변이체 *T. 레에세이*(딸) 세포(도 23, 세포 ID 번호: 2218, 2219, 2220, 2222 및 2223)의 액침 배양(즉, 진탕 플라스크)에 의해 생산된 분비 단백질의 SDS-PAGE를 보여 준다.

[생물학적 서열의 간단한 설명]

다음의 서열은 37 C.F.R. §§ 1.821-1.825("뉴클레오티드 서열 및/또는 아미노산 서열의 기재를 포함하는 특허 출원의 요건--서열 규정[Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures--the Sequence Rules")를 준수하며, WIP 표준 ST.25 (2009) 및 유럽 특허 조약(European Patent Convention, EPC) 및 특허 협력 조약(Patent Cooperation Treaty, PCT) 시행규칙 5.2 및 49.5(a-bis), 및 시행 세칙의 제208조 및 부칙 C의 서열 목록 요건과 일관된다. 뉴클레오티드 및 아미노산 서열 데이터에 사용되는 부호 및 포맷은 37 C.F.R. § 1.822에 개시된 규정을 따른다.

서열 번호 1은 서열 번호 3의 *Ace3-S* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 야생형 균주 QM6a 핵산 서열이다.

서열 번호 2는 서열 번호 3의 *Ace3-S* 단백질을 암호화하는 핵산 서열 오픈 리딩 프레임(ORF)이다.

서열 번호 3은 이하에서 "Ace3-S"로 표기된, *트리코더마 레에세이*(균주 QM6a) *Ace3* 단백질의 아미노산 서열이다.

서열 번호 4는 서열 번호 6의 *Ace3-L* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 균주 Rut-C30 핵산 서열이다.

서열 번호 5는 서열 번호 6의 *Ace3-L* 단백질을 암호화하는 핵산 서열 ORF이다.

서열 번호 6은 이하에서 "Ace3-L"로 표기된, *트리코더마 레에세이*(균주 Rut-C30) *Ace3* 단백질의 아미노산 서열이다.

서열 번호 7은 서열 번호 8의 *Ace3-SC* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 핵산 서열이다.

서열 번호 8은 이하에서 "Ace3-SC"로 표기된, *트리코더마 레에세이* *Ace3* 단백질의 아미노산 서열이다.

서열 번호 9는 서열 번호 10의 *Ace3-LC* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 핵산 서열이다.

서열 번호 10은 이하에서 "Ace3-LC"로 표기된, *트리코더마 레에세이* *Ace3* 단백질의 아미노산 서열이다.

서열 번호 11은 서열 번호 12의 *Ace3-EL* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 핵산 서열이다.

서열 번호 12는 이하에서 "Ace3-EL"로 표기된, *트리코더마 레에세이* *Ace3* 단백질의 아미노산 서열이다.

서열 번호 13은 서열 번호 14의 *Ace3-LN* 단백질을 암호화하는 유전자를 포함하는 *트리코더마 레에세이* 핵산 서열이다.

서열 번호 14는 이하에서 "Ace3-LN"으로 표기된, *트리코더마 레에세이* *Ace3* 단백질의 아미노산 서열이다.

서열 번호 15는 *rev3* 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 16은 β -xy1 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 17은 tk11 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 18은 PID104295 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 19는 dld1 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 20은 xyn4 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 21은 PID72526 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 22는 axe3 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 23은 hxk1 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 24는 dic1 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 25는 opt 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 26은 gut1 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 27은 pk11 프로모터 서열을 포함하는 핵산 서열이다.

서열 번호 28은 프라이머 TP13의 핵산 서열이다.

서열 번호 29는 프라이머 TP14의 핵산 서열이다.

서열 번호 30은 프라이머 TP15의 핵산 서열이다.

서열 번호 31은 프라이머 TP16의 핵산 서열이다.

서열 번호 32는 프라이머 TP17의 핵산 서열이다.

서열 번호 33은 프라이머 TP18의 핵산 서열이다.

서열 번호 34는 프라이머 TP19의 핵산 서열이다.

서열 번호 35는 프라이머 TP20의 핵산 서열이다.

서열 번호 36은 프라이머 TP21의 핵산 서열이다.

서열 번호 37은 프라이머 TP22의 핵산 서열이다.

서열 번호 38은 프라이머 TP23의 핵산 서열이다.

서열 번호 39는 프라이머 TP24의 핵산 서열이다.

서열 번호 40은 프라이머 TP25의 핵산 서열이다.

서열 번호 41은 프라이머 TP26의 핵산 서열이다.

서열 번호 42는 플라스미드 pYL1의 핵산 서열이다.

서열 번호 43은 플라스미드 pYL2의 핵산 서열이다.

서열 번호 44는 플라스미드 pYL3의 핵산 서열이다.

서열 번호 45는 플라스미드 pYL4의 핵산 서열이다.

서열 번호 46은 T. 레에세이 xyr1(A824V) 변이체 단백질의 아미노산 서열이다.

서열 번호 47은 T. 레에세이 Ace2 단백질의 아미노산 서열이다.

서열 번호 48은 T. 레에세이 야생형 xyr1 단백질의 아미노산 서열이다.

서열 번호 49는 프라이머 핵산 서열이다.

서열 번호 50은 프라이머 핵산 서열이다.

서열 번호 51은 프라이머 핵산 서열이다.

서열 번호 52는 프라이머 핵산 서열이다.

서열 번호 53은 프라이머 핵산 서열이다.

서열 번호 54는 프라이머 핵산 서열이다.

서열 번호 55는 프라이머 핵산 서열이다.

서열 번호 56은 프라이머 핵산 서열이다.

서열 번호 57은 프라이머 핵산 서열이다.

서열 번호 58은 프라이머 핵산 서열이다.

서열 번호 59는 프라이머 핵산 서열이다.

서열 번호 60은 프라이머 핵산 서열이다.

서열 번호 61은 프라이머 핵산 서열이다.

서열 번호 62는 프라이머 핵산 서열이다.

서열 번호 63은 프라이머 핵산 서열이다.

서열 번호 64는 프라이머 핵산 서열이다.

서열 번호 65는 프라이머 핵산 서열이다.

서열 번호 66은 프라이머 핵산 서열이다.

서열 번호 67은 프라이머 핵산 서열이다.

서열 번호 68은 프라이머 핵산 서열이다.

서열 번호 69는 프라이머 핵산 서열이다.

서열 번호 70은 프라이머 핵산 서열이다.

서열 번호 71은 프라이머 핵산 서열이다.

서열 번호 72는 프라이머 핵산 서열이다.

서열 번호 73은 프라이머 핵산 서열이다.

서열 번호 74는 프라이머 핵산 서열이다.

서열 번호 75는 프라이머 핵산 서열이다.

서열 번호 76은 프라이머 핵산 서열이다.

서열 번호 77은 프라이머 핵산 서열이다.

서열 번호 78은 프라이머 핵산 서열이다.

서열 번호 79는 프라이머 핵산 서열이다.

서열 번호 80은 프라이머 핵산 서열이다.

서열 번호 81은 프라이머 핵산 서열이다.

서열 번호 82는 인공 핵산 서열이다.

서열 번호 83은 인공 핵산 서열이다.

서열 번호 84는 인공 핵산 서열이다.

서열 번호 85는 인공 핵산 서열이다.

서열 번호 86은 인공 핵산 서열이다.

서열 번호 87은 인공 핵산 서열이다.

서열 번호 88은 인공 핵산 서열이다.

서열 번호 89는 인공 핵산 서열이다.

서열 번호 90은 인공 핵산 서열이다.

서열 번호 91은 인공 핵산 서열이다.

서열 번호 92는 프라이머 핵산 서열이다.

서열 번호 93은 프라이머 핵산 서열이다.

서열 번호 94는 프라이머 핵산 서열이다.

서열 번호 95는 프라이머 핵산 서열이다.

서열 번호 96은 프라이머 핵산 서열이다.

서열 번호 97은 프라이머 핵산 서열이다.

서열 번호 98은 서열 번호 12의 Ace3-EL 단백질의 N 말단 서열의 45개 아미노산의 단편을 포함하는 아미노산 서열이다.

서열 번호 99는 Ace3-SC 단백질을 암호화하는 핵산 서열 ORF이다.

서열 번호 100은 Ace3-LC 단백질을 암호화하는 핵산 서열 ORF이다.

서열 번호 101은 Ace3-EL 단백질을 암호화하는 핵산 서열 ORF이다.

서열 번호 102는 Ace3-LN 단백질을 암호화하는 핵산 서열 ORF이다.

발명을 실시하기 위한 구체적인 내용

I. 개요

[0038] 본 발명의 일부 구현예는 변이체 사상 진균 세포, 이의 조성물 및 이의 1종 이상의 관심 단백질의 증가된 생산을 위한 방법을 대상으로 한다. 더욱 상세하게는, 본 발명의 일부 구현예는 유도성 먹이(즉, 락토오스, 소포로오스, 젠티오비오스, 셀룰로오스 등과 같은 유도 기질)의 부존재 하에 1종 이상의 관심 단백질을 생산할 수 있는 변이체 사상 진균 세포를 대상으로 한다. 따라서, 본 발명의 일부 구현예는 모 사상 진균 세포로부터 유래된 변이체 사상 진균 (숙주) 세포를 대상으로 하며, 이때, 변이체 숙주 세포는 유도 기질의 부존재 하에 (관심 단백질을 암호화하는) 관심 유전자의 발현을 가능하게 하는 유전자 변형을 포함한다. (관심 단백질을 암호화하는) 관심 유전자는 내인성 사상 진균 세포 유전자(예컨대, cbh1, chb2, xyn1, xyn2, xyn3, egl1, egl2, egl3, bg11, bg12 등) 또는 이 사상 진균 세포에 대해 이종성인 유전자일 수 있다.

[0040] 따라서, 일부 다른 구현예에서, 본 발명의 변이체 진균 숙주 세포는 Ace3-L 단백질(서열 번호 6), Ace3-EL 단백질(서열 번호 12) 및 Ace3-LN 단백질(서열 번호 14)로 구성된 군으로부터 선택된 Ace3 단백질을 암호화하는 "셀룰라아제 발현 활성화 인자(activator of cellulase expression) 3"(ace3) 유전자의 발현을 증가시키는 유전자 변형을 포함한다. 다른 구현예에서, 본 발명의 변이체 진균 숙주 세포는 Ace3-L 단백질(서열 번호 6), Ace3-EL 단백질(서열 번호 12) 및 Ace3-LN 단백질(서열 번호 14)로 구성된 군으로부터 선택된 Ace3 단백질을 암호화하는 폴리뉴클레오티드 오픈 리딩 프레임(ORF)의 발현을 증가시키는 유전자 변형을 포함한다. 따라서, 일부 구현예에서, 본 발명의 변이체 진균 숙주 세포는 Ace3-L 단백질(서열 번호 6), Ace3-EL 단백질(서열 번호 12) 및 Ace3-LN 단백질(서열 번호 14)로 구성된 군으로부터 선택된 Ace3 단백질에 대해 약 90% 내지 약 99% 서열 동일성을 포함하는 Ace3 단백질을 암호화하는 Ace 유전자 또는 이의 ORF의 발현/생산을 증가시키는 유전자 변형을 포함한다.

[0041] 일부 구현예에서, Ace3 단백질(즉, Ace3-L, Ace3-EL 또는 Ace3-LN 단백질)의 발현을 증가시키는 유전자 변형은 사상 진균 숙주 세포의 계놈(염색체) 내로 통합된 ace3 발현 카세트이다. 다른 구현예에서, 사상 진균 세포에서 Ace3 단백질의 발현을 증가시키는 유전자 변형은 ace3 발현 카세트(즉, Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는)를 포함하는 에피솜 유지된 플라스미드 구성체이다. 다른 구현예에서, 사상 진균 세포에서 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 ace3 유전자의 발현을 증가시키는 유전자 변형은 텔로미어 부위에 통합된 텔로미어 벡터/플라스미드이다. 일부 구현예에서, 이러한 발현 카세트 또는 플라스미드는 두 카피 이상 존

재한다. 일부 다른 구현예에서, *ace3* 유전자, 또는 *ace3* ORF는 이종성 프로모터에 작동 가능하게 연결된다. 다른 구현예에서, 사상 진균 세포에서 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 *ace3* 유전자(또는 이의 ORF)의 발현을 증가시키는 유전자 변형은 천연 *ace3* 프로모터(즉, *ace3* 유전자와 천연적으로 관련이 있는 *ace3* 프로모터 영역)의 변형이며, 이 변형은 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 *ace3* 유전자의 발현을 바꾼다.

[0042] II. 정의

[0043] 본 발명의 조성물 및 방법을 더욱 상세하게 기술하기 전에, 다음의 용어 및 어구가 정의된다. 정의되지 않은 용어는 당해 분야에서 사용되고 당업자에게 공지된 통상적인 의미를 따라야 한다.

[0044] 본 명세서에 인용된 모든 간행물 및 특허는 본 설명에 참조로 포함된다.

[0045] 일정 범위의 값이 제공되는 경우, 문맥이 명백하게 달리 지시하지 않는 한, 그 범위의 상한 및 하한, 그리고 그 언급된 범위 내의 임의의 기타 언급된 값 또는 개별 값의 사이에 있는 각각의 값은 하한 단위의 소수점 첫째 자리까지, 본 발명의 조성물 및 방법에 포함된다. 또한, 이러한 더 작은 범위의 상한들 및 하한들은, 이러한 더 작은 범위 내에 독립적으로 포함될 수 있으며, 언급된 범위 내의 임의의 구체적으로 제외된 한계치를 조건으로, 본 발명의 조성물 및 방법에 포함된다. 언급된 범위가 한계치 중 하나 또는 둘 다를 포함하는 경우, 그렇게 포함되는 한계치 중 하나 또는 둘 다를 제외하는 범위들 또한, 본 발명의 조성물 및 방법에 포함된다.

[0046] 일부 범위는 용어 "약"이 선행되는 수치 값으로 본 설명에 제시된다. 용어 "약"은 뒤에 오는 정확한 수뿐만 아니라, 이 용어 뒤에 오는 수에 근사한 수 또는 대략 이 용어 뒤에 오는 수를 문자 그대로 뒷받침하기 위하여 본 설명에서 사용된다. 소정의 수가 구체적으로 인용된 수에 근사한 수 또는 대략 구체적으로 인용된 수인지를 결정하는 데 있어서, 근사한 수 또는 근사값의 인용되지 않은 수는, 그것이 제시된 문맥에서, 구체적으로 인용된 수의 실질적인 등가를 제공하는 수일 수 있다. 예를 들어, 수치와 관련하여, 용어 "약"은, 이 용어가 문맥에서 달리 구체적으로 정의되지 않는 한, 이 수치의 -10% 내지 +10%의 범위를 지칭한다. 다른 예에서, 어구 "약 6의 pH 값"은, 이 pH 값이 구체적으로 달리 정의되지 않는 한, 5.4 내지 6.6의 pH 값을 지칭한다.

[0047] 본 설명에 제공된 표제는 본 명세서를 전체로서 참조하여 가질 수 있는 본 발명의 조성물 및 방법의 다양한 양태 또는 구현예의 제한이 아니다. 따라서, 바로 아래에 정의된 용어는 본 명세서를 전체로서 참조하여 더욱 완전하게 정의된다.

[0048] 이러한 발명을 실시하기 위한 구체적인 내용에 따라, 다음의 약어 및 정의가 적용된다. 단수형("a", "an" 및 "the")은, 문맥에서 명확하게 달리 지시하지 않는 한, 복수의 지시 대상을 포함한다는 것에 유의한다. 따라서, 예를 들어, "효소"에 대한 언급은 복수형의 그러한 효소를 포함하며, "투여량"에 대한 언급은 당업자에게 공지된 하나 이상의 투여량 및 이의 등가물에 대한 언급을 포함하며, 다른 것도 마찬가지이다.

[0049] 청구범위는 임의의 선택적 요소를 제외하도록 작성될 수 있다는 것 또한, 유의한다. 따라서, 이러한 서술은 청구항 요소의 설명과 관련하여 "오로지", "단지", "제외하는", "포함하지 않는" 등과 같은 배타적인 용어의 사용, 또는 "부정적" 제한의 사용에 대한 선행적 근거로서의 역할을 하기 위한 것이다.

[0050] 본 설명에서 사용되는 바와 같은 "포함하는"이라는 용어는 "포함하는"이라는 용어 앞의 성분(들)을 "포함하지만 이에 한정되는 것은 아님"을 의미함에 또한, 유의한다. "포함하는"이라는 용어 앞의 성분(들)은 요구되거나 의무적이지만, 이 성분(들)을 포함하는 조성물은 다른 비-의무적 또는 선택적 성분(들)을 추가로 포함할 수 있다.

[0051] 본 설명에서 사용되는 바와 같은 "~로 이루어진"이라는 용어는 "~로 이루어진"이라는 용어 앞의 성분(들)을 "포함하고 이에 한정됨"을 의미함에 또한, 유의한다. 따라서 "~으로 이루어진"이라는 용어 앞의 성분(들)은 요구되거나 의무적인 것이며, 다른 성분(들)은 조성물에 전혀 존재하지 않는다.

[0052] 본 개시 내용을 읽을 때 당업자에게 명백한 바와 같이, 본 설명에 기술되고 예시된 각각의 개별적인 구현예는 본 설명에 기술된 본 발명의 조성물 및 방법의 범위 또는 사상으로부터 벗어나지 않고서 다른 몇몇 구현예들 중 임의의 구현예의 특징과 쉽게 분리되거나 조합될 수 있는 개별 구성요소 및 특징을 갖는다. 임의의 언급된 방법은 언급된 사건의 순서로 또는 논리적으로 가능한 임의의 다른 순서로 수행될 수 있다.

[0053] 본 설명에 사용된 바와 같이, 용어 "자낭균 진균 세포"는 진균계의 자낭균 문(Division)의 임의의 유기체를 지칭한다. 자낭균 진균 세포의 예는 폐지조마이코티나 아문(subphylum Pezizomycotina)의 사상 진균, 예컨대 트리코더마 종(*Trichoderma* spp.), 아스퍼길러스 종(*Aspergillus* spp.), 마이셀리오프토라 종(*Myceliophthora*

spp.), 및 페니실리움 종(*Penicillium* spp.)을 포함하나, 이에 한정되지 않는다.

[0054] 본 설명에 사용된 바와 같이, 용어 "사상 진균"은 진균문(Eumycota) 및 난균문(Oomycota)의 아문의 모든 사상 형태를 지칭한다. 예컨대, 사상 진균은 아크레모늄(*Acremonium*), 아스페질러스(*Aspergillus*), 에메리셀라(*Emericella*), 푸사리움(*Fusarium*), 휴미콜라(*Humicola*), 무코르(*Mucor*), 마이셀리오프토라(*Myceliophthora*), 뉴로스포라(*Neurospora*), 페니실리움(*Penicillium*), 스키탈리디움(*Scyphalidium*), 티엘라비아(*Thielavia*), 톨리포클라디움(*Tolypocladium*), 또는 트리코더마(*Trichoderma*) 종을 제한 없이 포함한다. 일부 구현예에서, 사상 진균은 아스페질러스 아쿨레아투스(*Aspergillus aculeatus*), 아스페질러스 아와모리(*Aspergillus awamori*), 아스페질러스 포에티두스(*Aspergillus foetidus*), 아스페질러스 자포니쿠스(*Aspergillus japonicus*), 아스페질러스 니돌란스(*Aspergillus nidulans*), 아스페질러스 니게르(*Aspergillus niger*), 또는 아스페질러스 오리재(*Aspergillus oryzae*)일 수 있다.

[0055] 일부 구현예에서, 사상 진균은 푸사리움 박트리디오이데스(*Fusarium bactridioides*), 푸사리움 세레알리스(*Fusarium cerealis*), 푸사리움 크룩웰렌스(*Fusarium crookwellense*), 푸사리움 컬모룸(*Fusarium culmorum*), 푸사리움 그레미니아룸(*Fusarium graminearum*), 푸사리움 그레미눔(*Fusarium graminum*), 푸사리움 헤테로스포룸(*Fusarium heterosporum*), 푸사리움 네군디(*Fusarium negundi*), 푸사리움 옥시스포룸(*Fusarium oxysporum*), 푸사리움 레티쿨라툼(*Fusarium reticulatum*), 푸사리움 로세움(*Fusarium roseum*), 푸사리움 삼부시눔(*Fusarium sambucinum*), 푸사리움 사르코크룸(*Fusarium sarcochroum*), 푸사리움 스포로트리키오이데스(*Fusarium sporotrichioides*), 푸사리움 설푸레움(*Fusarium sulphureum*), 푸사리움 토룰로숨(*Fusarium torulosum*), 푸사리움 트리코테시오이데스(*Fusarium trichothecioides*), 또는 푸사리움 베네나툼(*Fusarium venenatum*)이다. 일부 구현예에서, 사상 진균은 휴미콜라 인솔렌스(*Humicola insolens*), 휴미콜라 라누지노사(*Humicola lanuginosa*), 뮤코어 미에헤이(*Mucor miehei*), 마이셀리오프토라 씨모필라(*Myceliophthora thermophila*), 뉴로스포라 크라사(*Neurospora crassa*), 스키탈리디움 씨모필룸(*Scyphalidium thermophilum*), 또는 티엘라비아 테레스트리스(*Thielavia terrestris*)이다.

[0056] 일부 구현예에서, 사상 진균은 트리코더마 하르지아눔(*Trichoderma harzianum*), 트리코더마 코닌기이(*Trichoderma koningii*), 트리코더마 롱기브라키아툼(*Trichoderma longibrachiatum*), 트리코더마 레에세이(*Trichoderma reesei*) 또는 트리코더마 비리데(*Trichoderma viride*)이다. 일부 구현예에서, 사상 진균은 트리코더마 레에세이 주 "Rut-C30"으로부터 유래된 *T. reesei* 세포로, 이는 아메리칸 타입 컬쳐 콜렉션(American Type Culture Collection)으로부터 트리코더마 레에세이 ATCC 수탁번호 56765로서 입수 가능하다. 일부 구현예에서, 사상 진균은 트리코더마 레에세이 주 "RL-P37"로부터 유래된 *T. reesei* 세포로, 이는 미국 농무부 북부 지역 연구소의 컬쳐 콜렉션으로부터 NRRL No. 15709로서 입수 가능하다.

[0057] 본 설명에 사용되는 바와 같이, "변이체 사상 진균 세포(들)", "변이체 진균 세포(들)", "변이체 세포(들)" 등의 어구는 페지조마이코티나 아문에 속하는 모 (대조군) 사상 진균 세포로부터 유래된(즉, 수득된) 사상 진균 세포를 지칭한다. 따라서, 본 설명에 정의된 바와 같은 "변이체" 사상 진균 세포는 "모" 사상 진균 세포로부터 유래된 것으로, 이때, "변이체" 세포는 "모" 세포에서 발견되지 않는 적어도 하나의 유전자 변형을 포함한다. 예를 들어, 본 발명의 "모 사상 진균 세포"에 대해 "변이체 사상 진균 세포"를 비교할 때, "모" 세포는 적어도 하나의 유전자 변형을 포함하는 "변이체" 세포와 비교하여 유전적으로 변형되지 않은 (모) "대조군" 세포로서 작용한다.

[0058] 본 설명에 사용되는 바와 같이, "유전자 변형"이라는 용어는 핵산 서열의 변경/변화를 지칭한다. 변형은 핵산 서열에서 적어도 하나의 뉴클레오티드의 치환, 결실, 삽입 또는 화학적 변형을 포함할 수 있으나, 이에 한정되지 않는다.

[0059] 본 설명에 정의된 바와 같이, 어구 "유전자 변형을 포함하는 변이체 세포(들)" 및 "Ace3-L 단백질, Ace3-EL 단백질 및/또는 Ace3-LN 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하는 변이체 세포(들)"은 적어도 하나의 카페의 Ace3-L 단백질, Ace3-EL 단백질 및/또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF의 사상 진균 세포 내로의 도입을 포함하나, 이에 한정되지 않는다. 따라서, 외인성으로 도입된 적어도 하나의 카페의 Ace3-L 단백질, Ace3-EL 단백질 및/또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF를 포함하는 사상 진균 세포는 (변형되지 않은) 모 진균 세포에 비해, 유전자 변형을 포함하는 변형체 진균 세포이다.

[0060] 다른 구현예에서, 본 발명의 모 진균 세포는 본 설명에 개시된 Ace3 단백질 중 임의의 것을 암호화하는 내인성 *ace3* 유전자(즉, Ace3-S 단백질, Ace3-SC 단백질, Ace3-L 단백질, Ace3-LC 단백질, Ace3-EL 단백질 및 Ace3-LN 단백질을 암호화하는 *ace3* 유전자)의 존재를 대상으로 스크리닝된다. 예를 들어, 모 진균 세포가 Ace3-L

단백질, Ace3-EL 단백질 또는 Ace3-LN 단백질을 암호화하는 *ace3* 유전자의 내인성 카피를 포함한다고 결정되는 경우, 이의 변이체 진균 세포는 예컨대, *ace3* 유전자의 내인성 프로모터를 이종성 프로모터로 교체하는 것에 의한 것과 같이, 유전자 변형에 의해 생성될 수 있다. 마찬가지로, 모 진균 세포가 Ace3-S 단백질, Ace3-SC 단백질 또는 Ace3-LC 단백질을 암호화하는 *ace3* 유전자의 내인성 카피를 포함한다고 결정되는 경우, 이의 변이체 진균 세포는 유전자 변형에 의해, 예컨대, 진균 세포 내로 본 발명의 Ace3-L 단백질, Ace3-EL 단백질 및/또는 Ace3-LN 단백질을 암호화하는 폴리뉴클레오티드 구성체로서 이의 Ace3-S, Ace3-SC 또는 Ace3-LC 단백질을 암호화하는 내인성 *ace3* 유전자의 유전자 변형을 더 포함할 수 있는 폴리뉴클레오티드 구성체를 도입함으로써, 생성될 수 있다.

[0061] 다른 구현예에서, 본 발명의 변이체 사상 진균 세포는 추가적인 유전자 변형을 포함할 것이다. 예를 들어, 일부 구현예에서, 그러한 변이체 사상 진균 세포(즉, 외인성으로 도입된 본 발명의 Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF의 카피를 포함하는 변이체 사상 진균 세포)는 탄소 이화대사산물 억제인자 단백질 "Cre1" 또는 Ace1 억제인자 단백질을 암호화하는 유전자의 발현 및/또는 활성을 감소시키는 유전자 변형을 더 포함한다.

[0062] 다른 구현예에서, 그러한 변이체 사상 진균 세포(즉, 외인성으로 도입된 본 발명의 Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF의 카피를 포함하는 변이체 사상 진균 세포)는 서열 번호 25 또는 서열 번호 27에 기재된 자일라나아제 조절인자 1(Xyr1)의 적어도 하나의 카피를 도입하는 유전자 변형을 더 포함한다.

[0063] 본 설명에 사용되는 바와 같이, "Ace3-L" 단백질 형태(서열 번호 6) 및 "Ace3-LN" 단백질 형태(서열 번호 14; 도 1b 참조)는 동일한 아미노산 서열을 포함한다. 그러나 항목 "Ace3-L" 및 "Ace3-LN"은 그러한 단백질 형태를 암호화하는 이의 일부 유전자와의 비교를 위해 본 발명의 일부 구현예에서 사용되며, 결코 본 발명을 제한하려는 것은 아니다.

[0064] 본 설명에 사용되는 바와 같이, 용어 "숙주 세포"는 본 설명에 기술되는 바와 같이 새로 들어오는 서열(즉, 세포 내로 도입된 폴리뉴클레오티드 서열)을 위한 숙주 또는 발현 비히클로서 작용하는 능력을 갖는 사상 진균 세포를 지칭한다.

[0065] "이종성" 핵산 구성체 또는 서열은 이것이 발현되는 세포에 고유하지 않거나 천연 형태로 존재하지 않는 서열의 일부를 갖는다. 제어 서열과 관련하여 이종성은 그것이 현재 조절하고 있는 유전자의 발현과 동일한 유전자를 조절하는 기능은 본질적으로 하지 않는 제어 서열(즉, 프로모터 또는 인핸서)을 지칭한다. 일반적으로, 이종성 핵산 서열은 이들이 존재하는 세포 또는 계놈의 일부에 대해 내인성이 아니며, 감염, 형질감염, 형질전환, 미세주입, 전기천공 등에 의해 세포에 첨가되었다. "이종성" 핵산 구성체는 천연 세포에서 발견되는 제어 서열/DNA 암호화 서열 조합과 동일하거나 상이한 제어 서열/DNA 암호화 서열 조합을 함유할 수 있다. 마찬가지로, 이종성 단백질은 종종 서로에 대해 본질적으로 동일한 관계에서 발견되지 않는 둘 이상의 하위 서열을 지칭할 것이다(예컨대, 융합 단백질).

[0066] 본 설명에서 사용되는 바와 같이, 용어 "DNA 구성체" 또는 "발현 구성체"는 핵산 서열을 지칭하며, 이는 적어도 두 개의 DNA 폴리뉴클레오티드 단편을 포함한다. DNA 또는 발현 구성체는 핵산 서열을 진균 숙주 세포 내로 도입하는 데 사용될 수 있다. DNA는 시험관 내에서(예컨대, PCR에 의해) 또는 임의의 기타 적합한 기법으로 생성될 수 있다. 일부 바람직한 구현예에서, DNA 구성체는 관심 서열(예컨대, Ace3-L 단백질을 암호화하는 서열)을 포함한다. 일부 구현예에서, 관심 폴리뉴클레오티드 서열은 프로모터에 작동 가능하게 연결된다. 일부 구현예에서, DNA 구성체는 적어도 하나의 선택 가능한 마커를 더 포함한다. 추가적인 구현예에서, DNA 구성체는 숙주 세포 염색체에 대해 상동인 서열을 포함한다. 다른 구현예에서, DNA 구성체는 숙주 세포 염색체에 대해 비-상동 서열을 포함한다.

[0067] 본 설명에서 사용되는 바와 같이, 용어 "셀룰라아제", "셀룰로오스 가수분해효소" 또는 "셀룰라아제 효소"는 엑소글루카나아제, 엑소셀로비오하이드롤라아제, 엔도글루카나아제 및/또는 β -글루코시다아제와 같은 세균 또는 진균 효소를 의미한다. 이러한 상이한 유형의 셀룰라아제 효소는 상승적으로 작용하여 셀룰로오스 및 그의 유도체를 글루코오스로 전환한다. 예를 들어, 목재 부폐 진균인 트리코더마, 퇴비 세균인 서모모노스포라 (Thermomonospora, 현재는 서모비피다(Thermobifida)), 바실루스(Bacillus), 및 셀로모나스(Cellulomonas); 스트렙토미세스(Streptomyces); 및 진균인 휴미콜라, 아스퍼지러스 및 푸사리움을 포함한 많은 미생물이 셀룰로오스를 가수분해하는 효소를 만든다. 이를 미생물이 만든 효소는 셀룰로오스의 글루코오스로의 전환에 유용한 세 가지 종류의 작용을 하는 단백질의 혼합물이다: 엔도글루카나아제(EG), 셀로비오하이드롤라아제(CBH) 및 β -글

루코시다아제(BG). 본 설명에서 정의되는 바와 같이, 용어 "엔도글루카나아제"(EG), "셀로비오하이드롤라아제"(CBH) 및 " β -글루코시다아제"(BG)는 각각 그의 약어 "EG", "CBH" 및 "BG"와 상호 교환적으로 사용된다.

[0068] 본 설명에서 사용되는 바와 같이, 용어 "탄소 제한"은 미생물이 원하는 단백질 생성물을 생산할 만큼의 탄소는 가지고 있으나 유기체의 필요조건, 예컨대, 성장 지속을 완전히 충족할 만큼의 탄소는 가지고 있지 않은 상태이다. 따라서, 최대량의 탄소가 단백질 생산에 쓰인다.

[0069] 본 설명에서 사용되는 바와 같이, 용어 "프로모터"는 하류 유전자 또는 이의 오픈 리딩 프레임(ORF)의 전사를 유도하는 기능을 하는 핵산 서열을 지칭한다. 프로모터는 일반적으로 표적 유전자가 발현되는 숙주 세포에 적절할 것이다. 프로모터는, 다른 전사 및 번역 조절 핵산 서열("제어 서열"로도 지칭됨)과 함께, 주어진 유전자의 발현에 필수적이다. 일반적으로, 전사 및 번역 조절 서열은 프로모터 서열, 리보솜 결합 부위, 전사 개시 및 종결 서열, 번역 개시 및 종결 서열, 및 인핸서 또는 활성화 인자 서열을 포함하나, 이에 한정되는 것은 아니다. 일부 구현예에서, 프로모터는 유도성 프로모터이다. 다른 구현예에서, 프로모터는 구성성 프로모터이다.

[0070] 본 설명에 사용되는 바와 같이, "프로모터 서열"은 발현 목적을 위해 특정 사상 진균에 의해 인식되는 DNA 서열이다. "프로모터"는 핵산의 전사를 유도하는 핵산 제어 서열의 어레이로서 정의된다. 본 설명에서 사용되는 바와 같이, 프로모터는 폴리머라아제 II형 프로모터의 경우에 있어서 TATA 요소와 같은, 전사 개시 부위 근처의 필수적인 핵산 서열을 포함한다. "구성성" 프로모터는 대부분의 환경적 및 발달상 조건 하에서 활성을 나타내는 프로모터이다. "유도성" 프로모터는 환경적 또는 발달상 조절 하에서 활성을 나타내는 프로모터이다.

[0071] 용어 "작동 가능하게 연결된"은 (프로모터, 또는 전사 인자 결합 부위의 어레이와 같은) 핵산 발현 제어 서열과 제2 핵산 서열 사이의 기능적 연결을 지칭하는데, 이때, 발현 제어 서열은 제2 서열에 해당하는 핵산의 전사를 유도한다. 따라서, 핵산이 또 다른 핵산 서열과 기능적 관계에 있도록 위치될 때 이 핵산은 "작동 가능하게 연결된" 것이다. 예를 들어, 분비 리더(즉 신호 웨პ티드)를 암호화하는 DNA는, 폴리펩티드의 분비에 참여하는 전단백질(pre-protein)로서 발현되는 경우 폴리펩티드를 암호화하는 DNA에 작동 가능하게 연결된다; 프로모터 또는 인핸서는, 서열의 전사에 영향을 미칠 경우, 암호화 서열에 작동 가능하게 연결된다; 또는 리보솜 결합 부위는 번역을 촉진하도록 위치되는 경우 암호화 서열에 작동 가능하게 연결된다. 일반적으로 "작동 가능하게 연결된"은 연결되는 DNA 서열들이 인접해 있음을 의미하고, 분비 리더의 경우, 연결되는 DNA 서열들이 인접해 있고 해독기(reading phase)에 있음을 의미한다. 그러나 인핸서는 인접해 있을 필요는 없다. 연결은 편리한 제한 부위에서의 라이케이션에 의해 달성된다. 그러한 부위가 존재하지 않는 경우, 합성 올리고뉴클레오티드 어댑터 또는 링커가 통상적인 관례에 따라 사용된다. 다른 구현예에서, 연결은 DNA가 서열 독립적이고 흔적이 없는 방식으로 이어진 심리스 클로닝(seamless cloning) 방법에 의해 달성된다. 심리스 클로닝은 전형적으로 김슨(Gibson) 어셈블리(NEB), 네빌더 하이파이 DNA 어셈블리(NEBuilder HiFi DNA Assembly, NEB), 골든 게이트 어셈블리(Golden Gate Assembly, NEB), 및 진아트 심리스 클로닝 및 어셈블리 시스템(GeneArt Seamless cloning and Assembly system, 서모피셔 사이언티픽(ThermoFisher Scientific))과 같은 상업적으로 이용 가능한 시스템을 이용하여(그러나 이에 한정되지 않음) 수행된다.

[0072] 본 설명에서 사용되는 바와 같이, 용어 "암호화 서열"은 (암호화된) 단백질 생성물의 아미노산 서열을 직접적으로 특정하는 뉴클레오티드 서열을 지칭한다. 암호화 서열의 경계는 일반적으로 오픈 리딩 프레임에 의해 결정되는데, 이는 보통 ATG 개시 코돈으로 시작된다. 암호화 서열은 전형적으로 DNA, cDNA, 및 재조합 뉴클레오티드 서열을 포함한다.

[0073] 본 설명에서 정의되는 바와 같이, "오픈 리딩 프레임"(이하, "ORF")은 (i) 개시 코돈, (ii) 아미노산을 나타내는 일련의 2개 이상의 코돈, 및 (iii) 종결 코돈으로 구성되는 연속된(uninterrupted) 리딩 프레임을 포함하는 핵산 또는 핵산 서열(자연 발생적, 비자연 발생적, 또는 합성이든 관계 없이)을 의미하며, ORF는 5'에서 3' 방향으로 해독(또는 번역)된다.

[0074] 본 설명에서 사용되는 바와 같이, 용어 "유전자"는 암호화 영역의 전후 영역(예컨대, 5' 비번역(5' UTR) 또는 "리더" 서열 및 3' UTR 또는 "트레일러" 서열뿐만 아니라, 개별적인 암호화 단편(엑손) 사이의 개재 서열(인트론)을 포함하거나 포함하지 않을 수 있는, 폴리펩티드 사슬을 생산하는 데 관여하는 DNA 단편)을 의미한다. 유전자는 상업적으로 중요한 산업 단백질 또는 웨პ티드, 예컨대 효소(예컨대, 프로테아제, 만나아제, 자일라나아제, 아밀라아제, 글루코아밀라아제, 셀룰라아제, 옥시다아제, 리파아제 등)를 암호화할 수 있다. 관심 유전자는 자연 발생적인 유전자, 돌연변이된 유전자 또는 합성 유전자일 수 있다.

[0075] 본 설명에서 사용되는 바와 같이, 용어 "재조합"은 세포, 핵산, 단백질, 또는 벡터에 관하여 사용될 때, 세포,

핵산, 단백질 또는 벡터가 이종성 핵산 또는 단백질의 도입, 또는 천연 핵산 또는 단백질의 변경에 의해 변형되었음을 의미하거나, 그 세포가 그렇게 변형된 세포로부터 유래함을 의미한다. 따라서, 예를 들어, 재조합 세포는 그 세포의 천연(비-재조합) 형태 내에서는 발견되지 않는 유전자를 발현하거나, 천연 유전자가 발현되거나 전혀 발현되지 않는 조건 하에서 재조합되지 않았다면 비정상적으로 발현되는 천연 유전자를 발현한다.

[0076] 용어 "벡터"는 하나 이상의 세포 유형으로 도입되는 핵산 서열을 수송하도록 설계된 폴리뉴클레오티드로서 본 설명에서 정의된다. 벡터는 클로닝 벡터, 발현 벡터, 서틀 벡터, 플라스미드, 파지 또는 바이러스 입자, DNA 구성체, 카세트 등을 포함한다. 발현 벡터는 프로모터, 신호 서열, 암호화 서열 및 전사 종결자와 같은 조절 서열을 포함할 수 있다.

[0077] 본 설명에 사용되는 바와 같은 "발현 벡터"는 적합한 숙주에서 단백질의 발현을 가져올 수 있는 적합한 제어 서열에 작동 가능하게 연결된 암호화 서열을 포함하는 DNA 구성체를 의미한다. 이러한 제어 서열은 전사를 가져오는 프로모터, 전사를 제어하는 선택적인 오퍼레이터 서열, mRNA 상의 적합한 리보솜 결합 자리를 암호화하는 서열, 인핸서 및 전사 및 번역의 종결을 제어하는 서열을 포함할 수 있다.

[0078] 본 설명에 사용되는 바와 같이, 용어 "분비 신호 서열"은, 더 큰 폴리펩티드의 구성요소로서, 그것이 합성되는 세포의 분비 경로를 통해 더 큰 폴리펩티드를 유도하는, 폴리펩티드(즉, "분비 펩티드")를 암호화하는 DNA 서열을 나타낸다. 더 큰 폴리펩티드는 보통 분비 경로를 통한 수송 중에 분비 펩티드를 제거하기 위하여 절단된다.

[0079] 본 설명에서 사용되는 바와 같이, 용어 "유도"는 "유도인자"(즉, 유도 기질)의 존재에 반응하여 현저하게 증가된 속도로 사상 진균 세포에서 관심 단백질(이하, "POI")의 합성을 가져오는 유전자의 증가된 전사를 지칭한다. "관심 유전자"(이하, "GOI") 또는 POI를 암호화하는 "관심 ORF"의 "유도"를 측정하기 위하여, 변이체 사상 진균(숙주) 세포는 후보 유도 기질(유도인자)로 처리되고, 유도 기질(유도인자)로 처리하지 않은 모 사상 진균(대조군, 비변형) 세포에 대해 비교된다. 따라서, (미처리) 모 (대조군) 세포는 100%의 상대적인 단백질 활성치를 부여받는데, 이때, 변이체 숙주 세포에서 POI를 암호화하는 GOI의 유도는, 활성치(즉, 대조군 세포에 대한)가 100% 초과, 110% 초과, 더욱 바람직하게는 150%, 더욱 바람직하게는 200 내지 500%(즉, 대조군에 비해 2 내지 5 배 초과), 또는 더욱 바람직하게는 1000 내지 3000% 초과일 때, 달성된다.

[0080] 본 설명에서 사용되는 바와 같이, 용어 "유도인자", "유도 기질"은 상호 교환적으로 사용되고, 사상 진균 세포로 하여금 유도 기질이 부존재한 경우에 생산하는 것보다 "증가된 양"의 폴리펩티드(예컨대, 효소, 수용체, 항체 등) 또는 기타 화합물/물질을 생산하게 하는 임의의 화합물을 지칭한다. 유도 기질의 예는 소포로오스, 락토오스, 젠티오비오스 및 셀룰로오스를 포함하나, 이에 한정되지 않는다.

[0081] 본 설명에서 사용되는 바와 같이, 용어 "유도성 먹이"는 사상 진균 세포에 공급되는 적어도 "유도 기질"을 포함하는 조성물로서, 이러한 유도성 먹이가 POI의 생산을 유도하는 것인, 조성물을 지칭한다.

[0082] 본 설명에서 사용되는 바와 같이, 용어 "단리된" 또는 "정제된"은 핵산 또는 아미노산이 자연적으로 회합되어 있는 적어도 하나의 구성요소로부터 제거되는 핵산 또는 아미노산을 지칭한다.

[0083] 본 설명에서 정의되는 바와 같이, 용어 "관심 단백질" 또는 "POI"는 사상 진균 세포에서 발현되는 것이 요망되는 폴리펩티드를 지칭한다. 그러한 단백질은 효소, 기질-결합 단백질, 표면 활성 단백질, 구조적 단백질 등일 수 있고, 높은 수준으로 발현될 수 있으며, 상업화의 목적을 위한 것일 수 있다. 관심 단백질은 내인성 유전자 또는 이종성 유전자(즉, 변이체 세포 및/또는 모 세포에 대해)에 의해 암호화될 수 있다. 관심 단백질은 세포 내에서 발현될 수 있거나 분비된 (세포 외) 단백질로서 발현될 수 있다.

[0084] 본 설명에서 사용되는 바와 같이, 용어 "폴리펩티드" 및 "단백질" (및/또는 그의 각각의 복수 형태)은 펩티드 결합에 의해 연결된 아미노산 잔기를 포함하는 임의의 길이의 중합체를 지칭하기 위해 상호 교환적으로 사용된다. 아미노산 잔기에 대한 통상적인 1문자 또는 3문자 코드가 본 설명에서 사용된다. 중합체는 선형 또는 분지형일 수 있고, 변형된 아미노산을 포함할 수 있으며, 비-아미노산이 개재될 수 있다. 이 용어는 또한, 자연적으로 또는 개입에 의해(예컨대, 이황화 결합 형성, 글리코실화, 지질화, 아세틸화, 인산화, 또는 표지(labeling) 성분과의 접합과 같은 임의의 다른 조작 또는 변형에 의해) 변형된 아미노산 중합체를 포함한다. 예를 들어, 아미노산의 하나 이상의 유사체(예를 들어, 비천연 아미노산 등을 포함함)뿐만 아니라 당해 분야에 공지된 다른 변형도 포함하는 폴리펩티드가 이 정의 내에 또한, 포함된다.

[0085] 본 설명에서 사용되는 바와 같이, 기능적으로 및/또는 구조적으로 유사한 단백질은 "관련 단백질"인 것으로 간주된다. 그러한 단백질은 상이한 속 및/또는 종의 유기체, 또는 심지어 상이한 강의 유기체(예컨대, 세균 및 진균)로부터 유래할 수 있다. 관련 단백질은 또한, 1차 서열 분석에 의해 결정되거나, 2차 또는 3차 구조 분석에

의해 결정되거나, 면역학적 교차반응성에 의해 결정된 상동물(homolog) 또한, 포함한다.

[0086] 본 설명에서 사용되는 바와 같이, 어구 "실질적으로 활성이 없는" 또는 유사한 어구는 소정의 활성이 혼합물에서 검출 불가능하거나 혼합물의 의도된 목적을 방해하지 않는 양으로 존재함을 의미한다.

[0087] 본 설명에서 사용되는 바와 같이, 용어 "유도체 폴리펩티드"는 N 말단 및 C 말단 중 어느 하나 또는 이들 둘 모두에 대한 하나 이상의 아미노산의 부가, 아미노산 서열에서의 하나의 또는 다수의 상이한 부위에서의 하나 이상의 아미노산의 치환, 단백질의 어느 하나 또는 양 말단에서의 또는 아미노산 서열에서의 하나 이상의 부위에서의 하나 이상의 아미노산의 결실, 및/또는 아미노산 서열에서의 하나 이상의 부위에서의 하나 이상의 아미노산의 삽입에 의해 단백질로부터 유래되거나 유래될 수 있는 단백질을 지칭한다. 단백질 유도체의 제조는 천연 단백질을 암호화하는 DNA 서열의 변형, 그 DNA 서열로 적합한 숙주를 형질전환시키는 것, 및 변형된 DNA 서열을 발현시켜 유도체 단백질을 형성하는 것에 의해 달성될 수 있다.

[0088] 관련(및 유도체) 단백질은 "변이체 단백질"을 포함한다. 변이체 단백질은 소수의 아미노산 잔기에서의 치환, 결실, 및/또는 삽입에 의해 기준/모 단백질(예컨대, 야생형 단백질)과 상이하다. 변이체 단백질과 모 단백질 사이에서 차이가 있는 아미노산 잔기의 수는 1개 이상, 예를 들어, 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 15개, 20개, 30개, 40개, 50개, 또는 그보다 많은 아미노산 잔기일 수 있다. 변이체 단백질은 기준 단백질과 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 또는 심지어 적어도 약 99%, 또는 그보다 많은 아미노산 서열 동일성을 공유할 수 있다. 변이체 단백질은 또한, 선택된 모티프, 도메인, 에피토프, 보존 영역 등에서 기준 단백질과 상이할 수 있다.

[0089] 본 설명에서 사용되는 바와 같이, 용어 "상동성 단백질"은 기준 단백질과 비슷한 활성 및/또는 구조를 갖는 단백질을 지칭한다. 상동물은 반드시 진화론으로 관련되어 있는 것은 아니다. 따라서, 이 용어는 상이한 유기체로부터 수득된 동일한, 유사한, 또는 상응하는 효소(들)(즉, 구조 및 기능면에서)을 포괄하고자 한 것이다. 일부 구현예에서, 기준 단백질과 유사한 4차, 3차 및/또는 1차 구조를 갖는 상동물을 확인하는 것이 바람직하다. 일부 구현예에서, 상동성 단백질은 기준 단백질과 유사한 면역학적 반응(들)을 유도한다. 일부 구현예에서, 상동성 단백질은 원하는 활성(들)을 갖는 효소를 생산하도록 유전자 조작된다.

[0090] 서열 사이의 상동성 정도는 당해 분야에 공지된 임의의 적합한 방법을 이용하여 결정될 수 있다(예컨대, Smith and Waterman, 1981; Needleman and Wunsch, 1970; Pearson and Lipman, 1988; 위스콘신 유전학 소프트웨어 패키지(제네티克斯 컴퓨터 그룹(Genetics Computer Group), 미국 위스콘신 주 매디슨 소재)의 GAP, BESTFIT, FASTA, 및 TFASTA와 같은 프로그램; 및 Devereux et al., 1984 참조).

[0091] 본 설명에서 사용되는 바와 같이, 어구 "실질적으로 유사한" 및 "실질적으로 동일한"은 적어도 2개의 핵산 또는 폴리펩티드의 맥락에서, 전형적으로 폴리뉴클레오티드 또는 폴리펩티드가 기준(즉, 야생형) 서열과 비교하여, 적어도 약 70% 동일성, 적어도 약 75% 동일성, 적어도 약 80% 동일성, 적어도 약 85% 동일성, 적어도 약 90% 동일성, 적어도 약 91% 동일성, 적어도 약 92% 동일성, 적어도 약 93% 동일성, 적어도 약 94% 동일성, 적어도 약 95% 동일성, 적어도 약 96% 동일성, 적어도 약 97% 동일성, 적어도 약 98% 동일성, 또는 심지어 적어도 약 99% 동일성 이상을 갖는 서열을 포함함을 의미한다. 서열 동일성은 표준적인 파라미터를 이용하여 BLAST, ALIGN, 및 CLUSTAL과 같은 공지된 프로그램을 이용하여 결정될 수 있다.

[0092] 본 설명에서 사용되는 바와 같이, 용어 "유전자"는 단백질 또는 RNA를 암호화하거나 이의 발현을 유도하는 핵산을 지칭한다는 점에서 용어 "대립 형질"과 같은 것을 의미한다. 사상 진균의 발육 형태는 일반적으로 반수체이며, 따라서 소정의 유전자의 단일 카페(즉, 단일 대립 형질)는 소정의 표현형을 부여하기에 충분하다.

[0093] 본 설명에서 사용되는 바와 같이, 용어 "야생형" 및 "천연형"은 상호 교환적으로 사용되고, 자연계에서 발견되는 것과 같은 유전자, 단백질, 진균 세포 또는 균주를 지칭한다.

[0094] 본 설명에서 사용되는 바와 같이, "유전자의 결실"은 숙주 세포의 계놈으로부터의 유전자의 제거를 지칭한다. 유전자가 유전자의 암호화 서열에 바로 인접하게 위치하지 않은 제어 요소(예컨대, 인핸서 요소)를 포함하는 경우, 유전자의 결실은 암호화 서열, 및 선택적으로, 예를 들어, 프로모터 및/또는 종결자 서열을 포함하나 이에 한정되지 않는, 인접한 인핸서 요소의 결실을 지칭한다.

[0095] 본 설명에서 사용되는 바와 같이, "유전자의 파괴"는 숙주 세포에서 세포가 기능성 유전자 생성물, 예컨대, 단백질을 생산하는 것을 실질적으로 방지하는, 임의의 유전적 또는 화학적 조작, 즉, 돌연변이를 광범위하게 지칭한다. 예시적인 파괴 방법은 폴리펩티드-암호화 서열, 프로모터, 인핸서, 또는 또 다른 조절 요소를 포함하는

유전자의 임의의 부분의 완전 또는 부분 결실, 또는 이의 돌연변이 유발을 포함하며, 여기서 돌연변이 유발은 치환, 삽입, 결실, 역전, 및 이의 조합 및 변이를 포함하고, 이 돌연변이 중 임의의 돌연변이는 기능 유전자 생성물의 생산을 실질적으로 방지한다. 유전자는 또한, RNAi, 안티센스, CRISPR/Cas9 또는 유전자 발현을 무효화하거나 감소시키거나 완화하는 임의의 기타 방법을 이용하여 파괴될 수 있다.

- [0096] 본 설명에서 사용되는 바와 같이, 어구 "본 발명의 Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 유전자의 발현을 증가시키는 '유전자 변형'을 포함하는 변이체 [숙주] 세포"는 숙주 세포 내로 그러한 Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 *ace3* 유전자 형태(또는 이의 ORF)를 포함하는 플라스미드 또는 염색체 통합 카세트를 도입(예컨대, 형질전환을 통해)하는 것을 포함한다. 일부 다른 구현예에서, 모진균 세포가 천연적으로 Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 내인성 유전자 형태를 포함할 때와 같이, 어구 "Ace3-L 단백질, Ace3-EL 및/또는 Ace3-LN 단백질을 암호화하는 유전자의 발현을 증가시키는 '유전자 변형'을 포함하는 변이체 [숙주] 세포"는 천연/야생형 *ace3* 유전자 프로모터를 이종성 프로모터로 교체하는 것을 포함한다.
- [0097] 본 설명에서 사용되는 바와 같이, "호기성 발효"는 산소의 존재 하에서의 성장을 지칭한다.
- [0098] 본 설명에서 사용되는 바와 같이, 용어 "세포 배양액(cell broth)"은 액체/액침 배양물 중의 배지 및 세포를 집합적으로 지칭한다.
- [0099] 본 설명에서 사용되는 바와 같이, 용어 "세포량(cell mass)"은 액체/액침 배양물에 존재하는 세포 구성요소(온전한 세포와 용해된 세포를 포함)를 지칭한다. 세포량은 건조 또는 습윤 중량으로 표현될 수 있다.
- [0100] 본 설명에서 사용되는 바와 같이, "기능성 폴리펩티드/단백질"은 효소 활성, 결합 활성, 표면 활성 등과 같은 활성을 보유하는 단백질로서, 그 활성을 무효화하거나 감소시키기 위하여 돌연변이화되었거나, 절단되었거나, 달리 변형되었거나 하지 않은 단백질이다. 기능성 폴리펩티드는 지정한 대로, 열에 안정적이거나 열에 불안정할 수 있다.
- [0101] 본 설명에서 사용되는 바와 같이, "기능성 유전자"는 활성 유전자 생성물, 전형적으로 단백질을 생산하기 위하여 세포 구성성분에 의해 사용될 수 있는 유전자이다. 기능성 유전자는 파괴된 유전자와는 반대되는 것으로, 파괴된 유전자는 활성 유전자 생성물을 생산하기 위하여 세포 구성성분에 의해 사용될 수 없거나 활성 유전자 생성물을 생산하기 위하여 세포 구성성분에 의해 사용되는 능력이 감소되도록 변형된 것이다.
- [0102] 본 설명에서 사용되는 바와 같이, "관심 단백질"은 액침 배양물 또는 사상 진균 세포에서 생산될 것이 요망되는 단백질이다. 일반적으로, 관심 단백질은 산업, 제약, 동물 건강, 및 식품 및 음료 용도를 위해 상업적으로 중요하므로, 대규모로 생산하는 것이 바람직하다. 관심 단백질은 사상 진균 세포에 의해 발현되는 무수한 다른 단백질과 구별되는데, 다른 단백질은 일반적으로 제품으로서 관심 대상이 아니고, 주로 배경 단백질 오염물질로 간주된다.
- [0103] 본 설명에서 사용되는 바와 같이, 변이체 세포에 의해 생산된 단백질의 양이 모 세포에 비해 적어도 5% 증가, 적어도 10% 증가, 적어도 15% 증가, 또는 그보다 많이 증가되는 경우, "변이체 진균 숙주 세포"는 "모 진균 세포"보다 "바이오매스 단위량당 실질적으로 더 많은 단백질"을 생산하며, 이때, 단백질의 양은 단백질 생산이 측정되는 세포의 총 바이오매스량에 대해 정규화되며, 바이오매스는 습윤 중량(예컨대, 세포 펠릿의 습윤 중량) 또는 건조 중량면에서 표현될 수 있다.
- [0104] **III. 셀룰라아제 발현 활성화 인자 3(ace3)**
- [0105] 최근, 셀룰라아제/헤미셀룰라아제 생산이 "유도된"(즉, 상이한 유도 조성물; 예컨대, 소포로오스, 락토오스의 첨가를 통해) 트리코더마 레에세이 배양물로부터의 전사 프로파일링 데이터(Hakkinen et al., 2014)는 셀룰라아제 및 헤미셀룰로아제 유전자 발현의 추정상의 "조절인자"를 확인하기 위하여 조사되었고, 조절 단백질을 암호화하는 후보 유전자가 확인되었다. Hakkinen 등 (2014)은 이 후보 유전자를 유전자 식별번호 77513(이때, 유전자 식별번호는 T. 레에세이 데이터베이스 2.0에서와 동일함)을 확인하였고(Hakkinen et al. 도 2 및 표 2 참조), 후보 유전자를 "셀룰라아제 발현 활성화 인자 3(activator of cellulase expression 3)"(이하, "ace3")으로 명명하고, 암호화된 단백질(즉, 후보 전사 인자)을 "Ace3"이라 명명하였다. 더욱 상세하게는, Hakkinen 등 (2014)의 연구에서는 T. 레에세이 주 QM6a의 공개적으로 이용 가능한 게놈 서열 (genome.jgi.doe.gov/Trire2/Trire2.home.html 참조)에 기초하여 예상 ace3 ORF(서열 번호 2)를 이용하였는데, 이때, QM6a 예상 주석(유전자 식별번호 77513)은 2개의 엑손과 1개의 인트론으로 구성된다(예컨대, 도 1 참조).

- [0106] 본 설명에서 기술되고, 아래 실시예 섹션에 추가로 기재되는 바와 같이, 본 발명의 출원인들은 *T. 레에세이* "RUT-C30 주" 주석을 기반으로 한 *ace3* ORF에 대한 Hakkinen 등에 기술된(즉, *Ace3*의 *T. 레에세이* "QM6a 주" 주석을 기반으로 한) 클로닝된 *ace3* ORF를 비교 및 평가하였을 때, 놀랍고 예상치 못한 결과를 발견하였다. 예를 들어, 아래 실시예 1에 기재된 바와 같이, 서열 번호 1의 *T. 레에세이* "QM6a 주" *ace3* 유전자(및 서열 번호 2의 ORF)는, 서열 번호 6의 더 긴 *Ace3* 단백질(이하, 본 설명에서 "Ace3-L"로 지칭됨)을 암호화하는 서열 번호 4의 *T. 레에세이* "RUT-C30 주" *ace3* 유전자(또는 서열 번호 5의 ORF)에 비해, 서열 번호 3의 더 짧은 *Ace3* 단백질(이하, 본 설명에서 "Ace3-S"으로 지칭됨)을 암호화한다.
- [0107] 대조적으로, *T. 레에세이* 주 Rut-C30의 공개적으로 이용 가능한 계놈 서열 (genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html 참조)(유전자 식별번호 98455)로부터 예측된 *ace3* ORF는 세 개의 엑손과 두 개의 인트론을 포함하는 더 긴 단백질 서열(즉, *T. 레에세이* QM6a로부터의 *Ace3-S*에 비해)을 포함한다(도 1a). 더욱 상세하게는, "RUT-C30" 모델에 의해 예측된 개시 코돈은 "QM6a" 모델에서의 그것의 상류에 위치해 있으며, C 말단에 넌센스 돌연변이가 있어(Poggi-Parodi et al., 2014), 더 긴 N 말단 서열과 더 짧은 C 말단 서열을 가져온다(예컨대, 도 1b 참조).
- [0108] 마찬가지로, 아래 실시예 6에 기술된 바와 같이, *ace3* 유전자 암호화 영역의 5' 말단의 위치가 명백하지 않아, 출원인은 본 설명에 기술된 바와 같이 *ace3* 유전자의 5' 말단을 더 조사하였다. 위에서 간략하게 언급된 바와 같이, DNA 서열이 동일하더라도, 합동 계놈 연구소(Joint Genome Institute, JGI)의 DNA 서열의 주석은 돌연변이 균주 Rut-C30과 야생형 균주 QM6a 사이에 차이가 있었다. QM6a 사례에서는, 암호화 영역의 5' 말단은 (도 11에 도시된 바와 같이) 엑손 3의 상류이자 인트론 2의 내부인 것으로 제시되었다. Rut-C30 사례에서는, 암호화 영역의 5' 말단은 엑손 2의 내부에 있다(도 11).
- [0109] 계놈 DNA 서열 및 추가적인 cDNA 서열의 추가적인 분석은 (도 11에 도시된 바와 같이) "엑손 1"과 "인트론 1"의 가능성 있는 존재를 시사하였다. 또한, Rut-C30의 *ace3* 암호화 영역의 3' 말단은 야생형 단리물 QM6a의 서열에 비해 미성숙 종결 코돈을 생성하는 돌연변이를 포함하였다(도 11). 따라서, 실시예 6에 기술된 바와 같이, 출원인은 도 12에 도시된 바와 같이 이렇게 상이한 가능성 있는 *ace3* 유전자 형태들의 과발현 효과를 조사하였다.
- [0110] 나아가, 아래의 실시예에 기재된 바와 같이, 출원인은 "pYL1", "pYL2", "pYL3" 및 "pYL4"(도 2a 내지 2d 플라스미드 맵 참조)라는 명칭의 네 개의 상이한 *ace3* 발현 벡터 중 하나로 *T. 레에세이* 세포를 형질전환시켜, *Ace3-S* 단백질(서열 번호 3) 및 *Ace3-L* 단백질(서열 번호 6)을 암호화하는 유전자를 구성하고(실시예 1) 시험하였다(실시예 2 내지 4). 더욱 구체적으로는(실시예 1), 이들 발현 벡터는 대장균(*E. coli*)에서의 복제 및 선택을 위해 세균의 *ColE1 ori* 및 *AmpR* 유전자를 갖는 벡터 백본을 함유한다. 또한, 이 발현 벡터(도 2a 내지 도 2d 참조)는 *T. 레에세이* *pyr2* 선택 마커 및 그것의 천연 종결자 서열과 함께 *ace3* ORF 암호화 서열(*ace3-L* 또는 *ace3-S*)에 작동 가능하게 연결된 이종성 *T. 레에세이* 프로모터 서열(즉, *hxl1* 또는 *pk1*의 프로모터)를 포함한다.
- [0111] 이후에, 생성된 안정적인 *T. 레에세이* 형질전환체(즉, 변이체 숙주 세포 A4-7, B2-1, C2-28 및 D3-1)를 "유도" 및 "비유도" 조건 모두에서 완효성(slow release) 미세적정 플레이트(srMTP)(실시예 2), 진탕 플라스크(실시예 3) 및 소규모 발효(실시예 4)에서 시험/스크리닝하였고, 배양 상청액을 수확하고 폴리아크릴아미드 젤 전기영동을 통해 분석하였다(도 3 내지 도 5 참조). 이를 실시예에서 제시된 바와 같이, 시험한 숙주 세포 모두(즉, 모세포, 변이체 A4-7, 변이체 B2-1, 변이체 C2-28 및 변이체 D3-1)는 유도 기질(즉, 소포로오스 또는 락토오스)의 존재 하에 다양한 단백질을 분비하였다. 이와 대조적으로, "글루코오스"가 유일한 탄소 공급원인, 유도 기질(즉, 소포로오스 또는 락토오스)의 부존재 하에서는, 오로지 *ace3-L* ORF를 발현하는 변이체 숙주 세포(즉, 변이체 A4-7 및 C2-28)만이 분비 단백질을 생산할 수 있었고, 모 (대조군) 세포와 *ace3-S* ORF를 발현하는 변이체 숙주 세포(즉, 변이체 B2-1 및 D3-1)는 임의의 검출 가능한 분비 단백질을 생산하지 않았다.
- [0112] 다른 구현예에서, 본 발명은 *ace3-L*(실시예 5) 발현 구성체의 발현을 추진하기 위하여 13종의 상이한 프로모터를 이용하는 "비유도" 조건 하에서의 증진된 단백질 생산을 추가로 증명한다. 예를 들어, 시험한 13종의 프로모터에는 (i) 포름아미다아제 유전자(*rev3*; 단백질 ID 103041) 프로모터(서열 번호 15), (ii) β -자일로시다아제 유전자(*bx1*; 단백질 ID 121127) 프로모터(서열 번호 16), (iii) 트랜스케톨라아제 유전자(*tk11*; 단백질 ID 2211) 프로모터(서열 번호 17), (iv) 미지의 기능의 유전자(단백질 ID 104295) 프로모터(서열 번호 18), (v) 옥시도리덕타아제 유전자(*dld1*; 단백질 ID 5345) 프로모터(서열 번호 19), (vi) 자일라나아제 IV 유전자(*xyn4*; 단백질 ID 111849) 프로모터(서열 번호 20), (vii) α -글루쿠로니다아제 유전자(단백질 ID 72526) 프로모터(서열 번호 21), (viii) 아세틸 자일란 에스테라아제 유전자 1(*axe1*; 단백질 ID 73632) 프로모터(서열 번호 22),

(ix) 헥소オス 키나아제 유전자(*hxk1*; 단백질 ID 73665) 프로모터(서열 번호 23), (x) 미토콘드리아 수송 단백질 유전자(*dic1*; 단백질 ID 47930) 프로모터(서열 번호 24), (xi) 올리고펩티드 수송 유전자(*opt*; 단백질 ID 44278) 프로모터(서열 번호 25), (xii) 글리세롤 키나아제 유전자(*gut1*; 단백질 ID 58356) 프로모터(서열 번호 26) 및 (xiii) 피루베이트 키나아제 유전자(*pki1*; 단백질 ID 78439) 프로모터(서열 번호 27)가 포함되었다. 표 3에 나타난 바와 같이, 모 *T. 레에세이* 세포는 소포로オス 유도인자의 존재 하에서만 분비 단백질을 생산하였다. 이와 대조적으로, 13종의 상이한 프로모터 중 임의의 하나로부터 추진된 *Ace3-L*을 포함하고 발현하는 변이체 (딸) *T. 레에세이* 세포는 유도 및 비유도 조건 둘 다에서 비슷한 양의 분비 단백질을 생산하였다. 또한, 실시예 5에서 기술된 바와 같이, *T. 레에세이* 모 균주 및 이의 형질전환체를 진탕 플라스크 실험 및 소규모 발효에서 추가로 시험하였다. 도 8에 나타난 바와 같이, 모 (대조군) *T. 레에세이* 세포는 소포로オス("Sop") 유도인자의 존재 하에서만 분비 단백질을 생산하였지만, 딸 균주 LT83은 유도("Sop") 및 비유도("Glu") 조건 둘 다에서 비슷한 양의 분비 단백질을 생산하였다. 마찬가지로, 도 9에 나타난 바와 같이, 모 (대조군) *T. 레에세이* 균주는 소포로オス 유도인자("Sop")의 존재 하에서만 분비 단백질을 생산하였지만, 딸 균주 LT83은 유도 ("Sop") 및 비유도("Glu") 조건 둘 다에서 비슷한 양의 단백질을 생산하였다.

[0113]

본 발명의 실시예 6은 *ace3* 유전자의 서로 다른 가능한 형태의 과발현 효과에 대한 실험적 연구를 설명한다(예컨대, 위의 돌연변이 균주 *Rut-C30* / 야생형 균주 *QM6a* 계놈 서열 주석 논의, 도 11 및 도 12 참조). 따라서, 도 12에 도시된 서로 다른 형태의 *ace3*을 *T. 레에세이*에서 과발현시켰는데, 이때, *T. 레에세이* *ace3* 유전자를 위한 과발현 벡터는 *T. 레에세이*의 글루코아밀라아제 유전자좌(*gla1*)에서 *ace3*의 표적화된 통합을 가능하게 하도록 설계되었다. 따라서, 표 5에 제시된 구성체는 *ace3* 유전자의 상이한 형태를 가짐으로써 차이가 있다. 마찬가지로, 표 7의 균주를 탄소 공급원으로서 2% 락토오스 또는 2% 글루코오스를 함유하는 액체 배지 중에서 24-웰 미세적정 플레이트에서 성장시켰는데, 이때, 총 분비 단백질의 양을 두 배지(즉, 2% 락토오스 또는 2% 글루코오스) 중의 배양 상청액으로부터 측정하였고, *ace3-L*, *ace3-EL* 및 *ace3-LN* 형태(즉, *RutC-30* C 말단 돌연변이 함유)의 과발현은 총 단백질의 생산을 향상시켰다(표 8). 탄소 공급원으로서 락토오스를 함유하는 배지에서, *ace3* 유전자의 모든 형태의 과발현은 총 단백질의 생산을 어느 정도 향상시켰으나, 향상 수준은 *ace3* 유전자의 *ace3-L*, *ace3-EL* 및 *ace3-LN* 형태를 과발현하는 균주에서 가장 높았다(표 8). 따라서, *ace3* 유전자의 *ace3-L*, *ace3-EL* 및 *ace3-LN* 형태를 과발현시킬 때 높은 수준의 분비 단백질은 "비유도 조건"(즉, 글루코오스가 탄소 공급원으로 사용되었을 때) 하에서 관찰됨은 분명하다.

[0114]

본 발명의 실시예 7은 *ace3* 유전자좌에 천연 프로모터의 5' 영역 상류를 포함하는 DNA 단편, *loxP*가 측면에 있는 하이그로마이신 B-저항성 선택 가능한 마커 카세트 및 *ace3* 오픈 리딩 프레임의 5' 말단에 작동 가능하게 융합된 관심 프로모터를 포함하는 단편을 융합시켜 제조된 프로모터 교체 구성체(도 6 참조)를 설명한다. 예를 들어, 일부 구현예에서, 프로모터 교체 구성체는 트리코더마 레에세이 세포 내의 내인성 *ace3* 유전자 프로모터를 대안적인 프로모터로 교체하는 데 사용된다.

[0115]

본 발명의 실시예 8은 내인성 비-리그노셀룰로오스 관심 유전자 프로모터를 리그노셀룰로오스 관심 유전자 프로모터로 교체하는 것을 설명한다. 예를 들어, *T. 레에세이* 글루코아밀라아제 발현 구성체를 DNA 폴리뉴클레오티드 단편으로부터 조립하였는데, 이때, *T. 레에세이* 글루코아밀라아제를 암호화하는 ORF 서열이 5'(상류) *T. 레에세이* *cbh1* 프로모터에 작동 가능하게 연결되었고, 3'(하류) *T. 레에세이* *cbh1* 종결자에 작동 가능하게 연결되었으며, 이 구성체는 선택 가능한 마커로서 *T. 레에세이* *pyr2* 유전자를 더 포함하였다. 변이체(딸) *T. 레에세이* 세포(즉, *Ace3-L* 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하는)를 글루코아밀라아제 발현 구성체로 형질전환시키고, 배양 중에 *T. 레에세이* 글루코아밀라아제 효소를 분비할 수 있는 형질전환체를 식별하기 위하여 탄소 공급원으로서 글루코오스를 함유하는(즉, 소포로오스 또는 락토오스와 같은 유도 기질이 없는) 액체 배지에서 형질전환체를 선택하고 배양하였다. 도 10에 제시된 바와 같이, 모 *T. 레에세이* 세포는 글루코오스/소포로오스를 함유하는 합성 배지(유도 조건)에서 1,029 $\mu\text{g/mL}$ 의 글루코아밀라아제를 생산하였고, 글루코오스를 함유하는 합성 배지(비유도 조건)에서는 겨우 38 $\mu\text{g/mL}$ 의 글루코아밀라아제를 생산한 반면, *dic1* 프로모터로부터 추진된 *ace3-L*을 포함하는 변형된 (딸) 균주 "LT88"은 "유도"("Sop") 조건 하에서 (즉, 모 (대조군) 균주에 비해) 3배 더 많은 글루코아밀라아제를 생산하였고, "비유도"("Glu") 조건 하에서(즉, 모 (대조군) 균주에 비해) 2.5배 더 많은 글루코아밀라아제를 생산하였다. 따라서, 이러한 결과는 *Ace3-L* ORF를 포함하는 변형된 (딸) 세포는 유도인자의 부존재 하에서 세포 외 단백질을 생산할 뿐만 아니라, 이를 변이체 세포는 이러한 유도 조건 하에서 모 (대조군) *T. 레에세이* 세포보다 더 많은 총 단백질을 생산함을 증명한다.

[0116]

실시예 9는 천연적으로 관련된 이종성 관심 유전자 프로모터의 리그노셀룰로오스 관심 유전자 프로모터로의 교체를 설명한다. 예를 들어, 부티옥셀라 종(*Buttiauxella* sp.) 피타아제를 암호화하는 ORF(즉, 이종성 GOI)는 *T.*

레에세이 *cbh1* 프로모터에 대해 5' 말단에서, 그리고 T. 레에세이 *cbh1* 종결자에 대해 3' 말단에서 작동 가능하게 연결되는데, 이때, DNA 구성체는 선택 가능한 마커를 더 포함한다. 변이체 T. 레에세이 세포(즉, *Ace3-L* 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하는)를 피타아제 발현 구성체로 형질전환시키고, 배양 중에 부티옥셀라 피타아제 효소를 분비할 수 있는 형질전환체를 식별하기 위하여 탄소 공급원으로서 글루코오스를 함유하는(즉, 소포로오스 또는 락토오스와 같은 유도 기질이 없는) 액체 배지에서 형질전환체를 선택하고 배양하였다.

[0117] 본 발명의 실시예 10은 천연 *ace3* 프로모터 교체 벡터의 구성을 설명하는데, 이때, 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) *cas9* 유전자를 함유하였고, T. 레에세이 *pki1* 프로모터 하에 발현하였으며, 가이드 RNA는 U6 프로모터 하에 발현하였다. 예를 들어, *cas9* 매개성 *ace3* 프로모터 교체 벡터(pCHL760 및 pCHL761)를 T. 레에세이 모 세포 내로 형질전환시켰고, *ace3* 프로모터 교체된 균주의 기능성을 시험하기 위하여, 진탕 플라스크의 50 ml 액침 배양액 중에서 유도 기질(소포로오스)의 존재 및 부존재 하에서 세포를 성장시켰다. SDS-PAGE에서 보여지는 바와 같이, 모 세포(도 23, ID 1275.8.1)는 글루코오스/소포로오스(유도)와 비교하여 글루코오스를 함유하는 합성 배지(비유도)에서 훨씬 적은 분비 단백질을 생산하였다. 이와 대조적으로, 형질전환체 2218, 2219, 2220, 2222 및 2223은 유도 조건 및 비유도 조건 하에서 비슷한 양의 분비 단백질을 생산하여, *hxk1* 또는 *dic1* 프로모터를 보유하는(즉, *ace3* 유전자좌에서 천연 *ace3* 프로모터를 교체하는) 변이체 세포는 유도인자의 부존재 하에 세포 외 단백질을 생산하였음을 증명하였다.

[0118] 따라서, 본 설명에서 고려되고 설명되는 바와 같이, 본 발명의 일부 양태는 1종 이상의 내인성 사상 진균 리그노셀룰로오스 분해 효소(즉, 셀룰로오스 분해 효소, 예컨대, 셀로비오하이드롤라아제, 자일라나아제, 엔도글루카나아제 등)의 생산을 대상으로 한다. 더욱 구체적으로, 본 발명의 일부 구현예는 유도 기질의 완전한 부존재 하에서 본 발명의 변이체 숙주 세포(즉, *Ace3-L* 단백질, *Ace3-EL* 및/또는 *Ace3-LN* 단백질의 발현을 증가시키는 유전자 변형을 포함하는 변이체 숙주 세포)에서의 이러한 내인성 효소의 생산을 대상으로 한다. 본 발명의 변이체 숙주 세포, 조성물 및 방법은 특히, 본 발명의 이러한 변이체 숙주 세포가 전술된 셀룰로오스 분해 효소를 생산하기 위하여 유도 기질을 필요로 하지 않는다는 사실 때문에(즉, 유도 기질의 존재 하에서만 이러한 셀룰로오스 분해 효소를 생산하는 모 세포와는 대조적으로), 이러한 셀룰로오스 분해 효소 생산비용을 상당히 감소시키는 데 있어서 특별한 효용이 있다.

[0119] 예를 들어, 일부 구현예에서, 본 발명은 유도 기질의 부존재 하에서 1종 이상의 내인성 관심 단백질 및/또는 유도 기질의 부존재 하에서 1종 이상의 이종성 관심 단백질을 발현/생산할 수 있는 변이체 진균 숙주 세포를 대상으로 한다. 따라서, 일부 구현예에서, 본 발명의 변이체 진균 숙주 세포(즉, *Ace3-L* 단백질, *Ace3-EL* 및/또는 *Ace3-LN*의 발현을 증가시키는 유전자 변형을 포함하는 진균 숙주 세포)는 내인성, 비-리그노셀룰로오스 관심 단백질 및/또는 이종성 관심 단백질을 발현하도록 더 변형된다. 예를 들어, 일부 구현예에서, 내인성 비-리그노셀룰로오스 관심 단백질을 암호화하는 유전자가 변이체 진균 숙주 세포에서 변형된다. 따라서, 일부 구현예에서, 내인성 비-리그노셀룰로오스 관심 단백질을 암호화하는 유전자와 천연적으로 관련된 프로모터는 리그노셀룰로오스 단백질을 암호화하는 사상 진균 유전자로부터의 프로모터(예컨대, 비-리그노셀룰로오스 관심 단백질을 암호화하는 내인성 유전자에 작동 가능하게 연결된 5'-리그노셀룰로오스 유전자 프로모터)로 교체된다. 마찬가지로, 일부 다른 구현예에서, 본 발명의 변이체 진균 숙주 세포(즉, *Ace3-L* 단백질, *Ace3-EL* 및/또는 *Ace3-LN*의 발현을 증가시키는 유전자 변형을 포함하는 진균 숙주 세포)는 이종성 관심 단백질을 발현하도록 변형된다. 따라서, 일부 다른 구현예에서, 이종성 관심 단백질을 암호화하는 유전자와 천연적으로 관련된 프로모터는 리그노셀룰로오스 단백질을 암호화하는 사상 진균 유전자로부터의 프로모터(예컨대, 이종성 관심 단백질을 암호화하는 이종성 유전자에 작동 가능하게 연결된 5'-리그노셀룰로오스 유전자 프로모터)로 교체된다.

[0120] 따라서, 일부 구현예에서, 본 발명은 모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포를 대상으로 하는데, 이때, 변이체 세포는 모 세포에 비해 *Ace-L* 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하며, 이때, 암호화된 *Ace3-L* 단백질은 서열 번호 6의 *Ace3-L* 단백질에 대해 약 90% 서열 동일성을 포함한다. 예를 들어, 일부 구현예에서, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 암호화된 *Ace3* 단백질은 마지막 4개의 C 말단 아미노산으로 "Lys-Ala-Ser-Asp"를 포함한다. 다른 구현예에서, 서열 번호 6에 대해 약 90% 서열 동일성을 포함하는 암호화된 *Ace3* 단백질은 서열 번호 6에 작동 가능하게 연결되며 이에 앞서는 서열 번호 98의 N 말단 아미노산 단편을 더 포함한다. 또 다른 구현예에서, *Ace3* 단백질은 서열 번호 12에 대해 약 90% 서열 동일성을 포함한다.

[0121] 일부 다른 구현예에서, 본 발명의 폴리뉴클레오티드는 서열 번호 4, 서열 번호 11 또는 서열 번호 13에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 다른 구현예에서, 본 발명의 폴리뉴클레오티드는

서열 번호 5, 서열 번호 101 또는 서열 번호 102에 대해 약 90% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다.

IV. 사상 진균 숙주 세포

본 발명의 일부 구현예에서, Ace3-L 폴리펩티드를 암호화하는 유전자 또는 ORF의 발현을 증가시키는 유전자 변형을 포함하는 변이체 사상 진균 세포(즉, 모 사상 진균 세포로부터 유래된 사상 진균 세포)가 제공된다. 더욱 구체적으로, 일부 구현예에서, 변이체 사상 진균 세포(즉, 모 (대조군) 세포에 대한)는 서열 번호 6의 Ace3-L 단백질을 암호화하는 유전자(또는 ORF)의 발현을 증가시키는 유전자 변형을 포함한다. 바람직한 구현예에서, 이러한 서열 번호 6의 Ace3-L 단백질을 암호화하는 유전자(또는 ORF)의 발현을 증가시키는 유전자 변형을 포함하는 변이체 진균 세포는 유도 기질의 부존재 하에 적어도 1종의 내인성 관심 단백질을 생산할 수 있다. 다른 구현예에서, 이러한 서열 번호 6의 Ace3-L 단백질을 암호화하는 유전자(또는 ORF)의 발현을 증가시키는 유전자 변형을 포함하는 변이체 진균 세포는 유도 기질의 부존재 하에 적어도 1종의 이종성 관심 단백질을 생산할 수 있다.

따라서, 일부 구현예에서, 본 발명에서 조작 및 사용을 위한 사상 진균 세포는 자낭균 문, 페지조마이코티나 아문의 사상 진균, 특히 영양 균사 상태를 갖는 진균을 포함한다. 이러한 유기체는 상업적으로 중요한 산업 단백질 및 제약 단백질의 생산에 사용되는 사상 진균 세포를 포함하며, 이는 트리코더마 종, 아스페질러스 종, 푸사리움 종, 스케도스포륨(*Scedosporium*) 종, 페니실리움 종, 크리소스포륨(*Chrysosporium*) 종, 세팔로스포륨(*Cephalosporium*) 종, 탈라로마이세스(*Talaromyces*) 종, 게오스미티아(*Geosmithia*) 종, 마이셀리오프토라(*Myceliophthora*) 종 및 뉴로스포라 종을 포함하나, 이에 한정되지 않는다.

특정 사상 진균은 트리코더마 레에세이(이전에는 트리코더마 롱기브라카이아툼 및 하이포크레아 제코리나로 분류됨), 아스페질러스 니게르, 아스페질러스 푸미가투스(*fumigatus*), 아스페질러스 이타코니쿠스(*itaconicus*), 아스페질러스 오리재(*oryzae*), 아스페질러스 니두란스, 아스페질러스 테레우스(*terreus*), 아스페질러스 소재(*sojae*), 아스페질러스 자포니쿠스(*japonicus*), 스케도스포륨 프롤리피칸스(*prolificans*), 뉴로스포라 크라사, 페니실리움 푸니쿨로솜(*funiculosum*), 페니실리움 크리소게눔(*chrysogenum*), 탈라로마이세스(게오스미티아) 에메르소니아(*emersonii*), 푸사리움 베네나툼(*Fusarium venenatum*), 마이셀리오프토라 써모필라(*thermophila*) 및 크리소스포륨 루크노웬세(*Chrysosporium lucknowense*)를 포함하나, 이에 한정되지 않는다.

V. 재조합 핵산 및 분자생물학

일부 구현예에서, 본 발명은 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF의 발현을 증가시키는 유전자 변형을 포함하는 변이체 사상 진균 숙주 세포를 대상으로 한다. 위에서 기재된 바와 같이, 이러한 변이체 숙주 세포는 유도 기질의 부존재 하에서(즉, 비변형 모 (대조군) 세포와 대조적으로) 1종 이상의 관심 단백질을 생산할 수 있다.

따라서, 일부 구현예에서, 본 발명은 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 유전자 또는 ORF를 포함하는 재조합 핵산을 대상으로 한다. 일부 구현예에서, 재조합 핵산은 사상 진균 숙주 세포에서의 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질의 생산을 위한 폴리뉴클레오티드 발현 카세트를 포함한다. 다른 구현예에서, 폴리뉴클레오티드 발현 카세트는 발현 벡터 내에 포함된다. 일부 구현예에서, 발현 벡터는 플라스미드이다. 다른 구현예에서, 재조합 핵산, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터는 서열 번호 4, 서열 번호 5, 서열 번호 11, 서열 번호 101, 서열 번호 13 또는 서열 번호 102에 대해 적어도 85% 서열 동일성을 포함하는 뉴클레오티드 서열을 포함한다. 또 다른 구현예에서, 재조합 핵산, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터는 서열 번호 6에 대하여 약 90% 서열 동일성을 포함하는 Ace3-L 단백질을 암호화하는 뉴클레오티드 서열을 포함한다.

일부 다른 구현예에서, 재조합 핵산(또는 이의 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터)는 1종 이상의 선택 가능한 마커를 더 포함한다. 사상 진균에 사용하기 위한 선택 가능한 마커는 *als1*, *amdS*, *hygR*, *pyr2*, *pyr4*, *pyrG*, *sucA*, 블레오마이신 내성 마커, 블라스티시딘 내성 마커, 피리티아민 내성 마커, 클로리뮤론 에틸 내성 마커, 네오마이신 내성 마커, 아데닌 경로 유전자, 트립토판 경로 유전자, 티미딘 키나아제 마커 등을 포함하나, 이에 한정되지 않는다. 특정 구현예에서, 선택 가능한 마커는 *pyr2*이며, 조성물 및 이용 방법은 PCT 공개 제W02011/153449호에 일반적으로 기재되어 있다. 따라서, 일부 구현예에서, 본 발명의 Ace3 단백질을 암호화하는 폴리뉴클레오티드 구성체는 이에 작동 가능하게 연결된 선택 가능한 마커를 암호화하는 핵산 서열을 포함한다.

[0130]

또 다른 구현예에서, 재조합 핵산, 폴리뉴클레오티드 구성체, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터는 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 유전자(또는 ORF)의 발현을 추진하는 이종성 프로모터를 포함한다. 더욱 상세하게는, 일부 구현예에서, 이종성 프로모터는 구성성 또는 유도성 프로모터이다. 특정 구현예에서, 이종성 프로모터는 *rev3* 프로모터, *bx1* 프로모터, *tk11* 프로모터, *PID104295* 프로모터, *dld1* 프로모터, *xyn4* 프로모터, *PID72526* 프로모터, *axe1* 프로모터, *hxk1* 프로모터, *dic1* 프로모터, *opt* 프로모터, *gut1* 프로모터 및 *pki1* 프로모터로 구성된 군으로부터 선택된다. 특정 이론 또는 작동 메커니즘에 의해 얹매이는 것을 바라지는 않지만, 본 설명에서는 글루코오스 제한 조건(즉, 과량의 글루코오스 농도에 비해) 하에서의 높은 발현 수준을 가져오는 *rev3*, *bx1*, *tk11*, *PID104295*, *dld1*, *xyn4*, *PID72526*, *axe1*, *hxk1*, *dic1*, *opt*, *gut1* 및 *pki1*과 같은 프로모터는 본 발명에서 특별한 효용성을 갖는다고 여겨진다. 따라서, 일부 구현예에서, 재조합 핵산(또는 폴리뉴클레오티드 구성체, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터)은 Ace3 단백질을 암호화하는 핵산 서열에 대해 5'이며 작동 가능하게 연결된 프로모터를 포함한다.

[0131]

또 다른 구현예에서, 재조합 핵산(또는 폴리뉴클레오티드 구성체, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터)은 천연 *ace3* 종결자 서열을 암호화하는 핵산 서열을 더 포함한다. 따라서, 일부 구현예에서, 재조합 핵산(또는 폴리뉴클레오티드 구성체, 폴리뉴클레오티드 발현 카세트 또는 이의 발현 벡터)은 Ace3 단백질을 암호화하는 핵산 서열에 대해 5'이며 작동 가능하게 연결된 프로모터 및 Ace3 단백질을 암호화하는 핵산 서열에 대해 3'이며 작동 가능하게 연결된 천연 *ace3* 종결자 서열(예컨대, 5'-Pro-ORF-Term-3', 여기서 "Pro"는 구성성 프로모터이고, "ORF"는 Ace3을 암호화하며, "Term"은 천연 *ace3* 종결자 서열이다)을 포함한다.

[0132]

따라서, 일부 구현예에서, 사상 진균의 형질전환 및 진균 배양을 위한 표준 기술(이는 당업자에게 잘 알려져 있다)이 본 발명의 진균 숙주 세포를 형질전환하는 데 사용된다. 따라서, DNA 구성체 또는 벡터를 진균 숙주 세포내로 도입하는 것은 형질전환, 전기천공법, 핵 미세주입법, 형질도입, 형질감염(예컨대, 리포액션 매개 및 DEAE-텍스트린 매개 형질감염), 인산칼슘을 이용한 인큐베이션에 의한 DNA 침전, DNA-코팅된 미세발사체(microparticle)를 이용한 고속 충격(bombardment), 유전자 총 또는 바이오리스틱 형질전환(biolistic transformation) 및 원형질 융합 등과 같은 기법을 포함한다. 일반적인 형질전환 기법은 당해 분야에 공지되어 있다(예컨대, Ausubel et al., 1987, Sambrook et al., 2001 및 2012, 그리고 Campbell et al., 1989 참조). 트리코더마에서의 이종성 단백질의 발현은 예를 들어, 미국 특허 제6,022,725호; 제6,268,328호; Harkki et al., 1991 및 Harkki et al., 1989에 기술되어 있다. 또한, 아스퍼지스 균주의 형질전환에 대해서는 Cao et al. (2000)을 참조한다.

[0133]

일반적으로, 트리코더마 종의 형질전환은 투과성 처리된 원형질 또는 세포를 통상적으로 10^5 내지 10^7 /mL, 특히 2×10^6 /mL의 밀도로 사용한다. 적절한 용액(예컨대, 1.2 M 솔비톨 및 50 mM CaCl₂) 중의 이러한 원형질 또는 세포의 100 μ L의 부피가 목적하는 DNA와 혼합된다. 일반적으로, 높은 농도의 폴리에틸렌 글리콜(PEG)이 흡수(uptake) 용액에 첨가된다. 첨가제, 예컨대 디메틸 셀록시드, 헤파린, 스페르미딘, 염화칼륨 등이 또한, 형질전환을 촉진하기 위해 흡수 용액에 첨가될 수 있다. 다른 진균 숙주 세포에 대해 유사한 절차가 이용 가능하다. 예컨대, 미국 특허 제6,022,725호 및 제6,268,328호(이들 모두는 참조로 포함됨) 참조.

[0134]

일부 구현예에서, 본 발명은 사상 진균 숙주 세포에 대해 내인성인 1종 이상의 관심 단백질(즉, 내인성 단백질은 Ace3-L의 발현을 증가시키는 유전자 변형을 포함하는 본 발명의 변이체 진균 숙주 세포에 의해 생산됨)의 발현 및 생산을 대상으로 한다. 다른 구현예에서, 본 발명은 사상 진균 숙주 세포에 대해 이종성인 1종 이상의 관심 단백질의 발현 및 생산을 대상으로 한다. 따라서, 본 발명은 일반적으로 재조합 유전학 분야의 일상적인 기술에 의존한다. 본 발명의 일반적인 이용 방법을 개시하고 있는 기본 문서는 Sambrook et al., (2nd Edition, 1989); Kriegler (1990) 및 Ausubel et al., (1994)을 포함한다.

[0135]

따라서, 일부 구현예에서, 관심 단백질을 암호화하는 이종성 유전자 또는 ORF는 사상 진균 (숙주) 세포 내로 도입된다. 일부 구현예에서, 이종성 유전자 또는 ORF는 전형적으로 복제 및/또는 발현을 위해 사상 진균 (숙주) 세포 내로 형질전환되기 전에 중간 벡터 내로 클로닝된다. 이러한 중간 벡터는 예컨대, 플라스미드와 같은 원핵 벡터, 또는 셀룰 벡터일 수 있다. 일부 구현예에서, 이종성 유전자 또는 ORF의 발현은 그것의 천연 프로모터의 제어 하에 있다. 다른 구현예에서, 이종성 유전자 또는 ORF의 발현은 이종성 구성성 프로모터 또는 이종성 유도성 프로모터일 수 있는 이종성 프로모터의 제어 하에 위치된다.

[0136]

당업자는 자연(천연) 프로모터가 그것의 기능을 변화시키지 않고 1개 이상의 뉴클레오티드의 교체, 치환, 첨가 또는 제거에 의해 변형될 수 있음을 알고 있다. 본 발명의 실시는 프로모터에 대한 이러한 변경을 포함하지만,

그에 의해 제한되지 않는다.

- [0137] 발현 벡터/구성체는 전형적으로 이종성 서열의 발현에 필요한 추가적인 요소 전부를 함유하는 전사 단위 또는 발현 카세트를 함유한다. 예를 들어, 전형적인 발현 카세트는 관심 단백질을 암호화하는 이종성 핵산 서열에 작동 가능하게 연결된 5' 프로모터를 함유하고, 전사체, 리보솜 결합 부위의 효율적인 폴리아데닐화 및 번역 종결에 필요한 서열 신호를 더 포함할 수 있다. 카세트의 추가적인 요소는 인핸서를 포함할 수 있고, 캐놈 DNA가 구조적 유전자로서 사용되는 경우, 기능적 스플라이스 공여자 및 수용자 자리가 있는 인트론을 포함할 수 있다.
- [0138] 프로모터 서열 이외에도, 발현 카세트는 또한, 효율적인 종결을 제공하기 위하여 구조적 유전자의 하류에 전사 종결 영역을 함유할 수 있다. 종결 영역은 프로모터 서열과 동일한 유전자로부터 수득될 수 있거나 상이한 유전자들로부터 수득될 수 있다. 임의의 진균 종결자가 본 발명에서 기능성일 가능성이 있지만, 바람직한 종결자는 트리코더마 *cbhI* 유전자로부터의 종결자, 아스퍼질러스 *niudans* *trpC* 유전자로부터의 종결자(Yelton et al., 1984; Mullaney et al., 1985), 아스퍼질러스 *아와모리*(awamori) 또는 아스퍼질러스 *니게르*(niger) 글루코아밀라아제 유전자(Nunberg et al., 1984; Boel et al., 1984) 및/또는 뮤코어 미에헤이(Mucor miehei) 카르복실 프로테아제 유전자(EPO 공개 제0215594호)가 포함된다.
- [0139] 세포 내로 유전자 정보를 수송하는 데 사용되는 구체적인 발현 벡터는 특별히 결정적이지는 않다. 진핵 또는 원핵 세포에서의 발현에 사용되는 통상적인 임의의 벡터가 사용될 수 있다. 표준적인 박테리아 발현 벡터에는 박테리오파지 λ 및 M13뿐만 아니라, 플라스미드, 예컨대 pBR322 기반 플라스미드, pSKF, pET23D, 및 융합 발현 시스템, 예컨대 MBP, GST, 및 LacZ가 포함된다. 에피토프 태그, 예컨대, c-myc 또한, 편리한 단리 방법을 제공하기 위해 재조합 단백질에 부가될 수 있다.
- [0140] 발현 벡터에 포함될 수 있는 요소는 또한, 레플리콘, 재조합 플라스미드를 보유하는 박테리아의 선택을 허용하는 항생제 내성을 암호화하는 유전자, 또는 이종성 서열의 삽입을 허용하는 플라스미드의 비필수 영역 내 고유한 제한 부위일 수 있다. 선택된 특정한 항생제 내성 유전자가 결정적인 것은 아니다; 당해 분야에 공지된 임의의 여러 내성 유전자가 적합할 수 있다. 원핵생물 서열은 바람직하게는 트리코더마 레에세이에서 DNA의 복제 또는 통합을 방해하지 않도록 선택된다.
- [0141] 본 발명의 형질전환 방법은 사상 진균 캐놈 내로의 형질전환 벡터의 전부 또는 일부의 안정적인 통합을 가져올 수 있다. 그러나 자가-복제성 염색체 외 형질전환 벡터의 유지를 가져오는 형질전환 또한, 고려된다.
- [0142] 여러 표준적인 형질감염 방법이 다량의 이종성 단백질을 발현하는 트리코더마 레에세이 세포주를 생산하는 데 이용될 수 있다. 트리코더마의 셀룰라아제 생산 군주 내로의 DNA 구성체의 도입을 위한 공개된 방법 일부에는 문헌 [Lorito, Hayes, DiPietro and Harman, 1993, Curr. Genet. 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, 1990, Curr. Genet. 17:169-174; Penttila, Nevalainen, Ratto, Salminen and Knowles, 1987, Gene 6: 155-164], 아스퍼질러스의 경우, 문헌 [Yelton, Hamer and Timberlake, 1984, Proc. Natl. Acad. Sci. USA 81: 1470-1474], 푸사리움의 경우, 문헌 [Bajar, Podila and Kolattukudy, 1991, Proc. Natl. Acad. Sci. USA 88: 8202-8212], 스트렙토미세스의 경우, 문헌 [Hopwood et al., 1985, The John Innes Foundation, Norwich, UK] 및 바실루스의 경우, 문헌 [Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, 1990, FEMS Microbiol. Lett. 55: 135-138]이 포함된다.
- [0143] 숙주 세포 내로 외래 뉴클레오파이드 서열을 도입하기 위한 임의의 널리 공지된 절차가 이용될 수 있다. 이에는 칼슘 포스페이트 형질감염, 폴리브렌, 원형질체 융합, 전기천공, 바이오리스틱(biolistics), 리포좀, 마이크로주사, 플라즈마 벡터, 바이러스 벡터 및 클로닝된 캐놈 DNA, cDNA, 합성 DNA 또는 외래 유전 물질의 숙주 세포 내로의 도입을 위한 임의의 다른 공지된 방법의 이용이 포함된다(예컨대, 문헌 [Sambrook et al., 위 문헌] 참조). 미국 특허 제6,255,115호에 기재된 것과 같은 아그로박테리움-매개 형질감염 방법 또한, 이용된다. 이용된 특정한 유전자 조작 절차가 이종성 유전자를 발현할 수 있는 숙주 세포 내로 적어도 하나의 유전자를 성공적으로 도입할 수 있기만 하면 된다.
- [0144] 발현 벡터가 세포 내로 도입된 후, 형질감염된 세포는 셀룰라아제 유전자 프로모터 서열의 제어 하에 유전자의 발현을 선호하는 조건 하에 배양된다. 대규모 배치의 형질전환된 세포가 본 설명에 기술된 바와 같이 배양될 수 있다. 마지막으로, 표준적인 기술을 이용하여 배양물로부터 생성물이 회수된다.
- [0145] 따라서, 본 설명의 발명은, 발현이 자연 발생적인 셀룰라아제 유전자, 융합 DNA 서열 및 다양한 이종성 구성체를 포함하는 셀룰라아제 유전자 프로모터 서열의 제어 하에 있는, 원하는 폴리펩티드의 발현 및 증진된 분비를 위해 제공된다. 또한, 본 발명은 이러한 높은 수준의 원하는 폴리펩티드를 발현 및 분비하는 방법을 제공한다.

[0146] VI. 관심 단백질

[0147] 위에서 언급된 바와 같이, 본 발명의 일부 구현예는 (모 사상 진균 세포로부터 유래된 변이체 사상 진균 세포)를 대상으로 하는데, 이때, 변이체 세포는 (모 세포에 비해) Ace3-L, Ace3-EL 또는 Ace3-LN 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하며, 이때, 암호화된 Ace3-L, Ace3-EL 또는 Ace3-LN 단백질은 서열 번호 6의 Ace3-L 단백질에 대해 약 90% 서열 동일성을 포함하고, 변이체 세포는 유도 기질의 부존재 하에 적어도 1종의 관심 단백질(POI)을 발현한다.

[0148] 본 발명의 일부 구현예는 유도 기질의 부존재 하에 단백질(즉, 관심 단백질)의 세포 내 및/또는 세포 외 생산을 증가시키는 데 특히 유용하다. 관심 단백질은 내인성 단백질(즉, 숙주 세포 내에서 내인성) 또는 이종성 단백질(즉, 숙주 세포에서 본래의 것이 아님)일 수 있다. 본 발명에 따라 생산될 수 있는 단백질은 호르몬, 효소, 성장 인자, 사이토카인, 항체 등을 포함하나, 이에 한정되지 않는다.

[0149] 예를 들어, 관심 단백질은 헤미셀룰라아제, 페록시다아제, 프로테아제, 셀룰라아제, 자일라나아제, 리파아제, 포스포리파아제, 에스테라아제, 큐티나아제, 펙티나아제, 케라티나아제, 리덕타아제, 옥시다아제, 페놀 옥시다아제, 리폭시케나아제, 리그니나아제, 폴루라나아제, 탄나아제, 펜토사나아제, 만나아제, β -글루카나아제, 히알루로니다아제, 콘드로이티나아제, 락카아제, 아밀라아제, 글루코아밀라아제, 아세틸 에스테라아제, 아미노펩티나아제, 아밀라아제, 아라비나아제, 아라비노시다아제, 아라비노푸라노시다아제, 카르복시펩티다아제, 카탈라아제, 테옥시리보뉴클레아제, 에피머라아제, α -갈락토시다아제, β -갈락토시다아제, α -글루카나아제, 글루칸라이사아제, 엔도- β -글루카나아제, 글루코오스 옥시다아제, 글루쿠로니다아제, 인버타아제, 이소머라아제 등을 포함하나, 이에 한정되지 않는다.

[0150] 일정 구현예에서, 관심 단백질은 PCT 출원 공개 제W003/027306호, 제W0200352118호, 제W0200352054호, 제W0200352057호, 제W0200352055호, 제W0200352056호, 제W0200416760호, 제W09210581호, 제W0200448592호, 제W0200443980호, 제W0200528636호, 제W0200501065호, 제W02005/001036호, 제W02005/093050호, 제W0200593073호, 제W0200674005호, 제W02009/149202호, 제W02011/038019호, 제W02010/141779호, 제W02011/063308호, 제W02012/125951호, 제W02012/125925호, 제W02012125937호, 제W0/2011/153276호, 제W02014/093275호, 제W02014/070837호, 제W02014/070841호, 제W02014/070844호, 제W02014/093281호, 제W02014/093282호, 제W02014/093287호, 제W02014/093294호, 제W02015/084596호 및 제W02016/069541호에 개시된 효소를 포함하나, 이에 한정되지 않는다.

[0151] 단백질 합성을 위한 최적 조건은 숙주 세포의 선택 및 발현되는 단백질(들)의 선택에 따라 달라질 것이다. 이러한 조건은 통상적인 실험 및/또는 최적화를 통해 당업자에 의해 용이하게 확인될 수 있다.

[0152] 관심 단백질은 발현 후 정제 또는 단리될 수 있다. 관심 단백질은 어떤 다른 구성성분들이 샘플에 존재하는지에 따라 당업자에 공지된 다양한 방법으로 단리 또는 정제될 수 있다. 표준적인 정제 방법은 전기천공 기술, 분자적 기술, 면역학적 기술 및 이온 교환, 소수성, 친화도 및 역상 HPLC 크로마토그래피를 포함한 크로마토그래피 기술, 및 크로마토포커싱을 포함한다. 예를 들어, 관심 단백질은 표준적인 관심 항-단백질 항체 컬럼을 이용하여 정제될 수 있다. 단백질 농도와 함께, 한외여과 및 정용여과 또한, 유용하다. 필요한 정제도는 관심 단백질의 목적했던 용도에 따라 다를 것이다. 일부 예에서, 단백질의 정제는 전혀 필요하지 않을 것이다.

[0153] 일부 다른 구현예에서, 본 발명의 유전자 변형된 진균 세포(즉, ace3-L의 발현을 증가시키는 유전자 변형을 포함하는 변이체 진균 숙주 세포)가 증가된 수준의 관심 단백질을 생산할 수 있는 능력이 있음을 확인하기 위하여, 다양한 스크리닝 방법이 수행될 수 있다. 발현 벡터는 검출 가능한 표지로서 작용하는 표적 단백질에 융합된 폴리펩티드를 암호화할 수 있거나, 표적 단백질 그 자체가 선택 가능한 또는 스크리닝 가능한 마커로서 작용할 수 있다. 표지된 단백질은 웨스턴 블로팅, 도트 블로팅(콜드 스프링 하버 프로토콜즈(Cold Spring Harbor Protocols) 웹사이트에서 이용 가능한 방법), ELISA를 통해, 또는 표지가 GFP인 경우 전체(whole) 세포 형광 또는 FACS를 통해 검출될 수 있다. 예컨대, 6-히스티딘 태그가 표적 단백질에 대한 융합체로서 포함되며, 이 태그는 웨스턴 블로팅에 의해 검출된다. 표적 단백질이 충분히 높은 수준으로 발현되는 경우, 쿠마시/은(Coomassie/silver) 염색과 결합된 SDS-PAGE가 모 (대조군) 세포에 비해 변이체 숙주 세포 발현에서의 증가를 검출하기 위해 수행될 수 있으며, 이러한 경우에는 표지가 필요치 않다. 또한, 향상된 수준의 관심 단백질을 확인하기 위해 다른 방법들, 예컨대 세포 당 단백질 활성 또는 양, 배양 또는 발효가 더 오랜 기간 동안 효율적으로 지속되도록 하는, 배지 밀리리터 당 단백질 활성 또는 양의 증가의 검출, 또는 이러한 방법들의 조합이 사용될 수 있다.

- [0154] 비 생산성(specific productivity)의 검출은 단백질 생산을 평가하는 또 다른 방법이다. 비 생산성(Q_p)은 다음 식에 의해 결정될 수 있다:
- [0155] $Q_p = gP/gDCW \cdot hr$
- [0156] 여기서, "gP"는 탱크에서 생산된 단백질의 그램이고, "gDCW"는 탱크 내 건조 세포 중량(DCW)의 그램이고, "hr"은 접종 시점으로부터의 발효 시간(hour)으로서, 이는 생산 시간 및 성장 시간을 포함한다.
- [0157] 일부 구현예에서, 변이체 진균 숙주 세포는 (비변형) 모 세포에 대하여 비교하여, 관심 단백질을 적어도 약 0.5%, 예컨대 적어도 약 0.5%, 적어도 약 0.7%, 적어도 약 1%, 적어도 약 1.5%, 적어도 약 2.0%, 적어도 약 2.5%, 또는 심지어 적어도 약 3% 이상으로 생산할 수 있다.
- [0158] **VII. 발효**
- [0159] 일부 구현예에서, 본 발명은 변이체 진균 세포를 발효시키는 단계를 포함하는, 관심 단백질의 생산 방법을 제공하는데, 이때, 변이체 진균 세포는 관심 단백질을 분비한다. 일반적으로, 당해 분야에 잘 알려져 있는 발효 방법이 변이체 진균 세포를 발효시키는 데에 사용된다. 일부 구현예에서, 진균 세포는 배치 또는 연속 발효 조건 하에서 성장시킨다. 전통적인 배치식 발효는 배지의 조성이 발효 개시 시 설정되고, 발효 중에는 변경되지 않는 폐쇄 시스템이다. 발효 개시 시, 배지에 요망되는 유기체(들)가 접종된다. 이 방법에서는, 이러한 시스템에 임의의 구성요소를 추가하지 않고도 발효가 일어날 수 있다. 전형적으로, 배치식 발효는 탄소 공급원의 첨가와 관련하여 "배치"로서의 자격이 있으며, pH 및 산소 농도와 같은 인자를 제어하기 위한 시도가 종종 이루어진다. 배치식 시스템의 대사 산물 및 바이오매스 조성은 발효가 중단되는 시점까지 지속적으로 변화한다. 배치 배양 내에서, 세포는 정적인 유도기(lag phase)에서 고 성장 대수기(log phase)로 진행하고, 최종적으로 정지기(stationary phase)가 되어, 성장률이 감소하거나 중단된다. 처리되지 않는 경우, 정지기에 있는 세포는 결국 사멸한다. 일반적으로, 대수기에 있는 세포는 생산물의 대량 생산의 원인이 된다.
- [0160] 표준적인 배치식 시스템에 대한 적합한 변형은 "유가식 발효" 시스템이다. 전형적인 배치식 시스템의 이러한 변형에서는, 발효가 진행됨에 따라 기질이 충분 첨가된다. 유가식 시스템은 이화대사산물 억제가 세포의 대사를 저해할 가능성이 있을 때, 그리고 배지 중에 제한된 양의 기질이 있는 것이 바람직한 경우 유용하다. 유가식 시스템에서 실제 기질 농도를 측정하는 것은 어렵고, 따라서 pH, 용존 산소 및 폐가스, 예컨대 CO_2 의 분압과 같은 측정 가능한 인자의 변화에 기초하여 계산된다. 배치식 및 유가식 발효는 당해 분야에서 일반적이며 널리 공지되어 있다.
- [0161] 연속식 발효는 합성 발효 배지가 생물 반응기에 연속적으로 첨가되고, 동량의 조정 배지가 프로세싱을 위하여 동시에 제거되는 개방 시스템이다. 연속식 발효는 일반적으로 배양물을 일정한 고밀도로 유지하는데, 이때 세포는 주로 대수기 성장에 있다. 연속식 발효는 세포 성장 및/또는 생성물 농도에 영향을 미치는 하나 이상의 인자의 조절을 허용한다. 예를 들어, 일부 구현예에서, 탄소원 또는 질소원과 같은 제한 영양소는 고정된 비율로 유지되고 다른 모든 파라미터는 적정 수준이 되도록 한다. 다른 시스템에서는, 배지 탁도에 의해 측정되는 세포 농도를 일정하게 유지하면서, 성장에 영향을 미치는 다수의 인자를 계속 변경할 수 있다. 연속식 시스템은 정상 상태의 성장 조건을 유지하려고 한다. 따라서, 배지의 배출로 인한 세포 손실은 발효에서 세포 성장 속도와 균형을 이룰 것이다. 연속식 발효 공정을 위한 영양소 및 성장 인자를 조절하는 방법뿐만 아니라, 생성물 형성 속도를 최대화하기 위한 기법은 산업 미생물학 분야에 잘 알려져 있다.
- [0162] 본 발명의 일부 구현예는 진균 배양을 위한 발효 절차와 관련이 있다. 셀룰라아제 효소의 생산을 위한 발효 절차는 당해 분야에 공지되어 있다. 예를 들어, 셀룰라아제 효소는 배치식, 유가식 및 연속 순환식 공정을 포함한 고체 또는 액침 배양에 의해 생산될 수 있다. 배양은 일반적으로 수성 미네랄 염 배지, 유기 성장 인자, 탄소 및 에너지 공급원 물질, 분자 산소, 그리고 물론, 이용되는 사상 진균 숙주의 개시 접종원을 포함하는 성장 배지 중에서 탄성된다.
- [0163] 탄소 및 에너지 공급원, 산소, 동화성 질소, 및 미생물 접종원 이외에도, 적절한 미생물 성장을 확실하게 하기 위하여, 미생물 전환 공정에서 세포에 의한 탄소 및 에너지 공급원의 동화를 최대화하기 위하여, 그리고 발효 배지에서 최대 세포 밀도로 최대 세포 수율을 달성하기 위하여, 적절한 양의 미네랄 영양소를 적당한 비율로 공급하는 것이 필수적이다.
- [0164] 수성 미네랄 배지의 조성은 당해 분야에 공지된 바와 같이, 부분적으로는 사용되는 미생물 및 기질에 따라, 넓은 범위에 걸쳐 달라질 수 있다. 미네랄 배지는 질소 이외에도, 적절한 양의 인, 마그네슘, 칼슘, 칼륨, 황, 및

나트륨을 적절한 가용성의 동화 가능한 이온 형태 및 조합된 형태로 포함해야 하고, 또한, 바람직하게는 일부 미량 원소, 예컨대, 구리, 망간, 몰리브덴, 아연, 철, 봉소, 및 요오드 등이 다시, 적절한 가용성의 동화 가능한 형태로 존재해야 하며, 모두 당해 분야에 공지된 바와 같다.

[0165] 발효 반응은 미생물 종이 번성하는 식으로 성장하도록 돋는 데 효과적인, 적절한 산소 부분압으로 발효 용기의 내용물을 유지하도록 제공된, 분자 산소를 함유하는 기체, 예컨대, 공기, 산소가 풍부한 공기, 또는 심지어 실질적으로 순수한 분자 산소에 의해 필요한 분자 산소가 공급되는 호기성 과정이다.

[0166] 발효 온도는 다소 달라질 수 있으나, 트리코더마 레에세이와 같은 사상 진균의 경우, 온도는 일반적으로 약 20 ° C 내지 40 ° C의 범위 내, 일반적으로 바람직하게는 약 25 ° C 내지 34 ° C의 범위 내일 것이다.

[0167] 미생물은 또한, 동화 가능한 질소의 공급원을 요구한다. 동화 가능한 질소원은 임의의 질소를 함유하는 화합물 또는 미생물에 의한 대사 이용에 적합한 형태로 질소를 방출할 수 있는 화합물일 수 있다. 단백질 가수분해물과 같은 다양한 유기 질소원 화합물이 이용될 수 있지만, 통상 값싼 질소 함유 화합물, 예컨대 암모니아, 수산화암모늄, 우레아, 및 다양한 암모늄염, 예컨대 인산암모늄, 황산암모늄, 피로인산암모늄, 염화암모늄, 또는 다양한 기타 암모늄 화합물이 이용될 수 있다. 암모니아 기체 자체는 대규모 작업에 편리하며, 적당한 양으로 수성 발효물(발효 배지)을 통한 버블링에 의해 이용될 수 있다. 동시에, 이러한 암모니아는 또한, pH 제어를 돋기 위해 이용될 수 있다.

[0168] 수성 미생물 발효물(발효 혼합물) 내 pH 범위는 약 2.0 내지 8.0의 예시적 범위 내에 있어야 한다. 사상 진균으로는, pH는 보통 약 2.5 내지 8.0의 범위 내이고; 트리코더마 레에세이로는, pH는 보통 약 3.0 내지 7.0의 범위 내이다. 미생물의 pH 범위에 대한 선호는 이용되는 배지에 어느 정도 의존적일 뿐만 아니라, 특정 미생물에도 의존적이며, 따라서 당업자에 의해 용이하게 결정될 수 있는 바와 같이, 배지의 변화에 따라 다소 변화한다.

[0169] 바람직하게는, 발효는 탄소 함유 기질이 제한 인자로서 제어될 수 있는 방식으로 수행되어, 탄소 함유 기질의 세포로의 양호한 전환을 제공하며 실질적인 양의 비전환 기질로 세포가 오염되는 것을 방지한다. 기질로의 세포 오염은 수용성 기질에 따른 문제는 아닌데, 임의의 잔존 미량물은 용이하게 세척되기 때문이다. 그러나 비수용성 기질의 경우 이는 문제가 될 수 있으며, 적합한 세척 단계와 같은 추가된 생성물 처리 단계를 필요로 한다.

[0170] 위에 기술된 바와 같이, 이 수준에 도달하는 시간은 중요하지 않으며 특정 미생물 및 수행하는 발효 공정에 따라 다를 수 있다. 그러나 발효 배지 내 탄소원 농도 및 원하는 수준의 탄소원이 달성되었는지를 어떻게 결정하는지는 당해 분야에 잘 알려져 있다.

[0171] 발효는 배치식 또는 연속식 작업으로서 수행될 수 있지만, 유가식 작업이 제어의 용이함, 생성물의 균일량 제조 및 모든 기기의 가장 경제적인 사용을 위해 많이 선호된다.

[0172] 원하는 경우, 수성 미네랄 배지를 발효기에 공급하기 전에, 일부 또는 모든 탄소원 및 에너지원 물질 및/또는 일부 동화성 질소원, 예컨대 암모니아가 수성 미네랄 배지에 첨가될 수 있다.

[0173] 반응기 내로 도입된 각각의 스트립은 바람직하게는 사전에 결정된 속도로, 또는 탄소 및 에너지 기질의 농도, pH, 용존 산소, 발효기로부터의 배출 기체 내 산소 또는 이산화탄소, 건조 세포 중량, 광 투과율에 의해 측정 가능한 세포 밀도 등과 같은 모니터링에 의해 결정할 수 있는 필요성에 따라, 제어된다. 다양한 물질의 공급 속도는, 탄소원 및 에너지원의 효율적 이용과 일치하여, 가능한 한 빠른 세포 성장 속도가 수득되도록, 기질 충전에 대해 가능한 한 높은 수율의 미생물 세포가 수득되도록, 달라질 수 있다.

[0174] 배치식, 또는 바람직한 유가식 작업에서, 모든 기기, 반응기, 또는 발효 수단, 관(vessel) 또는 용기, 배관, 부수 순환 또는 냉각 장치 등은, 통상 약 121°C에서와 같은 스텀을 적어도 약 15분 동안 이용하여 초기에 멸균된다. 그런 다음, 멸균된 반응기에 산소를 포함한 모든 필요한 영양소 및 탄소를 함유하는 기질의 존재 하에 선택된 미생물의 배양물을 접종한다. 이용되는 발효기의 종류는 중요하지 않다.

[0175] 발효 배양액으로부터의 (예컨대, 셀룰라아제) 효소의 수집 및 정제는 또한, 당업자에 공지된 절차에 의해 수행될 수 있다. 발효 배양액은 일반적으로 세포를 포함한 세포 잔해, 다양한 부유 고체 및 기타 바이오매스 오염물 뿐만 아니라, 원하는 셀룰라아제 효소 생성물을 함유할 것이며, 이는 바람직하게는 당해 분야에 공지된 수단에 의해 발효 배양액으로부터 제거된다.

[0176] 이러한 제거를 위한 적합한 방법은 통상적인 고체-액체 분리 기술, 예를 들어 원심분리, 여과, 투석, 미세여과, 회전 진공 여과, 또는 세포 불포함 여과액을 생산하는 기타 공지된 방법들을 포함한다. 한외여과, 증발 또는 침전과 같은 기술을 사용하여 결정화 전에 발효 배양액 또는 세포 불포함 여과액을 추가로 농축하는 것이 바람직

할 수 있다.

[0177] 상청액 또는 여과액의 단백질 성분을 침전시키는 것은 염, 예를 들어 황산암모늄을 이용하고, 다음으로 다양한 크로마토그래피 절차, 예를 들어 이온 교환 크로마토그래피, 친화도 크로마토그래피 또는 비슷한, 업계에서 인지된 절차에 의한 정제가 이어질 수 있다.

실시예

[0179] 다음 실시예는 본 발명의 구현예를 나타내지만, 단지 예시로서 제공되는 것임을 이해해야 한다. 위의 논의 및 이를 실시예로부터, 당업자는 다양한 용도 및 조건에 맞도록 본 발명을 다양하게 변경하고 변형할 수 있다. 이러한 변형은 또한, 청구된 발명의 범위 내에 속하는 것으로 의도된다.

실시예 1

사상 진균 세포에서의 *ace3* 과발현의 생성

1A. 개요

[0183] 본 실시예에서, *ace3* 유전자를 발현하는 변이체 트리코더마 레에세이 세포(즉, 예시적인 사상 진균)는 모 T. 레에세이 세포를 원형질체 형질전환을 이용하여 *pyr2* 유전자, 이종성 프로모터 및 *ace3* 유전자를 함유하는 핵산으로 형질전환시킴으로써 생성하였다. 도 1 및 도 2에 일반적으로 제시된 바와 같이, 두 개의 상이한 프로모터 및 두 개의 상이한 버전의 *ace3* ORF(도 1; *ace3*-SC 및 *ace3*-L)로 네 가지의 상이한 조합을 이용하여 네 개의 상이한 *ace3*-발현 벡터를 구성하였다(도 2a 내지 2d). *hxk1*(헥소키나아제를 암호화하는 유전자) 및 *pk1*(피루베이트 키나아제를 암호화하는 유전자)의 프로모터를 선택하여 *ace3*의 구성성 발현을 추진하였으나, 당업자에 의해 다른 프로모터 또한, 사용 및 선택될 수 있다.

1B. 트리코더마 레에세이 속주 세포

[0185] 다음 실시예에 기재된 T. 레에세이 모 속주 세포는 T. 레에세이 균주 RL-P37(NRRL 수탁번호 15709)로부터 유래되었는데, 이때, 문헌 [Sheir-Neiss and Montenecourt, 1984]에서 일반적으로 설명된 바와 같이, T. 레에세이 *pyr2* 유전자는 결실되었다.

1C. *Ace3* 발현 벡터의 구성

[0187] 위의 발명을 실시하기 위한 구체적인 내용에 기재된 바와 같이, *Ace3*은 최근에 유도 조건 하(즉, 락토오스의 존재 하)에서 셀룰라라아제 및 헤미셀룰라아제 생산에 필요하다고 밝혀진 T. 레에세이 전사 인자이다(Hakkinen et al., 2014). 더욱 상세하게는, Hakkinen 등 (2014)은 T. 레에세이 주 QM6a의 공개적으로 이용 가능한 게놈 서열(genome.jgi.doe.gov/Trire2/Trire2.home.html 참조)에 기초하여 예상 *ace3* ORF를 이용하였는데, 이때, QM6a 예상 주석(단백질 ID 77513)은 2개의 엑손과 1개의 인트론으로 구성된다(예컨대, 도 1 참조).

[0188] 또한, T. 레에세이 주 Rut-C30의 공개적으로 이용 가능한 게놈 서열(genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html 참조)(단백질 ID 98455)로부터 예측된 *ace3* ORF는 세 개의 엑손과 두 개의 인트론을 포함하는 더 긴 단백질 서열(즉, T. 레에세이 QM6a로부터의 (짧은) *ace3*에 비해)을 포함한다(도 1). 더욱 상세하게는, "RUT-C30" 모델에 의해 예측된 개시 코돈은 "QM6a" 모델에서의 그것의 상류에 위치해 있으며, C 말단에 넌센스 돌연변이가 있어(Poggi-Parodi et al., 2014), 더 긴 N 말단 서열과 더 짧은 C 말단 단백질 서열을 가져온다(도 1).

[0189] 본 실시예에서, (QM6a 주석을 기초로 하여, 단백질의 C 말단을 절단한 RUT-C30 넌센스 돌연변이를 포함하는) 짧은 *ace3* ORF(*Ace3*-S) 및 (RUT-C30 주석을 기초로 한) 긴 *ace3* ORF(*Ace3*-L) 둘 다를 클로닝하였다. 도 1에 기재된 바와 같이, 짧은 *ace3*(*Ace3*-S) 및 긴 *ace3*(*Ace3*-L) ORF는 둘 다 RUT-C30에서 발견되는 바와 같이, C 말단 넌센스 돌연변이를 포함한다(도 1). *ace3* ORF의 발현을 추진하기 위해, 이종성 헥소오스 키나아제(*hxk1*) 프로모터 및 이종성 피루베이트 키나아제(*pk1*) 프로모터를 시험하였다.

[0190] 따라서, 표준 분자생물학 절차를 이용하여, 4종의 *Ace3*-발현 벡터 pYL1, pYL2, pYL3 및 pYL4(도 2a 내지 2d)를 구성하였다. 이들 발현 벡터는 대장균에서의 복제 및 선택을 위해 세균의 *ColE1 ori* 및 *AmpR* 유전자를 갖는 벡터 백본을 함유한다. T. 레에세이 *pyr2* 선택 마커 이외에도, T. 레에세이 프로모터 서열(즉, *hxk1* 또는 *pk1*의 프로모터), 그리고 그것의 천연 종결자와 함께 *ace3* ORF(*ace3*-L 또는 *ace3*-SC) 또한, 존재한다. T. 레에세이 프로모터 및 *ace3* ORF를 Q5 고성능 DNA 폴리머라아제(뉴잉글랜드 바이오랩스) 및 아래 표 1에 기재된 프라이머를 이용하여 T. 레에세이 게놈 DNA로부터 PCR 증폭시켰다.

[0191] 각 벡터를 위한 단편의 PCR 증폭에 사용된 구체적인 프라이머는 다음과 같이 나열된다. 벡터 pYL1을 구성하기 위하여, *hxk1* 프로모터를 프라이머 쌍 TP13(서열 번호 7) 및 TP14(서열 번호 8)를 이용하여 증폭시키고, Ace3-L ORF는 프라이머 TP15(서열 번호 9) 및 TP16(서열 번호 10)을 이용하여 증폭시키고, 벡터 백본은 프라이머 쌍 TP17(서열 번호 11) 및 TP18(서열 번호 12)을 이용하여 증폭시켰다. 플라스미드 pYL1의 완전한 서열은 서열 번호 21로서 제공된다.

[0192] 벡터 pYL2를 구성하기 위하여, *hxk1* 프로모터를 TP13(서열 번호 7) 및 TP19(서열 번호 13)의 프라이머 쌍을 이용하여 PCR 증폭시키고, Ace3-SC ORF는 프라이머 TP20(서열 번호 14) 및 TP16(서열 번호 10)을 이용하여 증폭시키고, 벡터 백본은 프라이머 쌍 TP17(서열 번호 11) 및 TP18(서열 번호 12)을 이용하여 증폭시켰다. 플라스미드 pYL2의 완전한 서열은 서열 번호 22로서 제공된다.

[0193] 벡터 pYL3을 구성하기 위하여, *pki1* 프로모터를 TP21(서열 번호 15) 및 TP22(서열 번호 16)의 프라이머 쌍을 이용하여 PCR 증폭시키고, Ace3-L ORF는 프라이머 TP23(서열 번호 17) 및 TP16(서열 번호 10)을 이용하여 증폭시키고, 벡터 백본은 프라이머 쌍 TP17(서열 번호 11) 및 TP24(서열 번호 18)을 이용하여 증폭시켰다. 플라스미드 pYL3의 완전한 서열은 서열 번호 23으로서 제공된다.

[0194] 벡터 pYL4를 구성하기 위하여, *pki1* 프로모터를 TP21(서열 번호 15) 및 TP25(서열 번호 19)의 프라이머 쌍을 이용하여 PCR 증폭시키고, Ace3-SC ORF는 프라이머 TP26(서열 번호 20) 및 TP16(서열 번호 10)을 이용하여 증폭시키고, 벡터 백본은 프라이머 쌍 TP17(서열 번호 11) 및 TP24(서열 번호 18)을 이용하여 증폭시켰다. 플라스미드 pYL4의 완전한 서열은 서열 번호 24로서 제공된다.

[0195] 각 벡터를 위해, 위에 기술된 3개의 PCR 단편을 조립하고, 제조사의 프로토콜에 따라 길은 어셈블리 클로닝 키트(뉴잉글랜드 바이오랩스; 카탈로그 번호: E5510S)를 이용하여 NEB DH5 α 컴파턴트 세포 내로 형질전환시켰다. 생성된 벡터를 생어(Sanger) 시퀀싱을 이용하여 시퀀싱하고, 그것들의 맵을 도 2a 내지 2d에 나타내었다.

표 1

구성체 어셈블리 프라이머

프라이머	서열	서열번호
TP13	TCAGGGTTATTGTCTCATGGCCATTAGGCCTGGCAGGCAGTGGCTCGGACGACATGT	7
TP14	AGAGCCCTGGGCCGGAGCTGCTGAGCCCATTGTTGAATTCTGGCGGGGTAGCTGTTGA	8
TP15	TCAACAGCTACCCGCCAGAATTCAACAATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	9
TP16	TCGTAAATAAACAAAGCGTAACTAGCTAGCGTAGGTTATGCGAGCAACATTGCAAGAAC	10
TP17	GTTTCGTGCAATGTTGCTCGCATAACCTACGCTAGCTAGTTACGCTTGTATTACGA	11
TP18	ACATGTCGTCCGAGCCAGTGCCTGCCAGGCCATAATGCCATGAGACAATAACCTGA	12
TP19	AGGTGTAAGACGGGGAGTAGCGCAGCATTGTTGAATTCTGGCGGGTAGCTGTTGA	13
TP20	TCAACAGCTACCCGCCAGAATTCAACAATGCTGCGCTACTCCCCGCTTACACCT	14
TP21	TCAGGGTTATTGTCTCATGGCCATTAGGCCTAGACTAGCGCCGGTCCCTTATCCCA	15
TP22	AGAGCCCTGGGCCGGAGCTGCTGAGCCATGGTGAAGGGGGCGGCCGGAGCCT	16
TP23	AGGCTCCCGGCCGCCCTTCACCATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	17
TP24	TGGGATAAGGGGACCGGCCGCTAGTCTAGGCCTAAATGCCATGAGACAATAACCTGA	18
TP25	TGTAAGACGGGGAGTAGCGCAGCATGGTGAAGGGGGCGGCCGGAGCCT	19
TP26	AGGCTCCCGGCCGCCCTTCACCATGCTGCGCTACTCCCCGCTTACCA	20

[0196]

1D. *T. 레에세이*의 형질전환

[0197] pYL1, pYL2, pYL3 및 pYL4의 발현 벡터를 *PacI* 효소(뉴잉글랜드 바이오랩스)를 이용하여 선형화하고, 폴리에틸렌 글리콜(PEG) 매개성 원형질체 형질전환(Ouedraogo et al., 2015; Penttila et al., 1987)에 의해 *T. 레에세이* 모 숙주 세포 내로 형질전환시켰다. 형질전환체를 보겔 최소 배지 아가 플레이트 상에서 성장시켜 *pyr2* 마커

에 의해 획득된 우리딘 원형양성을 대상으로 선택하였다. 2회 연속하여 보겔 아가 플레이트에 옮겨 안정적인 형질전환체를 획득하였고, 그 후, 포자 혼탁액의 희석액을 플레이팅하여 단일 콜로니를 획득하였다. pYL1, pYL2, pYL3 및 pYL4를 보유하는 변이체(즉, 변형된) 숙주 세포를 각각 변이체 A4-7, 변이체 B2-1, 변이체 C2-28 및 변이체 D3-1이라 명명하였다.

[0199] **실시예 2**

완효성 미세적정 플레이트(srMTP)에서의 단백질 생산

[0201] 본 실시예는 비유도 조건 하에서 효소를 분비하는 형질전환체(실시예 1 참조)를 확인하는 데 사용되는 스크리닝 법을 설명한다. 예를 들어, 실시예 1에서 획득된 안정적인 형질전환체를 완효성 미세적정 플레이트(slow release microtiter plate, srMTP)에서 시험하였다. 사용된 srMTP는 PCT 국제 공개 제WO2014/047520호에 기술된 바와 같이 제조된 20% 글루코오스(wt/wt) 또는 20% 락토오스(wt/wt)를 함유하는 24-웰 PDMS 엘라스토머 플레이트였다.

[0202] 실시예 1에 기술된 모 *T. 레에세이* 숙주 세포와 변이체 *T. 레에세이* 숙주 세포를 "비유도" 조건 및 "유도" 조건들 다에서 시험하였다. "비유도 조건"에서는, 세포를 20% 글루코오스(wt/wt)를 함유하는 srMTP 중 2.5% 글루코오스(wt/vol)가 첨가된 합성 배지의 1.25 ml 액체 배양액에서 성장시켰다. "유도 조건"에서는, 세포를 20% 락토오스(wt/wt)를 함유하는 srMTP 중 2.5% 글루코오스/소포로오스(wt/vol)가 첨가된 합성 배지의 1.25 ml 액체 배양액에서 성장시켰는데, 여기서 소포로오스와 락토오스는 셀룰라아제 효소 발현을 위한 강력한 유도인자로서 작용한다.

[0203] 글루코오스/소포로오스의 제조는 미국 특허 제US7,713,725호에 기술된 바와 같이 수행하였다. 9 g/L 카사미노산, 5 g/L $(\text{NH}_4)_2\text{SO}_4$, 4.5 g/L KH_2PO_4 , 1 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 33 g/L PIPPS 완충액(pH 5.5에서), 0.25 ml/L *T. 레에세이* 미량 원소를 포함하는 합성 배지를 PCT 국제 공개 제WO2013/096056호에 일반적으로 기술된 바와 같이 제조하였다. *T. 레에세이* 미량 원소는 191.41 g/l 시트르산. H_2O , 200 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 16 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.56 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1.2 g/L $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 및 0.8 g/L H_3BO_3 을 함유한다. 모든 srMTP를 280 rpm에서 계속하여 진탕하면서 대략 120시간 동안 28°C에서 인큐베이션시켰다.

[0204] 인큐베이션 후, 모든 배양물로부터의 상청액을 수확하고, 폴리아크릴아미드 겔 전기영동(PAGE)을 이용하여 분석하였다. 동일한 부피의 배양 상청액을 90°C에서 15분 동안 환원 환경을 겪게 한 후, 로딩 염료를 첨가한 후, MOPS-SDS 완충액을 이용하여 4 내지 12% NuPage™(인비트로젠(Invitrogen), 미국 캘리포니아주 칼스배드 소재) 폴리아크릴아미드 겔 상에서 분리하였다. 겔을 SimplyBlue™(인비트로젠)으로 염색하고, 영상화하였다(도 3 참조).

[0205] 도 3에 도시된 바와 같이, 시험한 숙주 세포 전부가 유도인자(즉, 배지 중의 소포로오스 및 srMTP 중의 락토오스)의 존재 하에 다량의 단백질을 분비하였다. 그러나 글루코오스가 유일한 탄소원인, 유도인자(즉, 소포로오스 또는 락토오스)의 부존재 하에서는, Ace3-L ORF를 발현하는 변이체 숙주 세포만이(즉, 변이체 A4-7 및 변이체 C2-28) 분비 단백질을 생산하였고, 모 숙주 세포 또는 Ace3-SC ORF를 발현하는 변이체 세포(즉, 변이체 B1-1 및 변이체 D3-1)는 PAGE의 검출 한계 이상으로 세포 외 단백질을 생산하지 않았다. 이 결과는 Ace3-L(즉, Ace3-S와는 대조적으로)은 *T. 레에세이*에서 유도인자 없이 단백질을 생산할 수 있음을 명백히 증명한다.

[0206] 정제된 효소를 기준물질로서 이용하여 Zorbax C3 역상(RP) 분석에 의해 분비 단백질의 상대적인 농도를 결정하였다. 예를 들어, 위에 기술된 숙주 세포의 분비 단백질 프로파일을 이 방법을 이용하여 분석하였는데, 이때, 모든 숙주 세포(즉, 모 세포 및 변이체 세포)가 유도 조건 하에서는 비슷한 셀룰라아제 단백질 프로파일을 생성하였으며, 셀룰라아제는 대략 40% CBH1, 20% CBH2, 10% EG1 및 7%의 EG2로 구성되었음이 관찰되었다. 비유도 조건 하에서는, 셀룰라아제 효소는 모 세포 및 변이체 Ace3-S 발현 숙주 세포에서는 검출 미만이었다. 대조적으로, 놀랍게도 변이체 Ace3-L 발현 숙주 세포(즉, 변이체 A4-7 및 C2-28)는 유도 조건 하에서와 비슷한 비율의 셀룰라아제 효소를 생산하였음이 발견되었다.

[0207] 간략하게는, 이 분석 방법은 다음과 같이 수행되었다: 상청액 샘플을 50 mM 아세트산 나트륨 완충액, pH 5.0 중에 희석시키고, 20 ppm EndoH를 첨가하여 탈글리코실화하고, 3시간 동안 37°C에서 인큐베이션시켰다. 10 μl 90% 아세토니트릴을 100 μL EndoH 처리 샘플에 첨가하고, 주입 전에 0.22 μm 필터를 통과시켰다. 애질런트(Agilent) Zorbax300 SB C3 RRHD 1.8um (2.1x100mm) 컬럼을 구비한 DAD 검출(애질런트 테크놀로지스) HPLC와 함께 애질런트 1290을 이용하였다. 컬럼은 전개 완충액(running buffer) A로서 밀리큐(MilliQ) 물 중 0.1% 트리

플루오로아세트산(TFA), 전개 완충액 B로서 아세토니트릴 중 0.07% TFA를 이용하여 1.0 mL/분의 유속으로 60 °C에서 작동시켰다. DAD 검출기는 4 nm 창과 함께, 220 nm 및 280 nm에서 작동시켰다. 주입 부피는 10 μL였다.

[0208] 추가적으로, 변이체 Ace3-L 발현 숙주 세포는 글루코오스를 함유하는 srMTP 중에서 대략 20 내지 30% 총 세포의 단백질을(즉, 락토오스를 함유하는 srMTP 중에서 생산된 총 세포 외 단백질과 비교하여) 생산하였음이 눈에 띄었다. 이러한 상대적으로 낮은 발현은 글루코오스를 함유하는 srMTP에서의 높은 글루코오스 공급 속도로 인한 것일 수 있다. 예를 들어, 가장 높은 셀룰라아제 및 헤미셀룰라아제 생산 속도는 종종 낮은 성장 속도에서 관찰된다는 점은 잘 확립되어 있다(Arvas et al., 2011). 그럼에도, srMTP 성장 분석법은 단백질 생산을 위한 안정적인 콜로니를 스크리닝하는, 상대적으로 고성능 분석법이었다.

[0209] 종합하면, Ace3-L ORF를 발현하는 변이체 T. 레에세이 숙주 세포는 비록 낮은 단백질 생산 속도라도, 유도인자의 부존재 하에 셀룰라아제 및 헤미셀룰라아제를 생산할 수 있었다. 더욱 상세하게는, 이 낮은 생산 속도는 아래의 실시예 3 및 실시예 4에 나타난 바와 같이, 숙주 세포의 생산 능력보다는, srMTP 성장법과 관련이 있었다.

실시예 3

진탕 플라스크에서의 단백질 생산

[0210] 위에서 제시된 srMTP 결과를 더 분석하고 검증하기 위하여, 모 숙주 세포 및 변이체 Ace3-L 발현 숙주 세포를 진탕 플라스크 내 50 mL 액침 배양액 중 유도인자 기질의 존재 및 부존재 하에서 성장시켰다. 더욱 구체적으로는, 모 T. 레에세이 숙주 세포, 변이체 A4-7 세포 및 변이체 C2-28 세포를 액침(액체) 배양액 중에서 유도 조건(즉, 탄소원으로서 글루코오스/소포로오스) 및 비유도 조건(즉, 탄소원으로서 글루코오스) 하에서 성장시키고, 각각의 세포 외(분비) 단백질 생산 수준을 비교하였다. 간략하게는, 각각의 숙주 세포(즉, T. 레에세이 모 숙주 세포, 변이체 A4-7 숙주 세포 및 변이체 C2-28 숙주 세포)의 균사를 바탕 칸막이가 있는 250-mL 엘렌마이어 플라스크 내의 50-mL의 YEG 배양액에 따로따로 첨가하였다. YEG 배양액은 5g/L 효모 추출물 및 22 g/L 글루코오스를 함유한다. 세포 배양물을 48시간 동안 성장시키고, 이어서 또 다른 24시간 동안 새로운 YEG 내로 계대배양하였다. 그런 다음, 이들 종 배양물을 바탕 칸막이가 있는 250 mL 진탕 플라스크 내의 1.5% 글루코오스(비유도 조건)가 첨가된 50 mL의 합성 배지, 또는 1.5% 글루코오스/소포로오스(유도 조건)가 첨가된 50 mL의 합성 배지 내로 접종하였다.

[0211] 모든 진탕 플라스크를 200 rpm에서 계속하여 진탕하면서 28 °C에서 인큐베이션시켰다. 3일의 인큐베이션 후, 모든 세포 배양물로부터의 상청액을 수확하고, 위의 실시예 2에 기술된 바와 같이 PAGE를 이용하여 분석하였다. 상청액 중 총 단백질을 바이오라드(Bio-Rad) 시약(서모 사이언티픽(Thermo Scientific®); 카탈로그 번호: 23236) 및 표준물질로서 소 혈청 알부민(BSA)의 다섯 가지 희석액을 이용하여 595 nm에서 브래드포드 염료 결합 분석법으로 측정하였다. 글루코오스 농도를 고성능 액체 크로마토그래피(HPLC) 분석으로 측정하였고, 3일간의 인큐베이션 후 어떠한 글루코오스도 배양물에서 검출되지 않았다.

[0212] 도 4에 도시된 바와 같이, 모 (대조군) T. 레에세이 세포는 글루코오스/소포로오스를 함유하는 합성 배지(유도)에서는 464 μg/mL 총 분비 단백질을 생산하였고, 글루코오스를 함유하는 합성 배지(비유도)에서는 오로지 140 μg/L의 총 분비 단백질을 생산하였다. 이와 대조적으로, 변이체 A4-7 세포와 C2-28 세포는 유도 및 비유도 조건 하에서 비슷한 양의 분비 단백질을 생산하였는데, 둘 다 소포로오스(유도)를 이용한 모 (대조군) 세포에서 생산된 분비 단백질보다 더 많다. 따라서, 이러한 결과는 Ace3-L ORF를 보유하는 변형된 세포(즉, 변이체 A4-7 및 C2-28 세포)는 유도인자의 부존재 하에서 세포 외 단백질을 생산할 뿐만 아니라, 이들 변이체 세포는 이러한 유도 조건 하에서 모 (대조군) T. 레에세이 세포보다 더 많은 총 단백질을 생산함을 증명한다.

실시예 4

소규모 유가식 발효에서의 단백질 생산

[0213] 본 실시예는 변이체 Ace3-L 발현 세포(즉, 변이체 A4-7 및 C2-28 세포)가 소규모 발효에서 유도인자 기질의 존재 및 부존재 하에서 비슷한 양의 셀룰라아제 및 헤미셀룰라아제 효소를 생산하였음을 보여준다. 더욱 상세하게는, 2L 생물반응기에서 시트르산 최소 배지 내의 종 배양물을 이용하여, 일반적으로 미국 특허 제7,713,725호에 기술된 바와 같이, T. 레에세이 발효를 수행하였다. 더욱 구체적으로는, 발효 중에, 모든 배양물로부터의 상청액을 상이한 시점에서 수확하고, 동일한 부피의 배양 상청액을 대상으로 PAGE 분석하였다. 도 5에 나타난 바와 같이, 모 (대조군) T. 레에세이 세포는 소포로오스 유도인자의 존재 하에서만 분비 단백질을 생산하였다. 이와 대조적으로, 변이체 A4-7 및 C2-28 세포(도 5)는 유도 및 비유도 조건 둘 다에서 비슷한 양의 단백질을 생산하

였다.

[0218] **실시예 5**

[0219] ***Ace3* 발현을 위한 이종성 프로모터**

[0220] 본 실시예는 *ace3*-L의 발현을 추진하는 13종의 상이한 프로모터를 이용한 "비유도" 조건 하에서의 증진된 단백질 생산을 증명한다. 더욱 상세하게는, 원형질체 형질전환을 이용하여, *pyr2* 유전자, 이종성 프로모터 및 *ace3*-L 유전자를 함유하는 텔로미어 벡터로 모 *T. 레에세이* 세포를 형질전환하여 *ace3*-L 유전자를 발현하는 *T. 레에세이* 세포를 생성하였다.

[0221] 따라서, 13종의 *T. 레에세이* 프로모터가 *ace3*-L ORF의 발현을 추진하기 위하여 선택되었는데, 이때, 시험한 13종의 프로모터에는 (i) 포름아미다아제 유전자(*rev3*; 단백질 ID 103041) 프로모터(서열 번호 15), (ii) β -자일로시다아제 유전자(*bxl*; 단백질 ID 121127) 프로모터(서열 번호 16), (iii) 트랜스케틀라아제 유전자(*tkl1*; 단백질 ID 2211) 프로모터(서열 번호 17), (iv) 미지의 기능의 유전자(단백질 ID 104295) 프로모터(서열 번호 18), (v) 옥시도리덕타아제 유전자(*dld1*; 단백질 ID 5345) 프로모터(서열 번호 19), (vi) 자일라나아제 IV 유전자(*xyn4*; 단백질 ID 111849) 프로모터(서열 번호 20), (vii) α -글루코로니다아제 유전자(단백질 ID 72526) 프로모터(서열 번호 21), (viii) 아세틸 자일란 에스테라아제 유전자 1(*axe1*; 단백질 ID 73632) 프로모터(서열 번호 22), (ix) 헥소오스 키나아제 유전자(*hxk1*; 단백질 ID 73665) 프로모터(서열 번호 23), (x) 미토콘드리아 수송 단백질 유전자(*dic1*; 단백질 ID 47930) 프로모터(서열 번호 24), (xi) 올리고펩티드 수송 유전자(*opt*; 단백질 ID 44278) 프로모터(서열 번호 25), (xii) 글리세롤 키나아제 유전자(*gut1*; 단백질 ID 58356) 프로모터(서열 번호 26) 및 (xiii) 피루베이트 키나아제 유전자(*pk11*; 단백질 ID 78439) 프로모터(서열 번호 27)가 포함되나, 이에 한정되지 않는다. 단백질 ID(PID) 번호는 genome.jgi.doe.gov/Trire2/Trire2.home.html로부터 온 것이다. 따라서, 위에 기술된 13종의 프로모터를 발현을 추진하기 위해 선택하였는데, 이의 유전자들은 일반적으로 글루코오스 농도가 높을 때 성장 동안 낮은 수준으로 발현되고, 글루코오스 농도가 낮을 때, 또는 소포로오스 유도 조건 하에 있을 때는 높은 수준으로 발현되기 때문이다.

[0222] 아래 표 2는 13종의 프로모터 및 이의 발현 벡터를 요약한 것으로, 발현 벡터는 표준적인 분자생물학 절차를 이용하여 구성되었다. 더욱 상세하게는, 본 실시예에서 시험한 발현 벡터(표 2)는 대장균에서의 복제 및 선택을 위해 세균의 *ColE1 ori* 및 *AmpR* 유전자 및 사카로미세스 세레비시애(*Saccharomyces cerevisiae*)에서의 복제 및 선택을 위해 2μ ori 및 *Ura3* 유전자를 함유하는 벡터 백본을 포함한다. 또한, *T. 레에세이* 텔로미어 서열 ("TrTEL"), *T. 레에세이* *pyr2* 선택 마커, *T. 레에세이* 프로모터 서열 및 *ace3*-L ORF가 그것의 천연 종결자 서열과 함께 존재한다. *dic1* 프로모터를 함유하는 벡터 pYL8을 도시하는 대표적인 벡터 맵을 도 7에 나타내었다. 따라서, 나머지 벡터(예컨대, pYL9, pYL12 등)는 상이한 프로모터 서열을 제외하고는 도 7에 제시된 동일한 서열을 갖는다.

표 2

***ace3*-L 발현 추진을 위한 상이한 전균 프로모터를 활용한 *ace3*-L 발현 구성체**

벡터 #	프로모터	<i>ace3</i> ORF
pYL7	<i>opt</i> (서열 번호 25)	<i>ace3</i> -L
pYL8	<i>dic1</i> (서열 번호 24)	<i>ace3</i> -L
pYL9	<i>gut1</i> (서열 번호 26)	<i>ace3</i> -L
pYL12	<i>hxk1</i> (서열 번호 23)	<i>ace3</i> -L
pYL13	<i>pk11</i> (서열 번호 27)	<i>ace3</i> -L
pYL22	<i>rev3</i> (서열 번호 15)	<i>ace3</i> -L
pYL23	PID 104295 (서열 번호 18)	<i>ace3</i> -L
pYL24	<i>tkl1</i> (서열 번호 17)	<i>ace3</i> -L
pYL25	<i>bxl</i> (서열 번호 16)	<i>ace3</i> -L
pYL27	<i>dld1</i> (서열 번호 19)	<i>ace3</i> -L
pYL28	<i>xyn4</i> (서열 번호 20)	<i>ace3</i> -L
pYL29	PID 72526 (서열 번호 21)	<i>ace3</i> -L
pYL30	<i>axe1</i> (서열 번호 22)	<i>ace3</i> -L

[0224] 폴리에틸렌 글리콜(PEG) 매개성 원형질체 형질전환(Ouedraogo *et al.*, 2015; Penttila *et al.*, 1987)에 의해 (비기능성 *pyr2* 유전자를 포함하는) *T. 레에세이* 모 숙주 균주 내로 발현 벡터를 삽입하였다(형질전환시켰다). 형질전환체를 보겔 최소 배지 아가 플레이트 상에서 성장시켜 *pyr2* 마커에 의해 획득된 우리던 원영양성을 대상으로 선택하였다. 보겔 아가 플레이트 상에서의 2회 연속 이송에 이어서, 비선택적 PDA 플레이트 상에서의 2회 연속 성장 및 보겔 아가 플레이트 상에서의 1회로써 안정적인 형질전환체를 획득하였고, 그 후, 포자 혼탁액의 희석액을 플레이팅하여 단일 콜로니를 획득하였다.

[0225] 위에 기술된 모 *T. 레에세이* 숙주 세포와 형질전환된 (딸) *T. 레에세이* 숙주 세포를 "비유도" 조건 및 "유도" 조건 둘 다에서 시험하였다. 예를 들어, "비유도 조건"에서는, 세포를 보통의 24웰 미세적정 플레이트(MTP) 중 2.5% 글루코오스(wt/vol)가 첨가된 합성 배지의 1.25 ml 액체 배양액에서 성장시켰다. "유도 조건"에서는, 세포를 MTP 중 2.5% 글루코오스/소포로오스(wt/vol)가 첨가된 합성 배지의 1.25 ml 액체 배양액에서 성장시켰는데, 여기서 소포로오스는 셀룰라아제 효소 발현을 위한 강력한 유도인자로서 작용한다. 인큐베이션 후, 모든 배양물로부터의 상청액을 수확하고, 바이오라드 시약(서모 사이언티픽; 카탈로그 번호: 23236) 및 표준물질로서 소 혈청 알부민(BSA)의 다섯 가지 희석액을 이용하여 595 nm에서 브래드포드 염료 결합 분석법으로 총 분비 단백질을 측정하였다.

[0226] 표 3에 나타난 바와 같이, 모 *T. 레에세이* 세포는 소포로오스 유도인자의 존재 하에서만 높은 수준의 분비 단백질을 생산하였다. 이와 대조적으로, 13종의 상이한 프로모터로부터 추진된 *Ace3-L*을 포함하고 발현하는 변이체 (딸) *T. 레에세이* 세포는 유도 및 비유도 조건 둘 다에서 비슷한 양의 분비 단백질을 생산하였다. 표 3에 나타난 바와 같이, 각각의 변형된 (딸) 균주에 대한 단백질 수준을 글루코오스/소포로오스(Glu/Sop) 유도 조건 하에서의 모 균주(LT4)에 의해 생산된 단백질 (농도)에 대한 비율로서 제시하였다.

표 3

유도("GLU/SOP") 및 비유도("GLU") 조건 하에서 변형된 *T. 레에세이*(딸) 균주에 대한 *T. 레에세이* 모 균주(LT4)의 총 분비 단백질

주 ID	프로모터	Glu/Sop ¹	Glu ²
LT4(모)	해당 없음	1.00	0.20
LT82	<i>opt</i>	1.16	1.07
LT83	<i>dic1</i>	1.35	1.12
LT85	<i>gut1</i>	0.95	1.08
LT86	<i>hxk1</i>	0.96	0.73
LT87	<i>pkl1</i>	1.16	0.98
LT149	<i>rev3</i>	0.96	0.76
LT150	PID 104295	0.94	0.41
LT151	<i>tkl1</i>	1.02	1.01
LT152	<i>bxl</i>	1.00	1.04
LT154	<i>dld1</i>	1.00	0.58
LT155	<i>xyn4</i>	0.95	0.95
LT156	PID 72526	0.95	0.89
LT157	<i>axel</i>	1.01	0.85

Glu/Sop¹은 "글루코오스/소포로오스"; 유도 조건의 약어이다.

Glu²는 "글루코오스"; 비유도 조건의 약어이다.

[0227]

[0228] 추가적으로, 위에 기술된 *T. 레에세이* 모 균주 및 이의 형질전환체를 실시예 3에 일반적으로 기술된 바와 같이 진탕 플라스크 실험에서 추가로 시험하였다. 예를 들어, *T. 레에세이* 딸 균주 "LT83"의 대표적인 결과를 도 8에 나타내었는데, 이때, 딸 균주 LT83은 *dic1* 프로모터(서열 번호 28)로부터 추진된 *ace3-L* ORF를 포함한다. 도 8에 나타난 바와 같이, 모 (대조군) *T. 레에세이* 세포는 소포로오스("Sop") 유도인자의 존재 하에서만 분비 단백질을 생산하였지만, 딸 균주 LT83은 유도("Sop") 및 비유도("Glu") 조건 둘 다에서 비슷한 양의 분비 단백질을 생산하였다.

[0229]

마찬가지로, 실시예 4에 일반적으로 기술된 바와 같이 소규모 발효를 수행하였다. 더욱 상세하게는, 도 9에 나타난 바와 같이, 모 (대조군) *T. 레에세이* 균주는 소포로오스 유도인자("Sop")의 존재 하에서만 분비 단백질을

생산하였지만, 딸 균주 LT83은 유도("Sop") 및 비유도("Glu") 조건 둘 다에서 비슷한 양의 단백질을 생산하였다.

[0230] 실시예 6

[0231] *T. 레에세이* *ACE3* 과발현 구성체의 클로닝

*T. 레에세이*의 게놈 서열이 합동 게놈 연구소(Joint Genome Institute, <https://genome.jgi.doe.gov/>)에서 공개적으로 이용 가능하더라도, *ace3* 암호화 영역의 5' 말단의 위치는 명백하지 않다. 예를 들어, DNA 서열이 동일하더라도, 돌연변이 균주 Rut-C30(genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html)과 야생형 균주 QM6a(genome.jgi.doe.gov/Trire2/Trire2.home.html) 사이에서 합동 게놈 연구소의 DNA 서열의 주석은 차이가 있었다. QM6a 사례에서는, *ace3* 암호화 영역의 5' 말단은 도 11, 화살표 3으로 도시된 바와 같이, 엑손 3의 상류(5')이자 인트론 2의 내부인 것으로 제시되었다. Rut-C30 사례에서는, *ace3* 암호화 영역의 5' 말단은 엑손 2의 내부에 있다(도 11, 화살표 2). 게놈 DNA 서열 및 추가적인 cDNA 서열의 추가적인 분석은 도 11에 도시된 바와 같이 엑손 1과 인트론 1의 가능성 있는 존재를 시사하였다. 또한, Rut-C30의 *ace3* 암호화 영역의 3' 말단은 야생형 단리물 QM6a의 서열(도 11, 화살표 5)에 비해 미성숙 종결 코돈(도 11, 화살표 4)을 생성하는 돌연변이를 포함한다. 따라서, 본 실시예에서, *ace3* 유전자의 이러한 상이한 가능성 있는 형태들의 과발현 효과를 본 설명에 기술된 바와 같이 실험적으로 연구하였다(예컨대, 도 11, 도 12 및 도 13 내지 18 참조).

따라서, 도 12에 도시된 서로 다른 형태의 *ace3*을 *T. 레에세이*에서 과발현시켰는데, 이때, *T. 레에세이* *ace3* 유전자를 위한 과발현 벡터는 *T. 레에세이*의 글루코아밀라아제 유전자좌(*gla1*)에서 *ace3*의 표적화된 통합을 가능하게 하도록 설계되었다. 더욱 상세하게는, 모든 플라스미드(벡터) 구성체에서, *T. 레에세이* *ace3* 유전자를 *T. 레에세이* *dic1* 프로모터의 제어 하에, 그리고 천연 *ace3* 종결자로 발현시켰다. 이 벡터는 *T. 레에세이* 형질전환체의 선택을 위해 그것의 천연 프로모터 및 종결자와 함께 *pyr4* 마커를 더 포함한다. *pyr4* 프로모터의 반복부를 포함시켜, *gla1* 유전자좌에서의 통합 이후에 *pyr4* 유전자의 제거를 가능하게 하였다. 대장균에서의 복제 및 증폭을 가능하게 하는 벡터 백본은 EcoRI-XhoI 소화 pRS426이었다(Colot et al., 2006). 표적화된 통합에 필요한 *T. 레에세이* *gla1* 유전자좌의 5' 및 3' 플랭크, *dic1* 프로모터, 상이한 형태의 *ace3* 암호화 영역 및 종결자를 표 4에 제공된 프라이머를 이용하여 PCR에 의해 생산하였다. 플랭킹 단편을 위한 주형은 야생형 *T. 레에세이* QM6a로부터의 게놈 DNA였다(ATCC 수탁번호 13631).

표 4

DNA 단편 생성에 사용된 프라이머

프라이머	서열
Gla.5F (SID: 49)	GTAACGCCAGGGTTTCCCAGTCAGCAGGTTAAACTCCATACGCAGCAAACATGGGCTTGGGC
Gla.5R (SID: 50)	GTACGAGTACTAGGTGTGAAGATTCCGTCAAGCTTGGCGGAATGAAGGAGGATGTGTGAGAGG
DICprom.F (SID: 51)	CACACATCCTCCTCATCCGCCAAGCTTGACGGAATCTTACACCTAGTACTCGTAC
Ace3RutC.R (SID: 52)	TGACATTTTGTGTTCCAACACAGCATGCTTAGTCCGACGCCCTCGAGTC CAGCC
Ace3term.F (SID: 53)	CTGGACTCGAAGGCGTCGGACTAACAGCATGCTGTGTTGAAACAACAAAAATGTC
Ace3term.R (SID: 54)	GCAGAGCAGCAGTAGTCGATGCTATTAAATTAAAGTAGGTTATGCGAGCAACATTG
Gla.3F (SID: 55)	CTCAGCCTCTCTCAGGCCATCAGCCGCCGCTGAATCGCAAGGGTAGTACTAG
Gla.3R (SID: 56)	GCGGATAACAATTACACAGGAAACAGCGTTAAACCACATGCCAGAGTTCGATGCGCAAG
Ace3_nointron.F (SID: 57)	GTACCTCAGCGCTGTCGATAGCTGCACGCACTGCCGCGATGCCACGTGCAGTGCAC
Ace3_nointron.R (SID: 58)	GTGCACTGCACGTGGCATCGCGCAGTGCCTGCAGCTATCGACAGCGCTGAGGTACTC
Ace3QM.F (SID: 59)	GCGGCGCTTCCGCTGTCGTAACTATGCTGCGCTACTCCCCCGTCTTACGTAAGACGGGGAGTACGCGAGCATAGTTACGACAGCGGAAGCGCCGCCCTTAAAGTG
DICprom_QM.R (SID: 60)	GGCGGGCGCTTCCGCTGTCGTAACTATGGGCTCAGCAGCTCCGGCCCAGGGCTCGCTGGAGCTGAGCCCATAGTTACGACAGCGGAAGCGCCGCCCTTATAAAG
Ace3RutC.F (SID: 61)	GGCGGGCGCTTCCGCTGTCGTAACTATGGGCTCAGCAGCTCCGGCCCAGGGCTCGCTGGAGCTGAGCCCATAGTTACGACAGCGGAAGCGCCGCCCTTATAAAG
DICprom_rutc.R (SID: 62)	GGCGGGCGCTTCCGCTGTCGTAACTATGGGCTCAGCAGCTCCGGCCCAGGGCTCGCTGGAGCTGAGCCCATAGTTACGACAGCGGAAGCGCCGCCCTTATAAAG
Ace3cDNA.F (SID: 63)	CAGCTGCTGCCGCGGCCGCTGTGCCATAGTTACGACAGCGGAAGCGCCGCCCTTATAAAG
DICprom_cDNA.R (SID: 64)	TATAAG

위의 표에서 "SID"는 "서열 식별 번호"의 약어, 예컨대, "서열 번호"이다.

[0234]

[0235] *dic1* 프로모터, *ace3* 유전자 및 종결자의 경우, 주형은 이를 단편을 전달하는 초기의 플라스미드 pYL8(실시예 5 참조)이었다. 선택 마커(*pyr4*)는 초기 플라스미드로부터 *NotI* 소화로 획득하였다. 원하는 DNA 단편을 생성하는데 사용된 PCR 프라이머를 표 4에 나타내었다. PCR 생성물 및 소화된 단편을 아가로스 겔 전기영동을 이용하여 분리하였다. 제조사의 프로토콜에 따라 겔 추출 키트(퀴아젠(Qiagen))로 겔로부터 정화한 단편을 단리하였다. (WO2013/102674로 공개된) PCT/EP2013/050126에 기술된 바와 같이 효모 상동 재조합 방법을 이용하여 위에 기술된 단편들로 플라스미드를 구성하였다. 플라스미드를 효모로부터 구하여 대장균에 형질전환시켰다. 소수의 클론들을 선택하고, 플라스미드 DNA를 단리하고, 시퀀싱하였다. 플라스미드의 개요를 표 5에 제시하였다.

표 5

과발현 플라스미드

플라스미드	코드	클로닝된 유전자	선택	유전자좌	비고	유전자 서열 번호	단백질 서열 번호
B7683	SC	ace3 SC 형	PYR4	GLA	QM6a N-term; RutC-30 C-term	7	8
B7684	S	ace3 S 형	PYR4	GLA	QM6a N-term; QM6a C-term	1	3
B7709	L	ace3 L 형	PYR4	GLA	RutC-30 N-term; RutC-30 C-term	4	6
B7752	LC	ace3 LC 형	PYR4	GLA	RutC-30 N-term; QM6a C-term	9	10
B7778	EL	ace3 EL 형	PYR4	GLA	RutC-30 N-term; RutC-30 C-term	11	12
B7779	LN	ace3 LN 형	PYR4	GLA	RutC-30 N-term; RutC-30 C-term	13	14

[0236]

[0237] 따라서, 표 5에 제시된 구성체는 *ace3* 유전자의 상이한 형태를 가짐으로써 차이가 있다. SC 형태는 엑손 3 및 4뿐만 아니라, 인트론 3도 포함하는 유전자의 짧은 형태이다(도 12 및 도 13, 서열 번호 7 참조). 더욱 상세하게는, 서열 번호 7의 SC 형태는 1,713bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함한다. SC 형태의 (3' 말단) C 말단(즉, 엑손 4)은 *T. 레에세이* RutC-30 주와 동일한 돌연변이를 갖는다.

[0238]

S 형태는 엑손 3 및 4뿐만 아니라, 인트론 3도 포함하나, 엑손 4의 (3' 말단) C 말단의 돌연변이는 없는 유전자의 짧은 형태이다(도 12 및 도 14, 서열 번호 1 참조). 더욱 상세하게는, S 형태는 1,713bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4를 포함한다. 이러한 형태들(즉, "SC" 및 "S") 둘 다에서, 번역 개시 코돈은 *T. 레에세이* QM6a 주에 대해 주석 달린 바와 같이, 긴 인트론 2에 있고(도 12 참조), "SC" 및 "S" 형태 둘 다는 추정상의 DNA 결합 도메인에 대한 암호화 영역의 누락 부분이다.

[0239]

L 형태는 엑손 2, 3 및 4뿐만 아니라, 인트론 2(긴 인트론) 및 인트론 3도 포함하는 유전자의 긴 형태이다(예컨대, 도 12 및 도 15, 서열 번호 4 참조). 더욱 상세하게는, L 형태는 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함한다. "L" 형태의 (3' 말단) C 말단은 *T. 레에세이* RutC-30 주와 동일한 돌연변이를 갖는다.

[0240]

LC 형태는 엑손 2, 3 및 4뿐만 아니라, 인트론 2(긴 인트론) 및 인트론 3도 포함하는 유전자의 긴 형태이다(예컨대, 도 12 및 도 16, 서열 번호 9 참조). 더욱 상세하게는, LC 형태는 258 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 177 bp 엑손 4를 포함한다. LC 형태는 C 말단의 돌연변이가 없다. "L" 및 "LC" 형태를 둘 다에서, 번역 개시 코돈은 RutC-30 주에 대하여 JGI에서 주석 달린 바와 같이, 엑손 2 내에 있다.

[0241]

EL 형태는 엑손 1, 2, 3 및 4뿐만 아니라, 인트론 1, 인트론 2(긴 인트론) 및 인트론 3도 포함하는 *ace3* 유전자의 매우 긴 버전이다(예컨대, 도 12 및 도 17, 서열 번호 11 참조). 더욱 상세하게는, EL 형태는 61 bp 엑손 1, 142 bp 인트론 1, 332 bp 엑손 2, 423 bp 인트론 2, 1,635 bp 엑손 3, 148 bp 인트론 3 및 144 bp 엑손 4를 포함한다. EL 형태의 (3' 말단) C 말단은 *T. 레에세이* RutC-30 주와 동일한 돌연변이를 갖는다.

[0242]

LN 형태는 엑손 2, 3 및 4뿐만 아니라, 인트론 3도 함유하나, 인트론 2는 결여된 유전자의 긴 형태이다(예컨대, 도 12 및 도 18, 서열 번호 13 참조). LN 형태의 (3' 말단) C 말단은 *T. 레에세이* RutC-30 주와 동일한 돌연변이를 갖는다. 따라서, 위에 기술된 바와 같이, *ace3*의 L, LC, LN 및 EL 형태는 완전한 추정상의 DNA 결합 도메인을 암호화한다.

[0243]

T. 레에세이 RL-P37 균주 내로의 형질전환

[0244]

표 5에 제시된 플라스미드 전부를 *Msp*I로 소화시켜 표적화된 통합을 위한 단편들을 방출하고, 아가로스 겔 전기 영동으로 분리하였다. 예를 들어, 도 19는 *T. 레에세이*의 형질전환에 사용된 대표적인 단편 내의 DNA 서열들의 배열을 보여주는 도해를 제공한다. 제조사의 프로토콜에 따라 겔 추출 키트(퀴아젠)를 이용하여 겔로부터 정화

한 단편을 단리하였다. 대략 10 μ g의 정제된 단편을 이용하여 T. 레에세이 RL-P37 주의 *pyr4*- 돌연변이체의 원형질체를 형질전환하였다. 원형질체의 제조 및 형질전환은 *pyr4* 선택을 이용하여 PCT 공개 제WO2013/102674호에 기술된 바와 같이 수행하였다.

[0245]

형질전환체를 선택적 배지 플레이트에 스트리킹하였다. 표 6에 열거된 프라이머를 이용하여 PCR에 의한 정확한 통합을 위해 성장하는 클론들을 스크리닝하였다. 예상한 신호를 제공하는 클론들을 단일 세포 클론까지 정제하고, 표 6에 열거된 프라이머를 이용하여 PCR뿐만 아니라, 서던 블로팅(데이터 미도시)에 의해서도, 정확한 통합 및 클론 순도를 위해 다시 스크리닝하였다.

표 6

PCR 프라이머

프라이머	서열	서열 번호
Gla1_5creen.F	GCTGGAAAGCTGCTGAGCAGATC	65
DICprom.R	GTGCCAGCATTCCCCAGACTCG	66
T061_pyr4_orf_screen	TTAGGCGACCTCTTTTCCA	67
Gla1_3creen.R	GCCGCTCAGGCATACGAGCGAC	68
DICprom.F2	CTCTGGTCGGCCTGCCGTTG	69
ace3.R	TGAGTATAGCGGCTGACTTGTG	70

[0246]

상이한 *ace3* 형질전환체의 배양

[0247]

표 7의 균주들을 탄소원으로서 2% 락토오스 또는 2% 글루코오스를 함유하는 액체 배지 중에 24-웰 미세적정 플레이트에서 성장시켰다. 배지의 다른 구성요소들은 0.45% KH2PO4, 0.5% (NH4)2SO4, 0.1% MgSO4, 0.1% CaCl2, 0.9% 카사미노산, 0.048% 시트르산xH2O, 0.05% FeSO4x7H2O, 0.0003% MnSO4xH2O, 0.004% ZnSO4x7H2O, 0.0002% H3BO3 및 0.00014% CuSO4x5H2O였다. 100 mM PIPPS(칼바이오켐(Calbiochem))을 포함시켜 pH를 5.5로 유지시켰다.

표 7

트리코더마 레에세이 주

코드	균주 명칭	선택
M1904	RL-P37, 모 균주	
M2015	<i>ace3</i> SC 클론 2-1	<i>Pyr4</i>
M2016	<i>ace3</i> SC 클론 28-3	<i>Pyr4</i>
M2017	<i>ace3</i> S 클론 9-1	<i>Pyr4</i>
M2018	<i>ace3</i> S 클론 20-1	<i>Pyr4</i>
M2019	<i>ace3</i> L 클론 16-1	<i>Pyr4</i>
M2020	<i>ace3</i> L 클론 18-1	<i>Pyr4</i>
M2021	<i>ace3</i> LC 클론 52-1	<i>Pyr4</i>
M2022	<i>ace3</i> LC 클론 14-4	<i>Pyr4</i>
M2023	<i>ace3</i> EL 클론 3-3	<i>Pyr4</i>
M2024	<i>ace3</i> EL 클론 4-5	<i>Pyr4</i>
M2025	<i>ace3</i> LN 클론 3-3	<i>Pyr4</i>
M2026	<i>ace3</i> LN 클론 4-1	<i>Pyr4</i>

[0248]

80 % 습도의 인포스(Infors) HT 마이크로톤 진탕기에서 28°C 및 800 RPM에서 배양을 수행하였다. 배양물의 샘플링을 3 내지 7일에 수행하였다. 제조사의 프로토콜에 따라 바이오라드 단백질 분석을 이용하여 배양물 상청액으로부터 총 분비 단백질의 양을 측정하였다. 두 배지에서, RutC-30 C 말단 돌연변이가 있는 *ace3* L, EL 및 LN 형태의 과발현은 총 단백질의 생산을 향상시켰다. 탄소 공급원으로서 락토오스를 함유하는 배지에서, *ace3* 유전자의 모든 형태의 과발현은 총 단백질의 생산을 어느 정도 향상시켰으나, 향상 수준은 *ace3* 유전자의 L, EL 및 LN

형태를 과발현하는 균주에서 가장 높았다. 높은 수준의 분비 단백질(표 8)은 글루코오스(즉, 비유도 조건)가 탄소원으로서 사용되었을 때 *ace3* 유전자의 L, EL 또는 LN 형태가 과발현된 형질전환체에서 관찰되었음을 분명하다.

표 8

24-웰 플레이트 배양에서 상이한 균주에 의해 생산된 총 단백질

주	5d, mg/ml	7d, mg/ml
시어빠른	M2015	0.77
	M2016	0.54
	M2017	0.47
	M2018	1.23
	M2019	2.50
	M2020	1.97
	M2021	0.67
	M2022	1.19
	M2023	1.84
	M2024	2.05
	M2025	1.76
	M2026	1.93
시어빠른	M1904	0.40
	M2015	0.59
	M2016	0.12
	M2017	0.00
	M2018	0.02
	M2019	1.57
	M2020	1.79
	M2021	0.03
	M2022	0.06
	M2023	1.40
	M2024	1.37
	M2025	1.24
	M2026	1.49
	M1904	0.02
		0.37

[0251]

실시예 7

내인성 *Ace3* 이종성 프로모터 녹인(knock-in)

ace3 유전자좌에서 천연 프로모터의 5' 영역 상류를 포함하는 DNA 단편, *loxP*가 측면에 있는 하이그로마이신 B-저항성 선택 가능한 마커 카세트, 및 *ace3* 오픈 리딩 프레임의 5' 말단에 작동 가능하게 융합된 관심 프로모터를 포함하는 단편을 융합시켜 프로모터 교체 구성체(도 6 참조)를 제조한다(예컨대, 국제 PCT 출원 일련번호 PCT/US2016/017113 참조, 이는 사상 진균에서의 사용을 위한 유전자/프로모터 교체 카세트를 추가로 설명한다).

따라서, 트리코더마 레에세이 세포는 위에 기술된 프로모터 교체 구성체로 형질전환되는데, 이때, 형질전환체는 단리되고, 게놈 DNA는 진단 PCR을 위해 추출되어 천연 *ace3* 유전자좌에서의 프로모터 교체 구성체의 상동 재조합을 확인한다. 이 방법을 이용하여, 천연 *ace3* 프로모터가 하이그로마이신-B 저항성 카세트 및 임의의 관심 프로모터에 의해 교체된 형질전환체가 확인될 수 있다. 이후, 하이그로마이신 B-저항성 카세트는 *cre* 리콤비나이제의 작용에 의해 제거된다(Nagy, 2000).

ace3 유전자좌에서의 상동 통합의 효율은 국제 PCT 공개 제W02016/100272호, 제W02016/100571호 및 제W02016/100568호에 예시된 바와 같은 적절하게 설계된 가이드 RNA에 의해 천연 *ace3* 프로모터 내로 유도된 cas9

의 작용에 의해 증진될 수 있다.

[0257] 조건적 프로모터 교체(conditional promoter replacement, CPR) 전략은 Hu 등 (2007)의 발표에서 아스페질러스 푸미가투스에 대해 설명되었는데, Hu 등(2007)은 일반적으로 선택된 유전자의 내인성 프로모터를 결실시키고 교체하기 위하여 *A. 푸미가투스* *니이아*(*NiiA*) 질소 조절 가능 프로모터(p*NiiA*)를 이용하는 전략을 설명하고 있다. 따라서, 일부 구현예에서, 트리코더마 레에세이에서 *ace3* 유전자의 내인성 프로모터를 대안적인 프로모터로 교체하기 위하여 유사한 방법이 이용될 수 있다.

실시예 8

내인성 비-리그노셀룰로오스 관심 유전자 프로모터의 리그노셀룰로오스 관심 유전자 프로모터로의 교체

[0260] 트리코더마 레에세이 글루코아밀라아제 발현 구성체를 DNA 폴리뉴클레오티드 단편(예컨대, 미국 특허 제 7,413,879호 참조)으로부터 조립하였는데, 이때, *T. 레에세이* 글루코아밀라아제를 암호화하는 ORF 서열이 5' (상류) *T. 레에세이* *cbh1* 프로모터에 작동 가능하게 연결되었고, 3' (하류) *T. 레에세이* *cbh1* 종결자에 작동 가능하게 연결되었다. 이 DNA 구성체는 선택 가능한 마커로서 *T. 레에세이* *pyr2* 유전자를 더 포함하였다.

[0261] 따라서, 본 발명의 변이체 (딸) *T. 레에세이* 세포(즉, *Ace3-L* 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하는)가 글루코아밀라아제 발현 구성체로 형질전환되었다. 배양 중에 *T. 레에세이* 글루코아밀라아제 효소를 분비할 수 있는 형질전환체를 확인하기 위하여 탄소원으로서 글루코오스를 함유하는(즉, 소포로오스 또는 락토오스와 같은 유도 기질이 없는) 액체 배지에서 형질전환체를 선택하고 배양하였다.

[0262] 따라서, *T. 레에세이* 글루코아밀라아제 발현 (모) 균주(대조군) 및 변형된 *T. 레에세이*(딸) 글루코아밀라아제 발현 균주(즉, *Ace3-L* ORF를 포함하고 발현하는)를 진탕 플라스크에서 성장시키고, 모든 세포 배양물로부터의 상청액을 수확하고, 위의 실시예 2와 실시예 3에 일반적으로 기술된 바와 같이, PAGE를 이용하여 분석하였다.

[0263] 더욱 상세하게는, 도 10에 도시된 바와 같이, 모 *T. 레에세이* 세포는 글루코오스/소포로오스를 함유하는 합성 배지(유도 조건)에서 1,029 μ g/mL의 글루코아밀라아제를 생산하였고, 글루코오스를 함유하는 합성 배지(비유도 조건)에서는 오로지 38 μ g/mL의 글루코아밀라아제를 생산하였다. 이와 대조적으로, *dic1* 프로모터로부터 추진된 *ace3-L*을 포함하는 변형된 (딸) 균주 "LT88"은 "유도"("Sop") 조건 하에서 3배 더 많은(즉, 유도 조건 하의 모 (대조군) 균주에 비해) 글루코아밀라아제를 생산하였고, "비유도"("Glu") 조건 하에서 2.5배 더 많은(즉, 유도 조건 하의 모 (대조군) 균주에 비해) 글루코아밀라아제를 생산하거나 비유도 조건 하의 모 (대조군) 균주에 비해 "비유도"("Glu") 조건 하에서 67배 더 많은 글루코아밀라아제를 생산하였다. 따라서, 이러한 결과는 *Ace3-L* ORF를 포함하는 변형된 (딸) 세포는 유도인자의 부존재 하에서 세포 외 단백질을 생산할 뿐만 아니라, 이들 변이체 세포는 이러한 유도 조건 하에서도 모 (대조군) *T. 레에세이* 세포보다 더 많은 총 단백질을 생산함을 증명한다.

실시예 9

천연적으로 관련된 이종성 관심 유전자 프로모터의 리그노셀룰로오스 관심 유전자 프로모터로의 교체

[0266] 피타아제 발현 구성체를 다음과 같이 DNA 폴리뉴클레오티드 단편으로부터 조립하였다(예컨대, 미국 특허 제 8,143,046호 참조). 부티옥셀라(*Buttiauxella*) 종 피타아제를 암호화하는 ORF가 5' 말단에서 *T. 레에세이* *cbh1* 프로모터에, 그리고 3' 말단에서 *T. 레에세이* *cbh1* 종결자에 작동 가능하게 연결된다. 이 DNA 구성체는 선택 가능한 마커인 아스페질러스 *니두란스* *amdS* 유전자를 더 포함한다. 변이체 *T. 레에세이* 세포(즉, *Ace3-L* 단백질을 암호화하는 유전자의 발현을 증가시키는 유전자 변형을 포함하는)가 피타아제 발현 구성체로 형질전환된다. 배양 중에 부티옥셀라 피타아제 효소를 분비할 수 있는 형질전환체를 확인하기 위하여 탄소원으로서 글루코오스를 함유하는(즉, 소포로오스 또는 락토오스와 같은 유도 기질이 없는) 액체 배지에서 형질전환체를 선택하고 배양한다.

실시예 10

천연 *ace3* 프로모터 교체 벡터의 구성

[0269] 본 실시예에서는, 표준적인 문자생물학 절차를 이용하여 두 개의 *ace3* 프로모터 교체 벡터 pCHL760 및 pCHL761을 구성하였다. 벡터 베본 pMCM3282(도 20)는 대장균에서의 복제 및 선택을 위하여 *pMB1 ori* 및 *AmpR* 유전자를 함유하였다. 또한, *N. 크라사*(*crassa*) *cpc1* 프로모터 및 *A. 니두란스*(*nidulans*) *trpC* 종결자 하에서 발현된, 트리코더마 레에세이를 위한 *hph* 하이그로마이신 선택 마커가 포함되었다. 프로모터 교체를 위해, 벡터는 옥수

수를 위해 코돈 최적화된 스트렙토코커스 피오게네스 cas 9를 함유하였고, *T. 레에세이* *pki1* 하에서 발현하였으며, 가이드 RNA는 U6 프로모터 하에서 발현하였다(예컨대, PCT 공개 제WO2016/100568호 및 제WO2016/100272호 참조).

[0270] 따라서, pMCM3282를 EcoRV로 소화시키고, *ace3* 천연 프로모터를 교체하는 *T. 레에세이* *hxk1* 또는 *dic1* 프로모터 영역의 측면에 있는 *T. 레에세이* *ace3* 유전자좌로부터 대략 1kb의 5' 및 3' 상동성 서열을 갖는 3개의 단편을 긁은 어셈블리에 의해 pMCM3282/EcoRV로 클로닝하여 pCHL760(도 21) 및 pCHL761(도 22)가 얻어졌다.

[0271] *hxk1* 또는 *dic1* 프로모터와 함께, 5' 및 3' *ace3* 상동성 서열을 Q5 고성능 DNA 폴리머라아제(뉴잉글랜드 바이오랩스(New England Biolabs)) 및 아래 표 9에 기재된 프라이머를 이용하여 *T. 레에세이* 게놈 DNA로부터 PCR 증폭시켰다.

표 9

PCR 프라이머

CL1791 (SID: 71)	TCTAGTATGTACGAGTACTAGGTGTGAAGATTCCGTCATTCCTCGACAT GCGAATGCG
CL1792 (SID: 72)	TGCCATGCAAACCCCGCATTGCATGTCGAGGAAATGACGGAATCTTC CACCTAGTAC
CL1793 (SID: 73)	TGCAGCTACAGAGCCCTGGGCCGGAGCTGCTGAGCCATAGTTACGACA GCGGAAGCGC
CL1794 (SID: 74)	ATAGCACTTATAAGGCGCGCTTCCGCTGTCGTAACTATGGGCTCAGCA GCTCCGGC
CL1840 (SID: 75)	TAACAAATAGGGTTCCGCGCACATTCCCCGAAAAGTGCCACCTGGA TAGACTAGCA TCTGAGCCATTGCAGC
CL1786 (SID: 76)	AGTGGCACCGAGTCGGTGGTGTCTTTCTATCGAGAGCATTGGTCAG TGGTGGCAAG
CL1800 (SID: 77)	ACCAATATAACAAACATGTCGTCGAGCCAGTGCCTGCCATTCCCTCGAC ATGCGAATGC
CL1801 (SID: 78)	GTTGCCATGCAAACCCCGCATTGCATGTCGAGGAAATGGCAGGCACTG GCTCGGACGAC
CL1802 (SID: 79)	AGCTACAGAGCCCTGGGCCGGAGCTGCTGAGCCATTGTTGAATTCTGG CGGGGTAGCTG
CL1803 (SID: 80)	CTTTTACACTTCAACAGCTACCCGCCAGAATTCAACAATGGGCTCA GCAGCTCCGGC
CL1831 (SID: 81)	TAGTCGGTTATCAACTGAAAAAGTGGCACCGAGTCGGTGGTGTCTTTT TTTCTATCGAGATGTTGGATGGTGGAGAGG

위의 표에서 "SID"는 "서열 식별 번호"의 약어, 예컨대, "서열 번호"이다.

[0272]

[0273] 더욱 상세하게는, 각 벡터를 위한 단편의 PCR 증폭에 사용된 구체적인 프라이머는 다음과 같이 나열된다. pCHL760을 구성하기 위하여, 5' 상류 상동성 영역을 프라이머 쌍 CL1840 및 CL1791을 이용하여 증폭시켰고, 3' 하류 상동성 영역을 프라이머 CL1794 및 CL1831을 이용하여 증폭시켰고, *dic1* 프로모터를 프라이머 쌍 CL1792 및 CL1793을 이용하여 증폭시켰다.

[0274] pCHL761을 구성하기 위하여, 5' 상류 상동성 영역을 프라이머 쌍 CL1840 및 CL1800을 이용하여 증폭시켰고, 3' 하류 상동성 영역을 프라이머 CL1803 및 CL1831을 이용하여 증폭시켰고, *hxk1* 프로모터를 프라이머 쌍 CL1801 및 CL1802를 이용하여 증폭시켰다.

[0275] 벡터 pMCM3282(도 20)는 5'에서 3' 방향으로 *T. 레에세이* U6 프로모터, 대장균 *ccdB* 카세트, 및 U6 유전자로부터의 인트론을 포함하는, Cas9 결합에 관여하는 단일 가이드 RNA(sgRNA)의 구조적 영역을 포함한다. *ccdB* 카세트를 트리코더마 *ace3* 유전자 내의 다섯 개의 상이한 표적 부위에 특이적인 서열들로 교체하였다. 최종 *ace3* 프로모터 교체 벡터를 구성하기 위한 가이드 RNA 서열의 pCHL760(도 21) 및 pCHL761(도 22) 내로의 삽입.

[0276] 따라서, *AarI* 제한 부위가 있는, 표 10에 제시된 다음의 올리고뉴클레오티드를 상이한 sgRNA 서열의 생산을 위하여 설계하였다.

표 10

올리고뉴클레오티드 sgRNA 서열

올리고 ID	올리고뉴클레오티드 설명	올리고뉴클레오티드 서열	서열 번호
CL1821	TS1을 위한 상부 올리고	AGTCTATCGCAGCCTGCCTAGCTAATGTTT	82
CL1822	TS1을 위한 하부 올리고	TCTAAACATTAGCTAAGGCAAGGCTGCGATA	83
CL1823	TS4를 위한 상부 올리고	AGTCTATCGGCAGAGTCGCGTCTCCGGGTTT	84
CL1824	TS4를 위한 하부 올리고	TCTAAAACCCGGAAGACGCGACTCTGCCGATA	85
CL1825	TS5를 위한 상부 올리고	AGTCTATCGAATGAGTGTAGGTACGAGTAGTTT	86
CL1826	TS5를 위한 하부 올리고	TCTAAAACACTCGTACCTACACTCATTGCGATA	87
CL1827	TS8을 위한 상부 올리고	AGTCTATCGGCCGCAATAGCTTCCTAATGTTT	88
CL1828	TS8을 위한 하부 올리고	TCTAAAACATTAGGAAGCTATTGCGGCCGATA	89
CL1829	TS10을 위한 상부 올리고	AGTCTATCGCAGCGCAATCAGTGCAGTGGTTT	90
CL1830	TS10을 위한 하부 올리고	TCTAAAACCACTGCACTGATTGCGCTGCGATA	91

[0277]

더욱 상세하게는, CL1821과 CL1822, CL1823과 CL1824, CL1825와 CL1826, CL1827과 CL1828, CL1829와 CL1830을 어닐링하여 이중 가닥 DNA를 생성하였고, 이를 개별적으로 IIS형 심리스 클로닝 방법을 이용하여 *AarI* 부위에서 pCHL760 및 pCHL761 내로 클로닝하였다. 정확하게 삽입된 가이드 RNA 서열이 있는 최종 플라스미드는 독성이 있는 *ccdB* 유전자를 잃었다.

[0279]

T. 레에세이의 형질전환

[0280]

pCHL760 및 pCHL761의 cas9 매개성 ace3 프로모터 교체 벡터를 폴리에틸렌 글리콜(PEG) 매개성 원형질체 형질전환에 의해 *T. 레에세이* 모 세포 내로 형질전환시켰다. 하이그로마이신 저항성 형질전환체를 선택하기 위하여 하이그로마이신을 함유한 보겔 최소 배지 아가 상에 형질전환체를 성장시켰다. 이를 형질전환체 중 일부는 불안정하여 플라스미드를 흡수하였지만, 게놈 DNA 내로의 안정적인 통합은 없었다. 형질전환체를 보겔 비선택적 아가 배지로 옮겨 플라스미드 및 하이그로마이신-저항성 마커를 상실하게 하였다.

[0281]

dic1 프로모터 교체 형질전환체를 스크리닝하기 위하여, 게놈 DNA를 추출하고, 프라이머 쌍 CL1858 및 CL1848(원하는 통합에 대한 예상 크기의 생성물은 2,412 bp였음), 그리고 CL1853 및 CL1818(원하는 통합에 대한 예상 크기의 생성물은 2,431 bp였음)(예컨대, 표 11 참조)을 이용하여 PCR을 수행하였다. 이후, PCR 생성물을 DNA 시퀀싱으로 분석하여 원하는 프로모터 통합을 확인하였다.

[0282]

hxk1 프로모터 교체 형질전환체를 스크리닝하기 위하여, 프라이머 쌍 CL1858 및 CL1898(원하는 통합에 대한 예상 크기의 생성물은 1,784 bp였음), 그리고 CL1853 및 CL1850(원하는 통합에 대한 예상 크기의 생성물은 2,178 bp였음)을 이용하여 PCR을 수행하였고, 이후 정확한 통합을 PCR 생성물에 대한 DNA 시퀀싱에 의해 확인하였다. 예컨대, 표 11 참조.

표 11

PCR 프라이머

프라이머 ID	프라이머 서열	서열 번호
CL1858	TGGAGAGACTCGGAGAGGATAGG	92
CL1853	AGCGTGGAGGCAGTTGGAGTGG	93
CL1848	TGGACAAAGCCTGGGTCTGCTCC	94
CL1818	ATCCTGACTCGTCCTGTGTCGG	95
CL1898	AGTGCTTCGTTAGTGGACTTG	96
CL1850	CTCGGTAGCTGCTGAATATAG	97

[0283]

진탕 플라스크에서의 단백질 생산

[0285]

ace3 프로모터 교체 균주의 기능성을 시험하기 위하여, 진탕 플라스크 내의 50 mL 액침 배양액에서 유도인자 기질(소포로오스)의 존재 및 부존재 하에서 세포를 성장시켰다. 모 *T. 레에세이* 숙주 세포(ID 번호 1275.8.1)와 변이체 세포 ID 번호 2218, 2219, 2220, 2221, 2222 및 2223을 액체 배양액 중 유도 조건(탄소원으로서 글루코오스/소포로오스) 및 비유도 조건(탄소원으로서 글루코오스) 하에서 성장시키고, 각각의 세포 외 분비 단백질 생산 수준을 비교하였다. 간략하게는, 각각의 숙주 세포(즉, *T. 레에세이* 모 숙주 세포 및 이의 변이체 세포)의 균사를 바탕 칸막이가 있는 250 mL 엘렌마이어 플라스크 내의 50 mL의 YEG 배양액에 따로따로 첨가하였다. YEG 배양액은 5g/L 효모 추출물 및 22 g/L 글루코오스를 함유하였다. 세포 배양물을 48시간 동안 성장시키고, 이어서 또 다른 24시간 동안 새로운 YEG 내로 계대배양하였다. 그런 다음, 이를 종 배양물을 바탕 칸막이가 있는 250 mL 진탕 플라스크 내의 2.5% 글루코오스(비유도 조건)가 첨가된 50 mL의 합성 배지, 또는 2.5% 글루코오스/소포로오스(유도 조건)가 첨가된 50 mL의 합성 배지 내로 접종하였다. 모든 진탕 플라스크를 200 rpm에서 계속하여 진탕하면서 28 °C에서 인큐베이션하였다. 4일의 인큐베이션 후, 모든 세포 배양물로부터의 상청액을 수화하고, **도 23**에 제시된 바와 같이, SDS-PAGE를 이용하여 분석하였다.

[0286]

위의 SDS-PAGE에서 보여지는 바와 같이, 모 세포(**도 23**, ID 1275.8.1)는 글루코오스/소포로오스(유도)와 비교하여 글루코오스를 함유하는 합성 배지(비유도)에서 활чин 적은 분비 단백질을 생산하였다. 이와 대조적으로, 형질전환체 2218, 2219, 2220, 2222 및 2223은 유도 조건 및 비유도 조건 하에서 비슷한 양의 분비 단백질을 생산하였다. 따라서, 이러한 결과는 ace3 유전자좌에서 천연 ace3 프로모터를 교체하는 *hxk1* 또는 *dic1* 프로모터를 보유하는 변이체 세포가 유도인자의 부존재 하에서 세포 외 단백질을 생산하였음을 증명한다.

[0287]

참고문헌

[0288]

유럽 특허 공개 제215,594호

[0289]

유럽 특허 공개 제244,234호

[0290]

유럽 특허 공개 제0215594호

[0291]

PCT 국제 출원 제PCT/EP2013/050126호

[0292]

PCT 국제 출원 제PCT/US2016/017113호

[0293]

PCT 국제 공개 제W003/027306호

[0294]

PCT 국제 공개 제W01992/06183호

[0295]

PCT 국제 공개 제W01992/06209호

[0296]

PCT 국제 공개 제W01992/06221호

[0297]

PCT 국제 공개 제W01992/10581호

[0298]

PCT 국제 공개 제W01998/15619호

[0299]

PCT 국제 공개 제W02002/12465호

- [0300] PCT 국제 공개 제WO2003/52054호
- [0301] PCT 국제 공개 제WO2003/52055호
- [0302] PCT 국제 공개 제WO2003/52056호
- [0303] PCT 국제 공개 제WO2003/52057호
- [0304] PCT 국제 공개 제WO2003/52118호
- [0305] PCT 국제 공개 제WO2004/16760호
- [0306] PCT 국제 공개 제WO2004/43980호
- [0307] PCT 국제 공개 제WO2004/48592호
- [0308] PCT 국제 공개 제WO2005/001036호
- [0309] PCT 국제 공개 제WO2005/01065호
- [0310] PCT 국제 공개 제WO2005/028636호
- [0311] PCT 국제 공개 제WO2005/093050호
- [0312] PCT 국제 공개 제WO2005/28636호
- [0313] PCT 국제 공개 제WO2005/93073호
- [0314] PCT 국제 공개 제WO2006/074005호
- [0315] PCT 국제 공개 제WO2006/74005호
- [0316] PCT 국제 공개 제WO2009/149202호
- [0317] PCT 국제 공개 제WO2010/141779호
- [0318] PCT 국제 공개 제WO2011/038019호
- [0319] PCT 국제 공개 제WO2011/063308호
- [0320] PCT 국제 공개 제WO2011/153276호
- [0321] PCT 국제 공개 제WO2012/125925호
- [0322] PCT 국제 공개 제WO2012/125951호
- [0323] PCT 국제 공개 제WO2012125937호
- [0324] PCT 국제 공개 제WO2013/102674호
- [0325] PCT 국제 공개 제WO2014/047520호
- [0326] PCT 국제 공개 제WO2014/070837호
- [0327] PCT 국제 공개 제WO2014/070841호
- [0328] PCT 국제 공개 제WO2014/070844호
- [0329] PCT 국제 공개 제WO2014/093275호
- [0330] PCT 국제 공개 제WO2014/093281호
- [0331] PCT 국제 공개 제WO2014/093282호
- [0332] PCT 국제 공개 제WO2014/093287호
- [0333] PCT 국제 공개 제WO2014/093294호
- [0334] PCT 국제 공개 제WO2015/084596호
- [0335] PCT 국제 공개 제WO2016/069541호

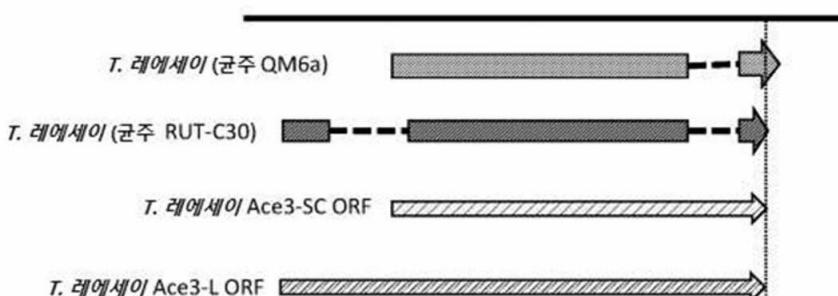
- [0336] PCT 국제 공개 제WO2016/100272호
- [0337] PCT 국제 공개 제WO2016/100568호
- [0338] PCT 국제 공개 제WO2016/100571호
- [0339] 미국 특허 제6,022,725호
- [0340] 미국 특허 제6,268,328호
- [0341] 미국 특허 제7,413,879호
- [0342] 미국 특허 제7,713,725호
- [0343] 미국 특허 제8,143,046호
- [0344] Alexopoulos, C. J., *Introductory Mycology*, New York: Wiley, 1962.
- [0345] Allen and Mortensen, *Biotechnol. Bioeng.*, 2641-45, 1981.
- [0346] Arvas, M., Pakula, T., Smit, B., Rautio, J., Koivistoinen, H., Jouhten, P., Lindfors, E., Wiebe, M., Penttila, M., and Saloheimo, M., "Correlation of gene expression and protein production rate-a system wide study", *BMC Genomics* 12, 616, 2011.
- [0347] Ausbel *et al.*, "Current Protocols in Molecular Biology", Green Publishing Associates/Wiley Interscience, New York, 1987.
- [0348] Ausubel *et al.*, Current Protocols in Molecular Biology, Green Publishing Associates/Wiley Interscience, New York, 1994.
- [0349] Boel *et al.*, *EMBO J.* 3:1581-1585, 1984.
- [0350] Campbell *et al.*, *Curr. Genet.*, 16: 53-56, 1989.
- [0351] Cao *et al.*, *Science*, 9: 991-1001, 2000.
- [0352] Colot *et al.*, *PNAS* 103(27):10352-10357, 2006.
- [0353] Devereux *et al.*, *Nucleic Acids Res.* 12:387-395, 1984.
- [0354] el-Gogary *et al.*, "Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*", *PNAS*, 86(16)-6138-6141, 1989.
- [0355] Hakkinnen, M., Valkonen, M.J., Westerholm-Parvinen, A., Aro, N., Arvas, M., Vitikainen, M., Penttila, M., Saloheimo, M., and Pakula, T.M., "Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production", *Biotechnol Biofuels* 7, 14, 2014.
- [0356] Harkki *et al.*, *BioTechnol.*, 7: 596-603, 1989.
- [0357] Harkki *et al.*, *Enzyme Microb. Technol.*, 13: 227-233, 1991.
- [0358] Hu *et al.*, "Essential gene identification and drug target prioritization in *Aspergillus fumigatus*" *PLoS Pathog.*, 3(3), 2007.
- [0359] Ilmen *et al.*, "Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*", *Applied and Environmental Microbiology*, 63(4)-1298-1306, 1997.
- [0360] Ju and Afolabi, *Biotechnol. Prog.*, 91-97, 1999.
- [0361] Kriegler, Gene Transfer and Expression: A Laboratory Manual, 1990.
- [0362] Martinez *et al.*, "Genome sequencing and analysis of the bio-mass-degrading fungi *Trichoderma reesei* (syn. *Hypocrea jecorina*)", *Nature Biotechnology*, 26:533-560, 2008.
- [0363] Mullaney *et al.*, *MGG* 199:37-45, 1985.

- [0364] Nagy, "Cre recombinase: the universal reagent for genome tailoring", *Genesis* 26(2), 99-109, 2000.
- [0365] Needleman and Wunsch, *J. Mol. Biol.*, 48:443, 1970.
- [0366] Nunberg *et al.*, *Mol. Cell Biol.* 4:2306, 1984.
- [0367] Ouedraogo, J.P., Arentshorst, M., Nikolaev, I., Barends, S., and Ram, A.F., "I-SceI-mediated double-strand DNA breaks stimulate efficient gene targeting in the industrial fungus *Trichoderma reesei"* *Applied microbiology and biotechnology* 99, 10083-10095, 2015.
- [0368] Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988.
- [0369] Penttila, M., Nevalainen, H., Ratto, M., Salminen, E., and Knowles, J., "A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei"*, *Gene* 61, 155-164, 1987.
- [0370] Poggi-Parodi, D., Bidard, F., Pirayre, A., Portnoy, T., Blugeon, C., Seiboth, B., Kubicek, C.P., Le Crom, S., and Margeot, A., "Kinetic transcriptome analysis reveals an essentially intact induction system in a cellulase hyper-producer *Trichoderma reesei* strain", *Biotechnol Biofuels* 7, 173, 2014.
- [0371] Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2nd Edition, *Cold Spring Harbor Laboratory Press*, Cold Spring, New York, 1989.
- [0372] Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 4th Edition, *Cold Spring Harbor Laboratory Press*, Cold Spring, New York, 2012.
- [0373] Seiboth, *et. al.*, *Mol. Genet. Genomics*, 124-32, 2002.
- [0374] Sheir-Neiss and Montenecourt, "Characterization of the secreted cellulases of *Trichoderma reesei* wild type and mutants during controlled fermentations", *Applied Microbiology and Biotechnology*, 20(1):46-53, 1984.
- [0375] Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981.
- [0376] Vaheri *et al.*, "Transglycosylation products of the cellulase system of *Trichoderma reesei"*, *Biotechnol. Lett.*, 1:41-46, 1979.
- [0377] Yelton *et al.*, *PNAS USA* 81:1470-1474, 1984.

도면

도면 1a

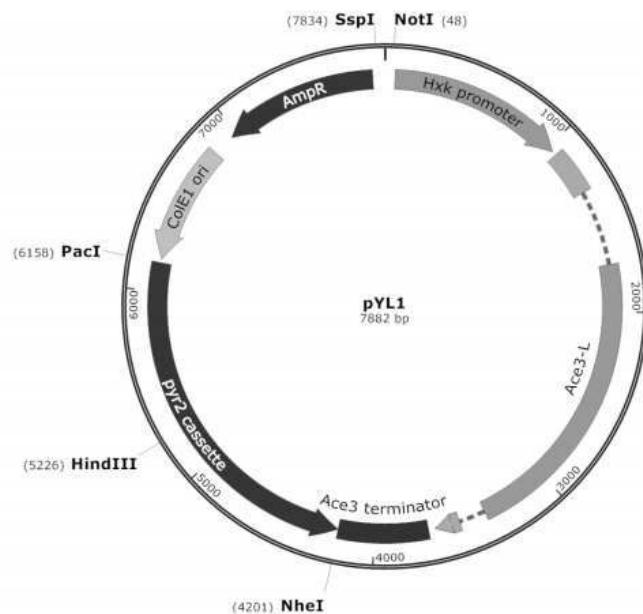
T. reesei 균주 QM6a 및 RUT-C30으로부터의 Ace3 단백질 암호화 영역의 개략도



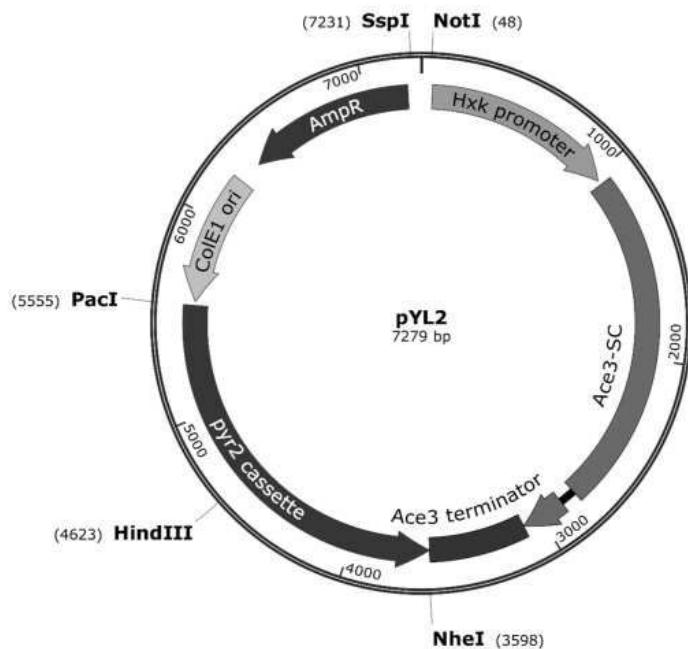
도면1b

	1	10	20	30	40	50	60
Ace3-S	-	-	-	-	-	-	-
Ace3-SC	-	-	-	-	-	MGSAAPAQGSVAAA	
Ace3-L	-	-	-	-	-	MGSAAPAQGSVAAA	
Ace3-LN	-	-	-	-	-	MGSAAPAQGSVAAA	
Ace3-LC	-	-	-	-	-	MGSAAPAQGSVAAA	
Ace3-EL	-	MATAAAAAAGGAAVAAGADTGAAGSSSTGPPGELGLFGLTGTGRTGTSVAMGSAAPAQGSVAAA					
Ace3-S	-	-	-	-	-	-	-
Ace3-SC	-	-	-	-	-	-	-
Ace3-L	-	-	-	-	-	-	-
Ace3-LN	-	-	-	-	-	-	-
Ace3-LC	-	-	-	-	-	-	-
Ace3-EL	-	-	-	-	-	-	-
Ace3-S	-	-	-	-	-	-	-
Ace3-SC	-	-	-	-	-	-	-
Ace3-L	-	-	-	-	-	-	-
Ace3-LN	-	-	-	-	-	-	-
Ace3-LC	-	-	-	-	-	-	-
Ace3-EL	-	-	-	-	-	-	-
Ace3-S	-	-	-	-	-	-	-
Ace3-SC	-	-	-	-	-	-	-
Ace3-L	-	-	-	-	-	-	-
Ace3-LN	-	-	-	-	-	-	-
Ace3-LC	-	-	-	-	-	-	-
Ace3-EL	-	-	-	-	-	-	-
Ace3-S	-	-	-	-	-	-	-
Ace3-SC	-	-	-	-	-	-	-
Ace3-L	-	-	-	-	-	-	-
Ace3-LN	-	-	-	-	-	-	-
Ace3-LC	-	-	-	-	-	-	-
Ace3-EL	-	-	-	-	-	-	-
Ace3-S	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-SC	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-L	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-LN	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-LC	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-EL	LSTAARPGQMPPPLTFSGPAAVAALQPFASSSSLPDAAWEPVEPLSDIDNGLPQRQLGDLPG						
Ace3-S	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-SC	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-L	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-LN	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-LC	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-EL	LSTIQNISTRQRWHLANAMTLLRERVKRCIDLFDYDYLPLTPLYVEPALRDVLAY						
Ace3-S	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-SC	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-L	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-LN	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-LC	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-EL	IFSQQLPVGVNQSPSLQSLTPDFTGTFLNAEWSAGFCQGPFSRTGVSRLAWFADSTTF						
Ace3-S	LVTAVCAEEAFLMPLKDF1PPEGESVS EIILEASRDC1LHQHLEADLENPTANS1I1RYFHSN						
Ace3-SC	LVTAVCAEEAFLMPLKDF1PPEGESVS EIILEASRDC1LHQHLEADLENPTANS1I1RYFHSN						
Ace3-L	LVTAVCAEEAFLMPLKDF1PPEGESVS EIILEASRDC1LHQHLEADLENPTANS1I1RYFHSN						
Ace3-LN	LVTAVCAEEAFLMPLKDF1PPEGESVS EIILEASRDC1LHQHLEADLENPTANS1I1RYFHSN						
Ace3-EL	LVTAVCAEEAFLMPLKDF1PPEGESVS EIILEASRDC1LHQHLEADLENPTANS1I1RYFHSN						
Ace3-S	CLHAAGKPKYSWHFGEAIRLAQVMQHLHEEEALEGLVPIEAEFRRCFWRILYLGDKSAAI						
Ace3-SC	CLHAAGKPKYSWHFGEAIRLAQVMQHLHEEEALEGLVPIEAEFRRCFWRILYLGDKSAAI						
Ace3-L	CLHAAGKPKYSWHFGEAIRLAQVMQHLHEEEALEGLVPIEAEFRRCFWRILYLGDKSAAI						
Ace3-LN	CLHAAGKPKYSWHFGEAIRLAQVMQHLHEEEALEGLVPIEAEFRRCFWRILYLGDKSAAI						
Ace3-EL	CLHAAGKPKYSWHFGEAIRLAQVMQHLHEEEALEGLVPIEAEFRRCFWRILYLGDKSAAI						
Ace3-S	LNRRPIT1IHYKCFDAGIT1LPLYSFGSIEDEFLSTASEPPRKSFSIGFNANWRV1WQSAAIDL						
Ace3-SC	LNRRPIT1IHYKCFDAGIT1LPLYSFGSIEDEFLSTASEPPRKSFSIGFNANWRV1WQSAAIDL						
Ace3-L	LNRRPIT1IHYKCFDAGIT1LPLYSFGSIEDEFLSTASEPPRKSFSIGFNANWRV1WQSAAIDL						
Ace3-LN	LNRRPIT1IHYKCFDAGIT1LPLYSFGSIEDEFLSTASEPPRKSFSIGFNANWRV1WQSAAIDL						
Ace3-EL	LNRRPIT1IHYKCFDAGIT1LPLYSFGSIEDEFLSTASEPPRKSFSIGFNANWRV1WQSAAIDL						
Ace3-S	EIRVLQDQMQQHFRGTMPPNHVLP SADQRQLHDLSLYVRFITCLDDLPPFLYQLQSCTLAMAAA						
Ace3-SC	EIRVLQDQMQQHFRGTMPPNHVLP SADQRQLHDLSLYVRFITCLDDLPPFLYQLQSCTLAMAAA						
Ace3-L	EIRVLQDQMQQHFRGTMPPNHVLP SADQRQLHDLSLYVRFITCLDDLPPFLYQLQSCTLAMAAA						
Ace3-LN	EIRVLQDQMQQHFRGTMPPNHVLP SADQRQLHDLSLYVRFITCLDDLPPFLYQLQSCTLAMAAA						
Ace3-EL	EIRVLQDQMQQHFRGTMPPNHVLP SADQRQLHDLSLYVRFITCLDDLPPFLYQLQSCTLAMAAA						
Ace3-S	EGNGSAESKQYV1CQ1NLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKS1EIVRDMRLV						
Ace3-SC	EGNGSAESKQYV1CQ1NLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKS1EIVRDMRLV						
Ace3-L	EGNGSAESKQYV1CQ1NLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKS1EIVRDMRLV						
Ace3-LN	EGNGSAESKQYV1CQ1NLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKS1EIVRDMRLV						
Ace3-EL	EGNGSAESKQYV1CQ1NLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKS1EIVRDMRLV						
Ace3-S	MNEAFTPWLQANGEPEVK1R1LIGASLL1IHRNQDSPLATRARSDFSVL1DILTRLDSK						
Ace3-SC	MNEAFTPWLQANGEPEVK1R1LIGASLL1IHRNQDSPLATRARSDFSVL1DILTRLDSK						
Ace3-L	MNEAFTPWLQANGEPEVK1R1LIGASLL1IHRNQDSPLATRARSDFSVL1DILTRLDSK						
Ace3-LN	MNEAFTPWLQANGEPEVK1R1LIGASLL1IHRNQDSPLATRARSDFSVL1DILTRLDSK						
Ace3-EL	MNEAFTPWLQANGEPEVK1R1LIGASLL1IHRNQDSPLATRARSDFSVL1DILTRLDSK						
Ace3-S	ASDQLRNRNTTTVVG (SID NO:3)						
Ace3-SC	ASD----- (SID NO:8)						
Ace3-L	ASD----- (SID NO:6)						
Ace3-EL	ASD----- (SID NO:14)						
Ace3-S	ASDQLRNRNTTTVVG (SID NO:10)						
Ace3-SC	ASD----- (SID NO:12)						

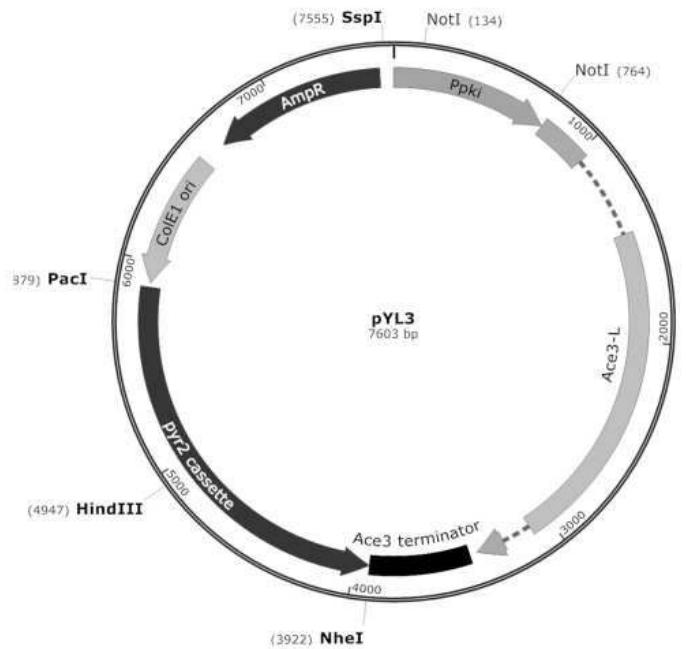
도면2a



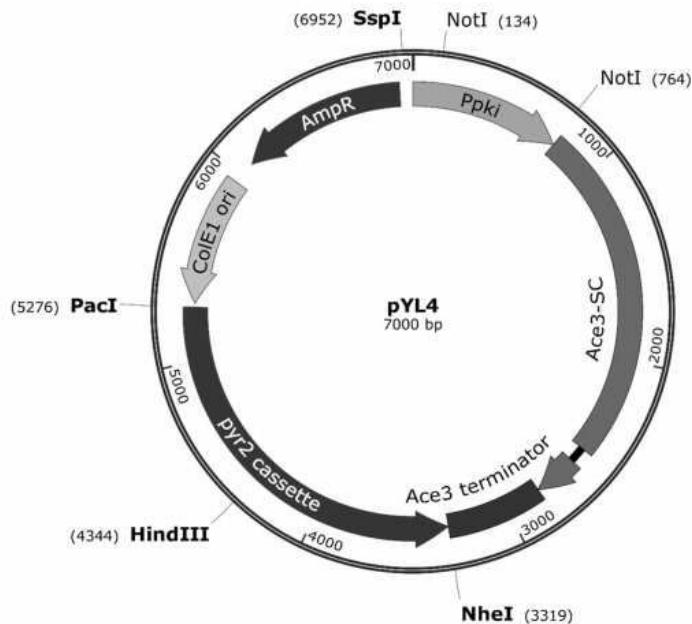
도면2b



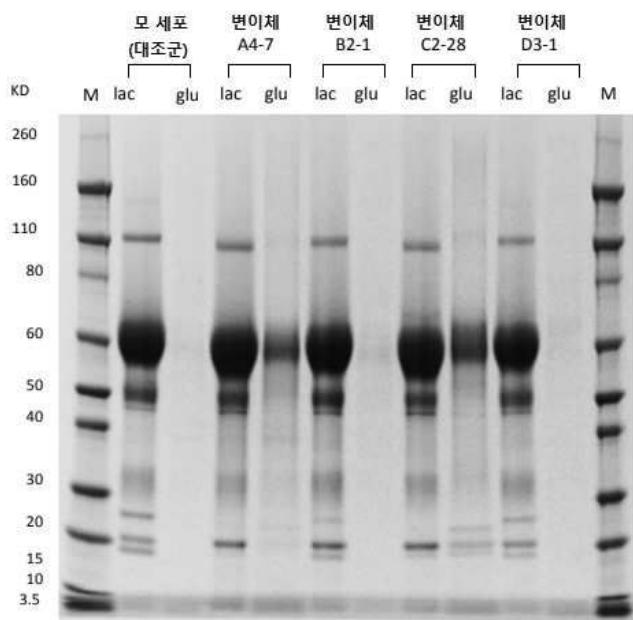
도면2c



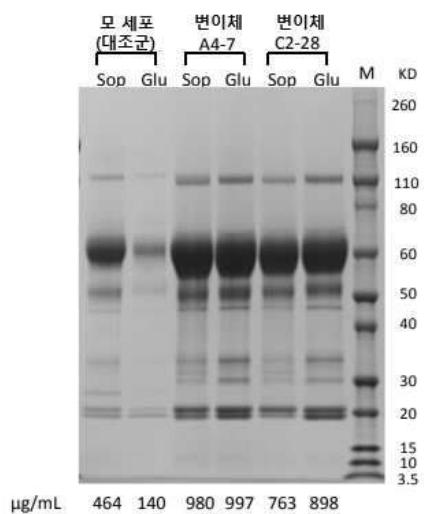
도면2d



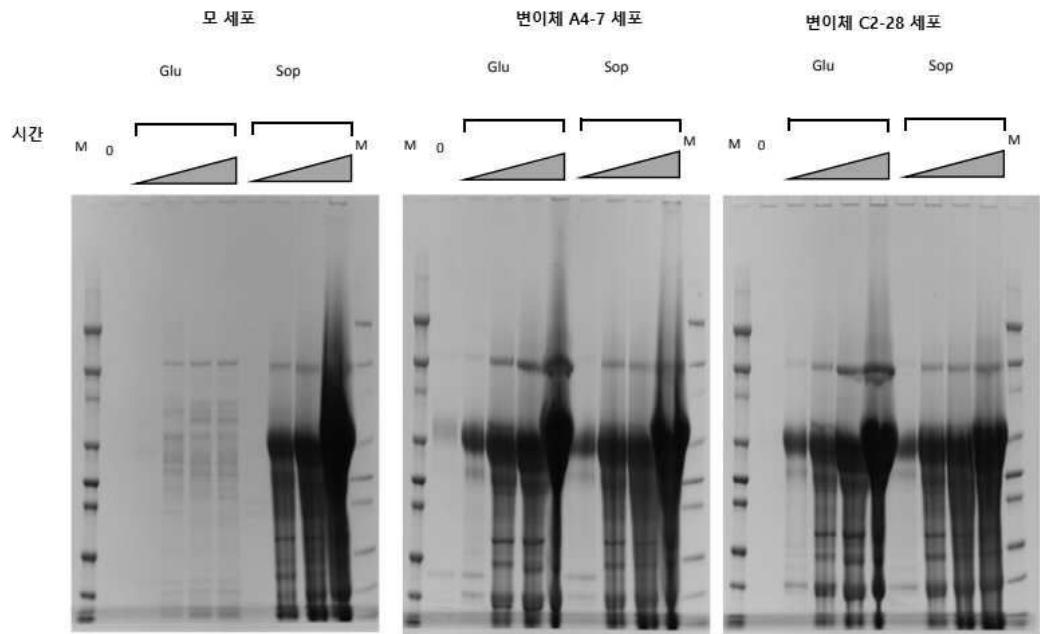
도면3



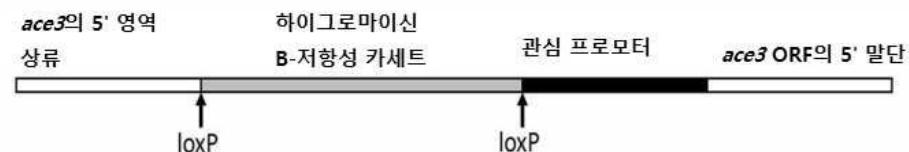
도면4



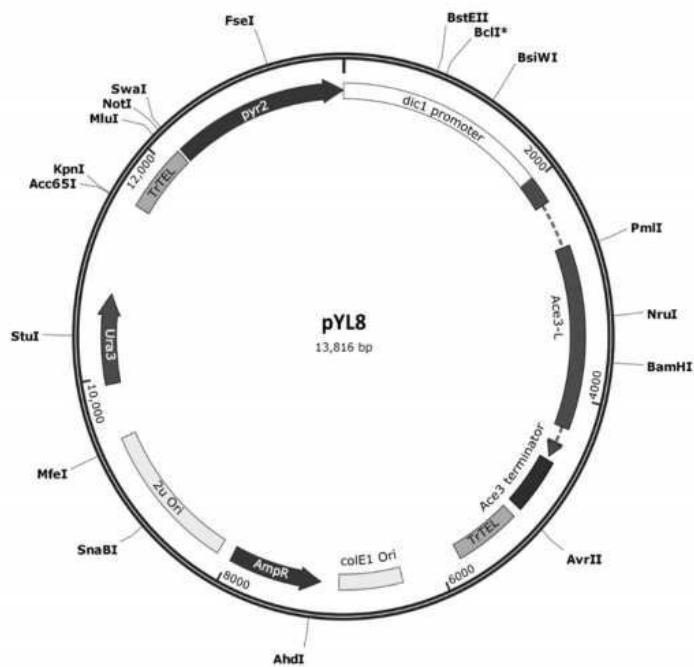
도면5



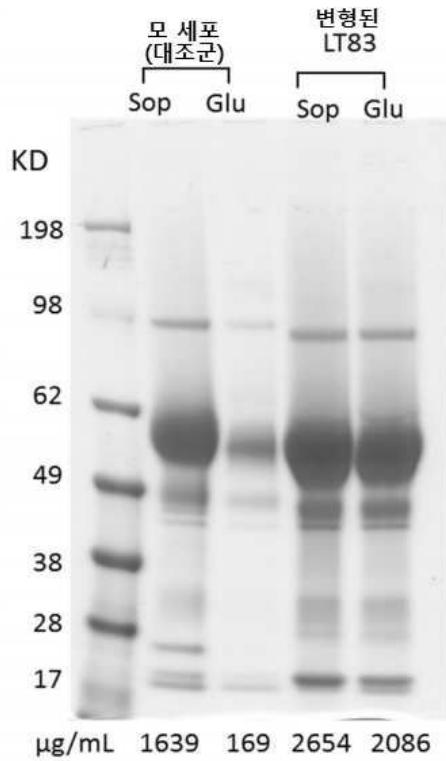
도면6



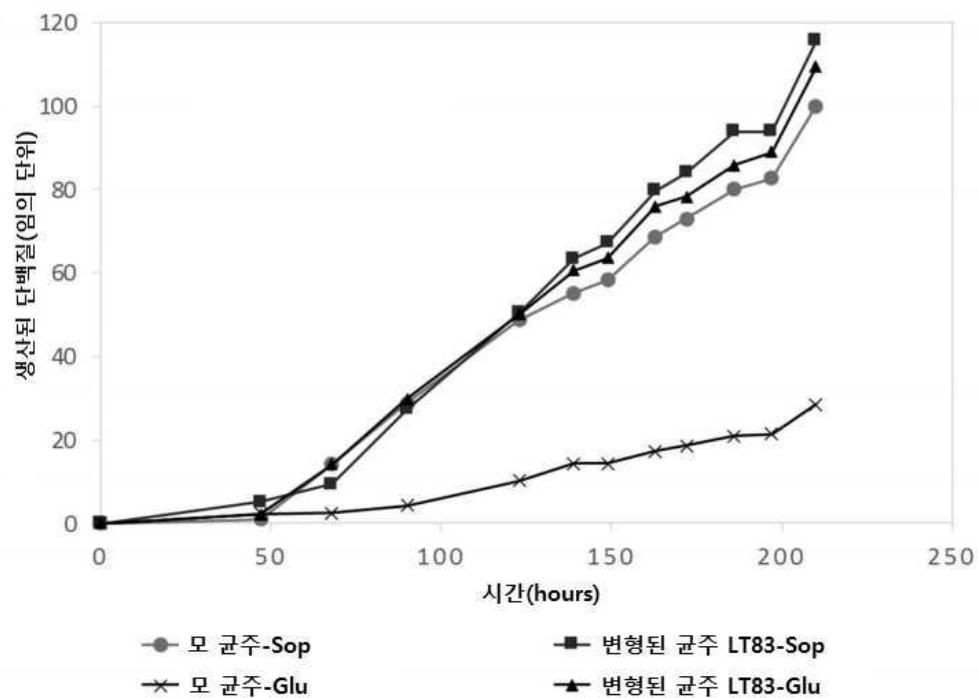
도면7



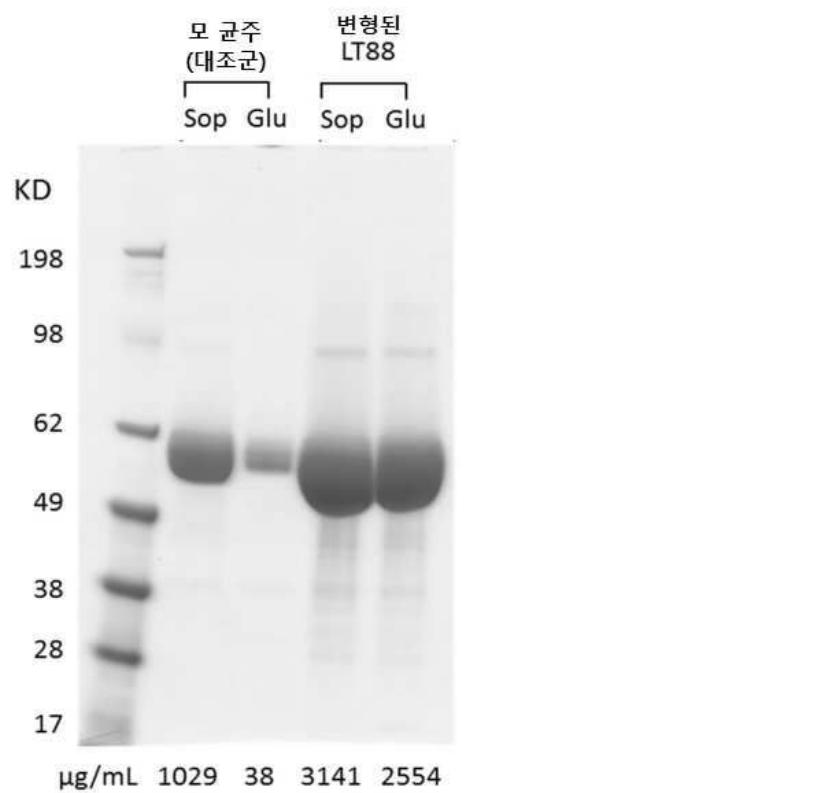
도면8



도면9



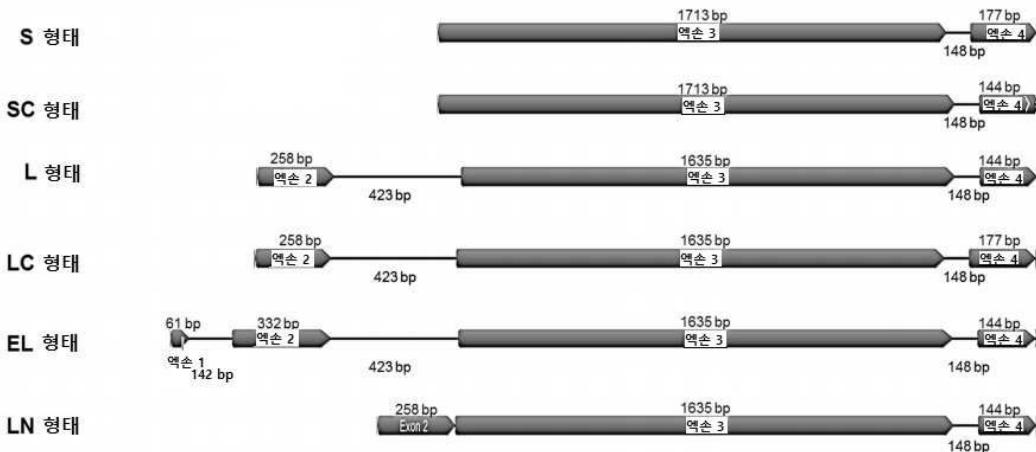
도면10



도면11



도면12



도면13

Ace3 SC 형태(서열 번호 7)

Ace3 단백질 SC 형태(서열 번호 8)

MLRYSPPVHLDTLSSLPLPNTNALPRPKCEYLSAVDSCTHRDAHVQCTFDLPLARRGPKARKKSQDPQQPPDPSLSTAARPQGMPPPLTFSGPAVAALQPFASSSSLSPDAWEPEV
EPLSISPLRPLQPLGDLGLSTQNISTRQRWHLAMTLRNLTRVLSRKCIDLFFDYLPTLPLVPEYALPDRDVLAYISQFLSPQLQVNPQSLPQLTDPDTGTTPLNAAESWAG
FQPGSQGSRTVGSRLPLAWDADSTFTLVTAVCAEEALMPKLDLIFPEGEVSSEIILLEASRDLQHLEADLENPLTANSIARAYFHNSLCHAAGPKYSHWFGEATLRAQLWMLHEEAAL
EGLVPIEIAEFRRRCFVWLIGDKSAAILNNRNPITIHKYCFDAGITLTLYPSGIEDEFLSTASEPPRKSFSIGFVNANRVLQSAADLLEIRVLQDQMMQHFRGTMVPLHVLPSADRQ
HDLSDYRFTICPLTADPLYPLQSCSTLAMMAEENGNSGCSQVYQIQCINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRLKESEIVRDLRMLRVMEAPFWGLQANGEPNEVKEKIRLGA
SLLAIHNRNDSPCITLPLARRSDFSVLILTLRDLKSAD

도면14

Ace3 S 형태(서열 번호 1)

Ace3 단백질 S 형태(서열 번호 3)

MLRYSFVLHLDLTLSPPLTNAFLPRCEYLSAVDSCTHCRDAHQVCTFDLPLARRGPKARKKSQDQPGQPFPDPSSLTAARPGQMPPPLTFSGPAVAALQPFASSSLSPDAAEWPEVPLSIDNGLPQRQDGLPGLSTIQNISTRQRWHLNAMANLTNRNLTTRVSKRC1DFDYYLPLTPLYVEPLRDLVYLAJTSQPLGPVNQSPSPLQLTPDTPTT1PLNAEWSWAGFGQPGSRTSGRSLPAPWDSTFTLVATCACEAALMPKDFPEGEVSSEILLEASRDLCHQHLEADLENPAINTSIAIRYHFSNCLHAAGKPGKWHIFGEAIRLAVQMLHEEAALEGVLPFIEAFRRRCFWILYLGDKSAIAILNNRNPITIHYKFCDAIGITLWYPSGIEDEFSLTASEPPRKSFSIGFNGANVRLWQSADLLEIRVLQDQMMQHFRGTMPPNNHVLPSADRHQLDLSLVFRITCDDLDRPLYQSLCSTLAMAEGNSAESKQVYQICNLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMLRVMNEAPFWGLQANGEPNVEKIRLIGASLLIAIRHNQDSPATRANSFSVLLIDLTLDSDAQSLRNLNTSTVVG

도면15

Ace3 L 형태(서열 번호 4)

Ace3 단백질 L 형태(서열 번호 6)

도면16

Ace3 LC 형태(서열 번호 9)

Ace3 단백질 LC 형태(서열 번호 10)

MGCSAAPAQGSVAAAAGCQGPPAAGAGGAGVHALTHSTPSESASASQPGSPTASTTPQNSLVSAAATSFFHHIPGRRLVSRACDRRRKAKCEYLSAVDSCTHCRDAHVQCTFDLPLARRGPKKRNKSD
QPGQPPQDPDFSSSLSTAARPGQMPPFLTQGPQFVAWAALQPFQFSSSLDPAWEPPFPLS1DQDGLPGLPGLTQIN1NISTRHLWHLANAM1LNTLRLTVEFSRKC1D1FFDYLQPLVYEP
DLRVLAYI1FQSPQLPQVGNQPSLQLTDPPTTGTFLNAESAGWAGFGPSGSRTVGRSFLAPWADSTFTVTAACAEAMFLPKD1FEGEVSSE1LLEASRDLQHCLAEADLNTPANTSIA1RYF
HSNCLHAAAGKPKYWSH1FGEAIRLAQVQLHEEAAALEG1VPI1EAAFRRCF2WLYLGCDKSA11LNRPNT1IHYKCYFDAGIT1LYPG1EDEFLSTASEFPRKSF1FSGFNNANVRWLWQSADLLE
1IRVLQDZQMMQHFRGTMPPHNLVSDAQRHQLDSD1LVRFTC1LDDLQFLYQSLCT1LAMAAGNGSAK1EASQK1VQ1N1LQTVFHCLRMVITQKFEDLSYFAPGVQADLRKSE1VROMLRVMNEA
FWFLQGANQENVKE1R1L1GSLA11HRRNQDLSPLARSD5FVLLD11LTDLSRASDQLRNLSTTVVGG

도면17

Ace3 EL 형태(서열 번호 11)

Ace3 단백질 EL 형태(서열 번호 12)

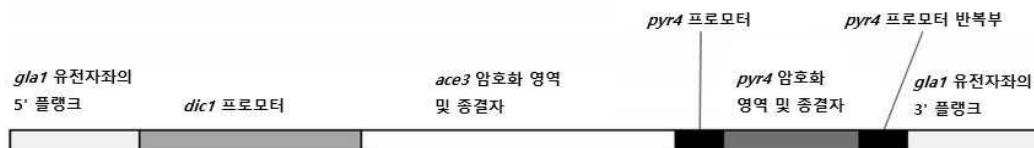
MATAAAAAAAGGAJAVAAGADTGAGSSSTGPGPLGLPGCITRGSVAMGSAAPQGSVAAAAGGPAAAGAGAGAVHALTTESESASQPGSPTASTTPPNQSLVSAATSFHHHPRGLVSRACD
MATAKRRKACELYSAVSDCSTCRRDAHQQVCFDTDLARLGPKARRKSDDQFGPPFDSLSTAARPGMPPLFTFSGPVAALQFQFNSLSDPAWEFVFTSDFLRLPQLGDLPLGTLSTQI
ISTRQRWHLANLMTLRLNTLERSVRCK1DFLDDYLYLPTLVEPALARLIVAYLQVNGPQNSPQLDTPTDGTGTLNAESWAQFGPQSRTVGSRFLAWDASTFTVTAACEA
AFMLPKDIFPEGEVSIELEASRDCLQRHLEADLENPTANSIAIRYFHNSCLHAAKGPKYWSH1FEA1R1LAQVMQLHEEAALEGV1PIEAEFRRCRFLWILYLGDKSAALNRRPIT1HYC
FDAGITTYLPSGSEDEFLTASEPFRKFISFVGPNANVRQWSAADDLLEIRV1LQDQMMQHFRGFTMPVNPVLWSADQRHLDL1YVRFTC1LDDLPYLQSCST1LAMAEMGNGSAESKQYV1QC
LN1QVFTCHRLMVT1QKFDLSSYFAGVEQDLRSE1VRMLVMNEAFTWGLQANGEPEV1R1L1GASL11HNRQDFPLATRARSDFSVL1L1LRTDLSKAD

도면18

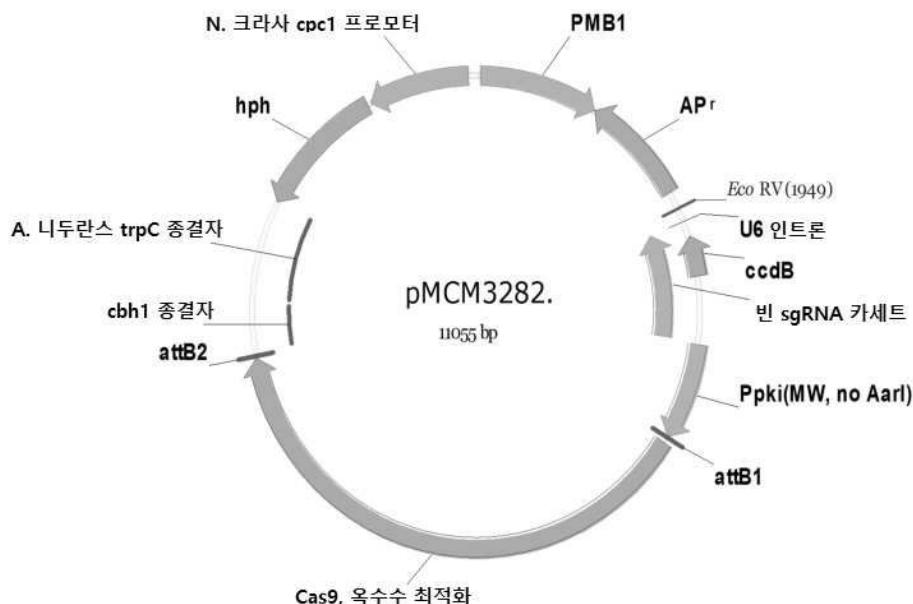
Ace3 LN 형태(서열 번호 13)

Ace3 단백질 LN 형태(서열 번호 14)

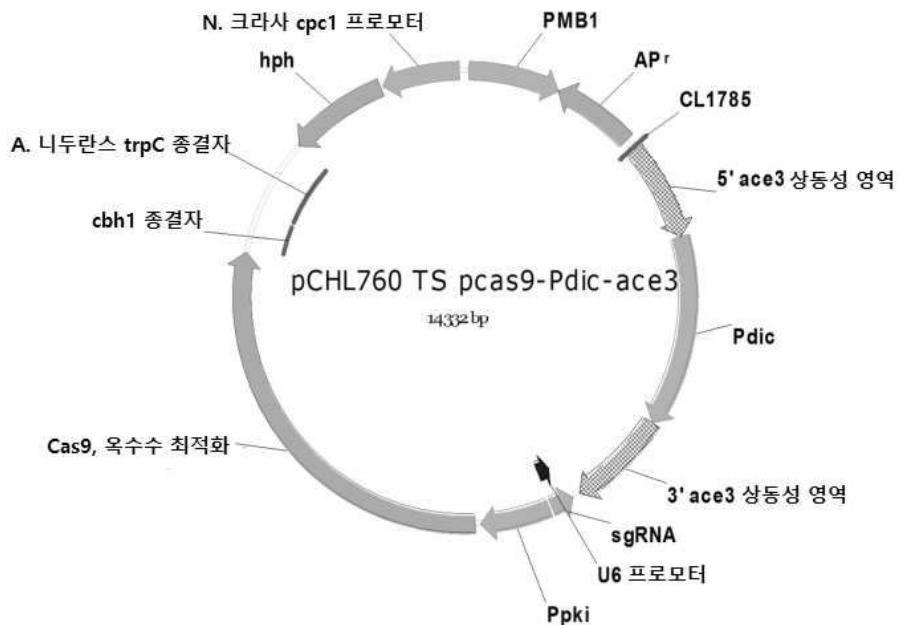
도면19



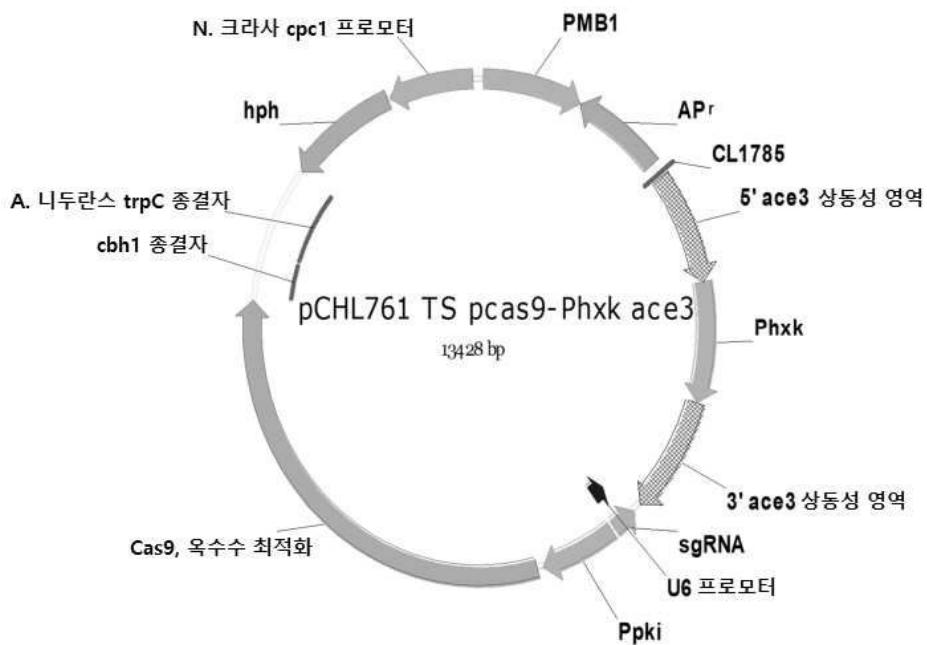
도면20



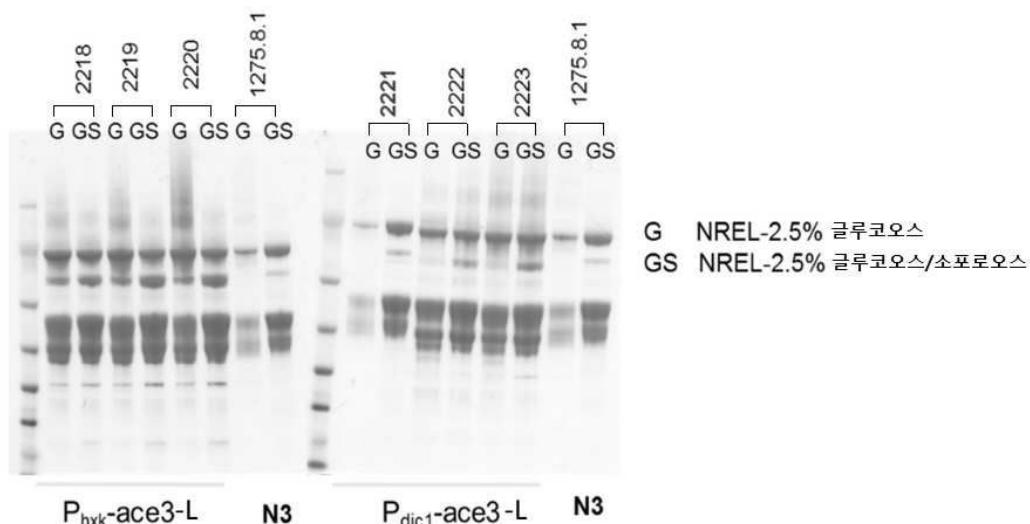
도면21



도면22



도면23



서 열 목 록

<110> Danisco
 <120> PROTEIN PRODUCTION IN FILAMENTOUS FUNGAL CELLS IN THE ABSENCE OF INDUCING SUBSTRATES

<130> NB41159-W0-PCT

<150> US 62/403,787

<151> 2016-10-04

<160> 102

<170> PatentIn version 3.5

<210> 1

<211> 2038

<212> DNA

<213> Trichoderma reesei

<400> 1

atgctgcgt actccccgt cttacacctg gatactctc cttgccacc actgaccaat 60

gctttcccc gcccaaagtg cgagtacctc agcgctgtcg atagctgcac gcactgccgc 120

gatgccacg tgcagtgcac tttcgacctg cccctggcgc gacgcggccc caaagcgagg 180

aagaagagcg accagcccg ccagccgcct cctgatccga gctcgcttc caccgcccgt 240

cgaccggcc agatgccgcc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag 300

cccttcgcct cgtcgctcgct gtcgcccgc acggcctggg agcccgctga gccgctcagc 360

attgacaacg gcctgccccg gcagccgctg ggcgacactc cccgccttc caccatccag 420

aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacacg 480

acgctagagc gcgctcgaa gcgtatgtac gacctttct tcgactacct ctacccctc	540
accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	600
ttgcctggcg tcaaccaacc atcgccgtcg tcacagctca cgccagaccc gaccacggc	660
accacccccc tcaacgctgc cgagtcgtgg gccggcttg gccagccag cggctcgca	720
accgtcgcca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	780
tgcgcagagg cagcattcat gctacccaag gacatttcc cgaaggaga atccgtctct	840
gagatcttc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctgca	960
gggaagccca agtactcgta gcacatattt ggcgaggcca tccgcctggc gcaggatcatg	1020
cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccgcgt	1080
cgctgcttt ggatcctgta ctgggcgac aagttagccg ctataactcaa caatcgccc	1140
atcaccatcc acaagttactg ctgcgtccgc ggcacatcacca cgctatacccg tcgggtatc	1200
gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagcttcat atccggcttc	1260
aacgcaaatg tgccgtctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgtc	1320
caagatcaga tgcgtcgac caattcgagg accatgcggc cgaaccatgt gtcgcctcc	1380
ccgcacagggc agcatctcgat ttctcttat gtccgtttca tcacctgttt ggacgatctc	1440
ccgcccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcactgtctg	1560
cgcatggtaa ttacgcgaa attcgaagac ctcttttatt ttgtcctgg cggtgagcag	1620
gctgatctca gaaagtccgaa gattgtgcga gacatgctga gggtgatgaa cgaggcgccc	1680
ttttggggcc tgcaggccaa tggcgagcca aacgtgagtc gttccctgt ctcttcttt	1740
ttctgcacac cttttcttc gacgacccccc cctctcttt tatatccctg cgatgtatgta	1800
tatcatcaag cctcggact tggctaat ctgtcctgtat tatgttgtct ggatgctgca	1860
ggttgaaaag attcgcctta tggagactag tttgtggcc atcatccatc gcaaccagga	1920
ttcaccccttg gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg	1980
gctggactcg aaggcgtcg accaactgag gaatacgtcc actaccgttg ttggctaa	2038
<210> 2	
<211> 1890	
<212> DNA	
<213> Trichoderma reesei	
<400> 2	

atgctgcgt actccccgt cttacacctg gatactctc cttgccacc actgaccaat	60
gctttcccc gcccaaagtg cgagtaccc agcgctgtcg atagctgcac gcactgccgc	120
gatgccacg tgcatgtcac tttcgacctg cccctggcgc gacgcggccc caaagcgagg	180
aagaagagcg accagcccg ccagccgcct cctgatccga gtcgctctc caccgcccgt	240
cgaccggcc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgccgtcgac	300
cccttcgcct cgtcgctcgct gtcgcccgc acggccctgg agcccgctga gccgctcagc	360
attgacaacg gcctgccccg gcagccgcgt ggcgacactc cccgcctctc caccatccag	420
aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacaacacg	480
acgctagagc gcgtctcgaa gcgatgtatc gaccctttct tcgactacct ctacccctc	540
accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	600
ttgcctggcg tcaaccaacc atgcgcgtc tcacagctca cggccagaccc gaccaccggc	660
accacccccc tcaacgcgtc cgagtcgtgg gccggcttg gccagcccag cggctcgca	720
accgtggca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	780
tgccgcagagg cagcattcat gtcacccaag gacatccc ccgaaggaga atccgtctct	840
gagatcttgc tcgaaggctc tcgggactgc ctgcaccagg acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctgcg	960
gggaagccca agtactcggt gcacatattt ggccgaggcca tccgcctggc gcaggtcatg	1020
cagctgcacg aggaggctgc ctcgagggg ctgtccccca tcgaggcaga gttccggcgt	1080
cgctgcgttt ggatcctgta cttggcgcac aagttagccg ctatactcaa caatcgcccc	1140
atcaccatcc acaagtactg ctgcacgcgc ggcataccca cgctatacccg tgcgggtatc	1200
gaggacgagt tcctgagcac ggcgtccgag ccgcggccga agagctcat atccggcttc	1260
aacgcaaatg tgcggctctg gcagtcgcgc gctgatttc tgctggaaat ccgcgtgctg	1320
caagatcaga tgaatcgacca cttcgaggg accatgcggc cgaaccatgt gtcgcctcc	1380
gccgacaggc agcatctcga ttctcttat gtccgcttca tcacctgctt ggacgatctc	1440
ccgcgttacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcactgtctg	1560
cgcattgtaa ttacgcgaa attcgaagac ctctttatt ttgtccctgg cgttggcag	1620
gctgatctca gaaagtccga gattgtgcga gacatgcgtga gggatgtgaa cgaggccccc	1680
ttttggggcc tgcaggccaa tggcgagccaa aacgttggaa agattgcct tatcgagct	1740
agtttgcgtgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1800

gactttccg tgctttgga tattctcagc cggctggact cgaaggcgtc ggaccaactg 1860
 aggaatacgt ccactaccgt tggtggctaa 1890
 <210> 3
 <211> 629
 <212> PRT
 <213> Trichoderma reesei
 <400> 3
 Met Leu Arg Tyr Ser Pro Val Leu His Leu Asp Thr Leu Ser Leu Pro
 1 5 10 15
 Pro Leu Thr Asn Ala Leu Pro Arg Pro Lys Cys Glu Tyr Leu Ser Ala
 20 25 30
 Val Asp Ser Cys Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe
 35 40 45
 Asp Leu Pro Leu Ala Arg Arg Gly Pro Lys Ala Arg Lys Lys Ser Asp
 50 55 60
 Gln Pro Gly Gln Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala
 65 70 75 80
 Arg Pro Gly Gln Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val
 85 90 95
 Ala Ala Leu Gln Pro Phe Ala Ser Ser Ser Leu Ser Pro Asp Ala Ala
 100 105 110
 Trp Glu Pro Val Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln
 115 120 125
 Pro Leu Gly Asp Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr
 130 135 140
 Arg Gln Arg Trp Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr
 145 150 155 160
 Thr Leu Glu Arg Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr
 165 170 175
 Leu Tyr Pro Leu Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val
 180 185 190

Leu Ala Tyr Ile Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser
 195 200 205
 Pro Leu Ser Gln Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu
 210 215 220
 Asn Ala Ala Glu Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg
 225 230 235 240
 Thr Val Gly Ser Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu
 245 250 255
 Val Thr Ala Val Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile
 260 265 270
 Phe Pro Glu Gly Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg
 275 280 285
 Asp Cys Leu His Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala
 290 295 300
 Asn Ser Ile Ala Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala
 305 310 315 320
 Gly Lys Pro Lys Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu
 325 330 335
 Ala Gln Val Met Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val
 340 345 350
 Pro Ile Glu Ala Glu Phe Arg Arg Cys Phe Trp Ile Leu Tyr Leu
 355 360 365
 Gly Asp Lys Ser Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His
 370 375 380
 Lys Tyr Cys Phe Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile
 385 390 395 400
 Glu Asp Glu Phe Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe
 405 410 415
 Ile Ser Gly Phe Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp
 420 425 430
 Leu Leu Leu Glu Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe
 435 440 445

Arg Gly Thr Met Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln

450 455 460

His Leu Asp Ser Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu

465 470 475 480

Pro Pro Tyr Leu Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu

485 490 495

Gly Asn Gly Ser Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn

500 505 510

Leu Gln Val Thr Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe

515 520 525

Glu Asp Leu Ser Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg

530 535 540

Lys Ser Glu Ile Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro

545 550 555 560

Phe Trp Gly Leu Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg

565 570 575

Leu Ile Gly Ala Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser

580 585 590

Pro Leu Ala Thr Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile

595 600 605

Leu Thr Arg Leu Asp Ser Lys Ala Ser Asp Gln Leu Arg Asn Thr Ser

610 615 620

Thr Thr Val Val Gly

625

<210> 4

<211> 2608

<212> DNA

<213> Trichoderma reesei

<400> 4

atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggcgg ccctccagct 60

gctggcgctg gcgctggcgc tgtccacgccc ctcaccacct cgcccgagtc tgcctcgccc 120

tgcagcccg gctcgccaaac cgccctcaacc acgcccggc agaactcact cgtgtcggt 180

gcaacctcg tccaccacca tcccagaggc cgtctggta gcagagcctg cgaccgctgc	240
cgccggcgca aggccaaggt cagtctagcc ccttgctgt tgcttgcatc tctgttgca	300
ttgcctcc tccgtctgct gctgatgctg ctgcctcc tccctctt cctccccgtc	360
tcctggtccc tggccctgc tcttcatatg tccttactgc cctgtctcc tctccccgtt	420
cccggtcccc ctcctccgt cctttctcc tgcgtgtc tcatgcgtac aaagcataca	480
tacaatacat cagcatacat ggcaagcggt gtgttgtt gagagtgtg tgtattgtat	540
tgcactgcct tcacaactcg ttcatactgc tgcagcctca cccaaacacc gacctcgct	600
tccatgtgc gctactcccc cgtcttacac ctggatactc ttccttgcc accactgacc	660
aatgccttc cccgccccaa gtgcgagttc ctcagcgctg tcgatagctg cacgcactgc	720
cgcgatgccc acgtcgatgt cacttcgac ctgcccctgg cgacgcgg cccaaagcg	780
aggaagaaga ggcgaccagcc cggccagccg ctcctgatc cgagctcgct ctccaccgcg	840
gctcgaccgc gccagatgcc gcccggctg accttctcg gccccgcagt agccgcgtg	900
cagcccttcg ctcgtcgctc gctgtcgccc gacgcccgtt gggagccgt cgagccgctc	960
agcattgaca acggcctgcc cggcagccg ctggcgacc tgccggct ctccaccatc	1020
cagaacatct cgacgcgcca gcatggata cacctggcca acgcccattgac gctgcgcaac	1080
acgacgctag agcgcgtctc gaagcgatgt atgcacccct tttcgacta ccttacccc	1140
ctcacccccc tgggtacga gcccggccctc cgggacgtgc tcgcatacat cttctccag	1200
cccttgctg ggtcaacca accatcgccg ctgtcacagc tcacgcaga cccgaccacc	1260
ggcaccaccc ccctcaacgc tgccgagtgc tggccggct ttggccagcc cagccgtcg	1320
cgaaccgtcg gcagcaggct ggctccctgg gccgactcga cttcacccct ggtcacggcc	1380
gtctgcccag aggccgtt catgtaccc aaggacattt tccccgaagg agaatccgtc	1440
tctgagatct tgctcgaagc ctctcggtac tgcctgcacc agcacctcga ggccgacctg	1500
gagaatccga cggccaaactc gattggcatt cgctacttcc actccaaactg cttccacgct	1560
gccccggaaagc ccaagtactc gtggcacata tttggcgagg ccatccgcct ggccgaggc	1620
atgcagctgc acgaggaggc tgcctcgag gggctcgatcc ccatcgaggc agagttccgc	1680
cgtcgctgtt ttggatct gtacttgggc gacaagttag cgcgtataact caacaatcg	1740
cccatcacca tccacaagta ctgcttcgac gcccgcata ccacgtata cccgtcggt	1800
atcgaggacg agttcctgag cacggcgatcc gagccgcctt ggaagagctt cttatccggc	1860
ttcaacgcaa atgtcggtt ctggcagtcg gcccgttatt tgctgctgaa aatccgcgtg	1920

ctgcaagatc agatgatgca gcacttcga gggaccatgc ccccgaacca tgtgctgcc	1980
tccggcaca ggcagcatct cgattctctc tatgtccgct tcatcacctg cttggacgat	2040
ctccggcgt acctccagtc gtgcactctg gcgatggcag ccatggcaga aggcaacggg	2100
tctggcgagt ccaaggcgtt cgtgatacag tgcataacc tgcaggtgac gtttactgt	2160
ctgcgcattgg taattacgca gaaattcgaa gacccctctt atttgtcc tggcggttag	2220
caggctgatc tcagaaagtc ggagattgtc cgagacatgc tgagggtgat gaacgaggcg	2280
ccctttggg gcctgcaggc caatggcag ccaaacgtga gtcgttccct tgtcttct	2340
ctttctgca caccctttc ttgcacgacc cccctctt ctttatatcc ctgcggatat	2400
gtatatcatc aagectcgcc acttgggtgt aatctgtcct gattatgttgc tctggatgt	2460
gcagggtgaa aagattcgcc ttatcgagc tagttgtcg gccatcatcc atcgcaacca	2520
ggattcaccc ttggctacgc gagccaggag cgactttcc gtgtttgg atattctcac	2580
gcggctggac tcgaaggcgt cgactaa	2608
<210> 5	
<211> 2037	
<212> DNA	
<213> Trichoderma reesei	
<400> 5	
atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggcgg ccctccagct	60
gctggcgctg gcgctggcgc tgtccacgccc ctcaccaccc cgcccgagtc tgcctcgcc	120
tcgcagcccg gctcgccaaac cgcctcaacc acgcccggc agaactcact cgtgtcggt	180
gcaacctcggt tccaccacca tcccagaggc cgtctggta gcagagcctg cgaccgctgc	240
cgcggcgca aggccaaatg cgagtaccc agcgctgtcg atagctgcac gcactggcgc	300
gatggccacg tgcagtgcac ttgcacccgt cccctggcgc gacgcccccaaaaggcgagg	360
aagaagagcg accagccccgg ccagccgcct cctgatccga gtcgttccct caccggcgt	420
cgaccggccc agatggccgc gccgctgacc ttctccggcc ccgcaggtagc cgccgtcgac	480
cccttcgcct cgtcgctcgat gtcgccccgac gcccgttgg agcccgatc gcccgtcagc	540
attgacaacg gcctgccccgg gcagccgctg ggcgaccctgc ccggcccttc caccatccag	600
aacatctcgaa cgcggccagcg atggatacacc ctggccaaacg ccatgacgct ggcgaacacg	660
acgcttagagc gcgctcgaa gcgatgtatc gaccccttct tcgactaccc ctacccctc	720
accccccctgg tgcacggacc ggcctccgg gacgtgctcg catacatctt ctcccgcccc	780
ttgcctggcg tcaaccaacc atgcggctg tcacagctca cgccagaccc gaccaccggc	840

accacccccc tcaacgctgc cgagtcgtgg gccggcttg gccagcccag cggctcgca	900
accgtcgcca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	960

tgccgagagg cagcattcat gctacccaag gacatttcc cccaaggaga atccgtctct	1020
gagatcttgc tcgaagcctc tcgggactgc ctgcaccagg acctcgaggc cgacctggag	1080
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctgcg	1140
gggaagccca agtactcgat gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1200
cagctgcacg aggaggctgc ctcgagggg ctcgtccca tcgaggcaga gttccgcgt	1260
cgctgcttt ggatcctgta cttggcgcac aagttagccg ctataactcaa caatcgccc	1320
atcaccatcc acaagtactg cttcgacgcc ggcacatcca cgtataaccc gtcgggtatc	1380

gaggacgagt tcctgagcac ggcgtccgag ccgcggccga agagcttcat atccggcttc	1440
aacgcaaatg tgccgctctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgctg	1500
caagatcaga ttagtcgcac ctttcgaggg accatgcccc cgaaccatgt gctccctcc	1560
ccgcacaggc agcatctcga ttctcttat gtccgttca tcacctgctt ggacgatctc	1620
ccgcgttacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680
cccgagtcca agcagtagctg gatacagtgc atcaacctgc aggtgacggt tcactgtctg	1740
cgcatggtaa ttacgcgaa attcgaagac ctcttttatt ttgctcctgg cggtgagcag	1800

gctgatctca gaaagtccgaa gattgtcgaa gacatgctga gggtgatgaa cgaggcgtcc	1860
ttttggggcc tgcaggccaa tggcgagcca aacgttggaa agattcgct tatcgagct	1920
agtttgcgttccatcatcca tcgcaaccag gattcacctt tggctacgcg agccaggagc	1980
gactttccg tgctttggaa tatttcacg cggctggact cgaaggcgtc ggactaa	2037

<210> 6

<211> 678

<212> PRT

<213> Trichoderma reesei

<400> 6

Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly

1 5 10 15

Gly Pro Pro Ala Ala Gly Ala Gly Ala Val His Ala Leu Thr

20 25 30

Thr Ser Pro Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala

35 40 45

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

290	295	300
Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val		
305	310	315
Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly		
325	330	335
Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His		
340	345	350
Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala		
355	360	365
Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys		
370	375	380
Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met		
385	390	395
Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala		
405	410	415
Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser		
420	425	430
Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Tyr Cys Phe		
435	440	445
Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe		
450	455	460
Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe		
465	470	475
Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Glu		
485	490	495
Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met		
500	505	510
Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser		
515	520	525
Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu		
530	535	540

Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser
 545 550 555 560
 Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr
 565 570 575
 Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser
 580 585 590

 Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile
 595 600 605
 Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu
 610 615 620
 Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala
 625 630 635 640
 Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr
 645 650 655
 Arg Ala Arg Ser Asp Phe Ser Val Leu Asp Ile Leu Thr Arg Leu

 660 665 670
 Asp Ser Lys Ala Ser Asp

675

<210> 7

<211> 2005

<212> DNA

<213> Trichoderma reesei

<400> 7

atgctgcgt actccccgt cttacacctg gatactctc cttgccacc actgaccaat 60
 gctttcccc gccaaagtg cgagtacctc agcgctgtcg atagctgcac gcactgccgc 120
 gatgccacg tgcaagtgcac tttcgacctg cccctggcgc gacgcggccc caaagcgagg 180
 aagaagagcg accagcccg ccagccgct cctgatccga gctcgctctc caccgcggct 240

cgaccggcc agatgccgcc gcccgtgacc ttctccggcc ccgcagtagc cgcgctgcagc 300
 cccttcgcct cgtcgctcgct gtcggccgac gcccgtggg agcccgtcga gccgctcagc 360
 attgacaacg gcctgccccg gcagccgctg ggcgacactc cccgcctctc caccatccag 420
 aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacacg 480
 acgctagagc gcgctcgaa gcgatgtatc gacctttct tcgactacact ctacccctc 540

acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	600
ttgcctggcg tcaaccaacc atgcccgtg tcacagctca cgccagaccc gaccaccggc	660
accacccccc tcaacgctgc cgagtcgtgg gccggcttg gccagccag cggctcgca	720
accgtcgca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	780
tgccagagg cagcattcat gctacccaag gacatttcc cccaaggaga atccgtctct	840
gagatcttgc tgaaggcctc tcggactgc ctgcaccagc acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctg	960
gggaagccca agtactcggt gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1020
cagctgcacg aggaggctgc ctcgagggg ctgcgtccca tcgaggcaga gttccgcgt	1080
cgctgcttt ggatcctgta ctgggcgac aagttagccg ctatactcaa caatcgccc	1140
atcaccatcc acaagtactg ctgcgacgcc ggcataccca cgctataccg gtcgggtatc	1200
gaggacgagt tcctgagcac ggctccgag ccgcggccga agagcttcat atccggcttc	1260
aacgcaaattg tgccgtctg gcagtcgctg gctgatttc tgctggaaat ccgcgtctg	1320
caagatcaga tgcgtcgac cttcgaggg accatgcccc cgaaccatgt gctgccctcc	1380
gccgacaggc agcatctcgat ttctcttat gtccgcttca tcacctgctt ggacgatctc	1440
ccgcgttacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
gccgagtcca agcagtacgt gatacagtc atcaacctgc aggtgacgtt tcactgtctg	1560
cgcattgttta ttacgcagaa attcgaagac ctctttatt ttgctcctgg cggtgagcag	1620
gctgatctca gaaagtccga gattgtccga gacatgtga gggtgatgaa cgaggccc	1680
ttttggggcc tgcaggccaa tggcgagccaa aacgtgagtc gttccttgc ctcttcctt	1740
ttctgcacac cttttcttc gacgaccccc cctctcttt tatatccctg cggatatgt	1800
tatcatcaag cctcggcact tggctaat ctgtcctgtat tatgttgtct ggatgctgca	1860
ggttgaaaag attcgcccta tcggagctag ttgcgtggcc atcatccatc gcaaccagga	1920
ttcaccccttg gctaegcgag ccaggagcga ctttccgtg ctggata ttctcacg	1980
gctggactcg aaggcgtcg actaa	2005
<210> 8	
<211> 618	
<212> PRT	
<213> Trichoderma reesei	
<400> 8	

Met Leu Arg Tyr Ser Pro Val Leu His Leu Asp Thr Leu Ser Leu Pro

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

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

 325 330 335
 Ala Gln Val Met Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val
 340 345 350
 Pro Ile Glu Ala Glu Phe Arg Arg Cys Phe Trp Ile Leu Tyr Leu
 355 360 365
 Gly Asp Lys Ser Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His
 370 375 380
 Lys Tyr Cys Phe Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile
 385 390 395 400

 Glu Asp Glu Phe Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe
 405 410 415
 Ile Ser Gly Phe Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp
 420 425 430
 Leu Leu Leu Glu Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe
 435 440 445
 Arg Gly Thr Met Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln
 450 455 460
 His Leu Asp Ser Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu

 465 470 475 480
 Pro Pro Tyr Leu Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu
 485 490 495
 Gly Asn Gly Ser Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn
 500 505 510

Leu Gln Val Thr Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe

515 520 525

Glu Asp Leu Ser Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg

530 535 540

Lys Ser Glu Ile Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro

545 550 555 560

Phe Trp Gly Leu Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg

565 570 575

Leu Ile Gly Ala Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser

580 585 590

Pro Leu Ala Thr Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile

595 600 605

Leu Thr Arg Leu Asp Ser Lys Ala Ser Asp

610 615

<210> 9

<211> 2641

<212> DNA

<213> Trichoderma reesei

<400> 9

atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggcgg ccctccagct	60
gctggcgctg gcgcgtggcgc tggccacgccc ctcaccacct cgcggagtc tgcctcgccc	120
tgcagcccg gctcgccaaac cgcctcaacc acgcccggcgc agaactcact cgtgtcggt	180
gcaacctcggt tccaccacca tcccagagggc cgtctggtga gcagagcctg cgaccgctgc	240
cgcggcgca aggccaaggt cagtctagcc ccttgctgt tgcttgatc tctgttgca	300

ttgtcttcc tcctgctgct gctgatgctg ctgtcttcc tcctcttcc cctccccgtc	360
tcctggtccc tggccctgc tttcatatg tccttactgc cctgtctcc tctccccgtt	420
cccggtcccc ctcccccgt cctttctcc tgcgtgtctg tcatgcgtac aaagcataca	480
tacaatacat cagcatacat ggcaagcggt gtgttgttt gagagttgtg tgtattgtat	540
tgcactgcttc tcacaactcg ttcatactgc tgcagctca ccccaacacc gacctcgct	600
tccatgctgc gctactcccc cgtcttacac ctggatactc ttccttgcc accactgacc	660
aatgctttc cccgccccaa gtgcgagttac ctcagcgctg tcgatagctg cacgcactgc	720

cgcgatgcc	acgtcagtg	cacttcgac	ctgcccctgg	cgcgacgcgg	ccccaaagcg	780
aggaagaaga	gcgaccagcc	cggccagccg	cctccgtatc	cgagctcgct	ctccaccgcg	840
gctcgacccg	gccagatgcc	gccgcccgtg	accttctccg	gccccgcagt	agccgcgtg	900
cagcccttcg	cctcgctc	ctgtcgccc	gacgcccgt	gggagcccgt	cgagccgtc	960
agcattgaca	acggcctgcc	ccggcagccg	ctggcgcacc	tgcccgccct	ctccaccatc	1020
cagaacatct	cgacgcgcca	gcgatggata	cacctggcca	acgccatgac	gctgcgcaac	1080
acgacgctag	agcgcgtctc	gaagcgatgt	atcgacctct	tcttcgacta	cctctacccc	1140
ctcacccccc	tggtgtacga	gccggccctc	cgggacgtgc	tcgcatacat	cttctccag	1200
cccttgctg	ggtcaacca	accatcgccg	ctgtcacagc	tcacgcaga	cccgaccacc	1260
ggcaccaccc	ccctcaacgc	tgccgagtgc	tggccggct	ttggccagcc	cagcggctcg	1320
cgaaccgtcg	gcagcaggct	ggctccctgg	gccgactcga	ctttacccct	ggtcaaggcc	1380
gtctgcgcag	aggcagcatt	catgttaccc	aaggacattt	tccccgaagg	agaatccgtc	1440
tctgagatct	tgtctgaagc	ctctcgggac	tgcctgcacc	agcacctcga	ggccgacctg	1500
gagaatccga	cggccaactc	gattgccatt	cgctacttcc	actccaactg	cctccacgct	1560
gcggggaagc	ccaagtactc	gtggcacata	tttggcgagg	ccatccgcct	ggcgcaggtc	1620
atgcagctgc	acgaggaggc	tgcctcgag	gggctcggtcc	ccatcgaggc	agagttccgc	1680
cgtcgctgct	tttggatct	gtacttgggc	gacaagtcag	ccgtataact	caacaatcg	1740
cccatcacca	tccacaagta	ctgcttcgac	gccggcatca	ccacgtata	cccgctgggt	1800
atcgaggacg	atgtcctgag	cacggcgtcc	gagccccc	ggaagagctt	catatccggc	1860
ttcaacgcaa	atgtcggtct	ctggcagtcc	gcccgttatt	tgcgtctgga	aatccgcgtg	1920
ctgcaagatc	agatgtatgc	gactttcga	gggaccatgc	ccccgaacca	tgtgtgcccc	1980
tccggccaca	ggcagcatct	cgattctctc	tatgtccgt	tcatcacctg	tttggacat	2040
ctccggccgt	acctccagtc	gtgcactctg	gcgtatggcag	cgatggcaga	aggcaacggg	2100
tctggccagt	ccaaggcgt	cgtgtatcag	tgcataacc	tgcaggtgac	gtttactgt	2160
ctgcccgtgg	taattacgca	gaaattcgaa	gacccctt	atttgtcc	tggcgtttag	2220
caggctgtac	tcaaaaatgc	ggagattgtg	cgagacatgc	tgagggtgt	gaacgaggcg	2280
ccctttggg	gcctgcaggc	caatggcag	ccaaacgtga	gtcggttct	tgtctttct	2340
ctttctgca	caccctttc	ttcgacgacc	ccccctctct	ctttatatcc	ctgcccgtat	2400
gtatatcatc	aaggcctcgcc	acttgggtct	aatctgtcct	gattatgttgc	tctggatgt	2460
gcaggttcaa	aagattcgcc	ttatcggtcc	tagttgtct	gccccatcc	atcgcaacca	2520
ggattcaccc	ttggctacgc	gagccaggag	cgactttcc	gtgttttgg	atattctcac	2580

gcggctggac tcgaaggcgt cgaccaact gaggaatacg tccactaccg ttgttggcta 2640
 a 2641
 <210> 10
 <211> 689
 <212> PRT
 <213> Trichoderma reesei
 <400> 10
 Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly

1 5 10 15
 Gly Pro Pro Ala Ala Gly Ala Gly Ala Val His Ala Leu Thr
 20 25 30
 Thr Ser Pro Glu Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala
 35 40 45
 Ser Thr Thr Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe
 50 55 60
 His His His Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys
 65 70 75 80
 Arg Arg Arg Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys
 85 90 95
 Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu
 100 105 110
 Ala Arg Arg Gly Pro Lys Ala Arg Lys Ser Asp Gln Pro Gly Gln
 115 120 125
 Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln
 130 135 140
 Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Leu Gln

145 150 155 160
 Pro Phe Ala Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val
 165 170 175
 Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp
 180 185 190
 Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp

195	200	205
Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg		
210	215	220
Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu		
225	230	235
Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile		
245	250	255
Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln		
260	265	270
Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu		
275	280	285
Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser		
290	295	300
Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val		
305	310	315
Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly		
325	330	335
Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His		
340	345	350
Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala		
355	360	365
Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys		
370	375	380
Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met		
385	390	395
Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala		
405	410	415
Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser		
420	425	430
Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Tyr Cys Phe		
435	440	445

Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe
 450 455 460
 Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe
 465 470 475 480
 Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Leu Glu
 485 490 495
 Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met
 500 505 510

Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser
 515 520 525
 Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu
 530 535 540
 Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser
 545 550 555 560
 Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr
 565 570 575
 Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser

580 585 590
 Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile
 595 600 605
 Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu
 610 615 620
 Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala
 625 630 635 640
 Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr
 645 650 655

Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu
 660 665 670
 Asp Ser Lys Ala Ser Asp Gln Leu Arg Asn Thr Ser Thr Thr Val Val
 675 680 685
 Gly

<210>	11					
<211>	2885					
<212>	DNA					
<213>	Trichoderma reesei					
<400>	11					
atggccacag	cgccgcggc	agcagctggc	ggcgccggg	ttgctgcggg	tgcagacaca	60
ggtgcgttga	gtcccgccc	gtccgctcg	gttccctccc	agctgccagc	ccgcgtgggt	120
ggcactggaa	cgcagtgcag	cgcaatcagt	gcagtgcggc	ccccccaact	aacgctgccc	180
cccggtggc	ctcgccaca	caggcgctgc	aggctccagc	tctacaggcc	ctccaggcct	240
tccagggttt	ccaggcaccc	ggacaggctc	cgtggcgatg	ggctcagcag	ctccggccca	300
gggcctgtta	gctgcagctg	caggcggccc	tccagctgt	ggcgctggcg	ctggcgctgt	360
ccacgcctc	accacctcg	ccgagtctgc	ctcgccctcg	cagccggct	cgccaaccgc	420
ctcaaccacg	ccgcccaga	actcactctgt	gtcggtcgca	acctcggtcc	accaccatcc	480
cagaggccgt	ctggtgagca	gaggctgcga	ccgctgccc	cggcgcaagg	ccaaggtcag	540
tctagccctt	ttgctgttgc	ttgcatctt	gttgcatatt	ctctcttcc	tgctgctgt	600
gatgctgctg	ctctcctcc	tcctcctct	ccccgtctcc	tggccctgg	tccctgtct	660
tcatatgtcc	ttactgccc	tgtctctct	ccccgttccc	gttccccc	ctcccgctt	720
cttcctctgc	gtgtctgtca	tgcgtacaaa	gcatacatac	aatacatcag	catacatggc	780
aagcgttgt	ttgtgtttag	agtttgtgt	attgtattgc	actgccttca	caactcgttc	840
atactgtgc	agcctcaccc	caacacccgac	ctcgcttcc	atgctgcgt	actccccgt	900
cttacacctg	gatactctt	cttgccacc	actgaccaat	gctttcccc	gccccaaagt	960
cgagtagctc	agcgctgtcg	atagctgcac	gcactgcgc	gatgcccaacg	tgcagtgac	1020
tttcgacctg	ccccctggcgc	gacgcggccc	caaagcgagg	aagaagagcg	accagccgg	1080
ccagccgcct	cctgatccga	gctcgcttc	caccgcggct	cgacccggcc	agatgccgc	1140
gccgctgacc	tttccggcc	cccgagtagc	cgcgctgcag	cccttcgcct	cgtcgctcg	1200
gtcgcccgac	gccccctggg	agcccgctcg	gccgctcagc	attgacaacg	gcctgcccc	1260
gcagccgctg	ggcgacctgc	ccggccttc	caccatccag	aacatctcg	cgcgccagcg	1320
atggatacac	ctggccaacg	ccatgacgt	gcgcaacacg	acgctagagc	gcgtctcgaa	1380
gcgatgtatc	gacctttct	tcgactacct	ctaccccttc	acccccc	tgtacgagcc	1440
ggccctccgg	gacgtgctcg	catacattt	ctcccgcccc	ttgctggcg	tcaaccaacc	1500
atcgccgcgt	tcacagctca	cgccagaccc	gaccacccgc	accacccccc	tcaacgctgc	1560

cgagtcgtgg	gccggcttg	gccagcccag	cggtcgca	accgtcgca	gcaggctggc	1620
tccctggcc	gactcgacct	tcaccctggt	cacggccgtc	tgcgcagagg	cagcattcat	1680
gctacccaag	gacatttcc	ccgaaggaga	atccgtctct	gagatcttgc	tcgaagcctc	1740
tcgggactgc	ctgaccaggc	acctcgaggc	cgaccctggag	aatccgacgg	ccaactcgat	1800
tgccattcgc	tacttccact	ccaactgcct	ccacgctgct	gggaagccca	agtactcgta	1860
gcacatattt	ggcgaggcca	tccgcctggc	gcaggtcatg	cagctgcacg	aggaggctgc	1920
cctcgagggg	ctcgccccca	tcgaggcaga	gttccgcgt	cgctgccttt	ggatccctgt	1980
cttggcgac	aagttagccg	ctatactcaa	caatggccc	ataccatcc	acaagtactg	2040
cttcgacgcc	ggcatcacca	cgctatacc	gtcgggtatc	gaggacgagt	tcctgagcac	2100
ggcgtccgag	ccgccccgga	agagcttcat	atccggcttc	aacgcaaatg	tgcggctctg	2160
gcagtcgcg	gctgatttgc	tgctggaaat	ccgcgtgt	caagatcaga	tgatgcagca	2220
ctttcgaggg	accatgcccc	cgaaccatgt	gctgcctcc	gccgacaggc	agcatctcga	2280
ttctctctat	gtccgcttca	tcacctgttt	ggacgatctc	ccggcgta	tccagtcgt	2340
cactctggcg	atggcagcga	ttggcagaagg	caacgggtct	gccgagtcca	agcagtacgt	2400
gatacagtgc	atcaacctgc	aggtgacgtt	tcactgtctg	cgcatggtaa	ttacgcagaa	2460
attcgaagac	ctctttatt	ttgctcctgg	cgttgagcag	gctgatctca	gaaagtgcgga	2520
gattgtgcga	gacatgctga	gggtgatgaa	cgaggcgccc	tttggggcc	tgcaggccaa	2580
tggcgagcca	aacgtgagtc	gtttccttgt	ctttcttctt	ttctgcacac	cctttcttc	2640
gacgaccccc	cctctctctt	tatatccctg	cggatatgta	tatcatcaag	cctcggact	2700
tgttgcata	ctgtcctgat	tatgttgtct	ggatgctgca	ggttggaaag	attcgcccta	2760
tcggagctag	tttgcggcc	atcatccatc	gcaaccagga	ttcacccctt	gctacgcgag	2820
ccaggagcga	ctttccgtg	ctttggata	ttctcacgctg	gctggactcg	aaggcgtcg	2880
actaa						2885
<210>	12					
<211>	723					
<212>	PRT					
<213>	Trichoderma reesei					
<400>	12					

Met Ala Thr Ala Ala Ala Ala Ala Gly Gly Ala Ala Val Ala Ala

1 5 10 15

Gly Ala Asp Thr Gly Ala Ala Gly Ser Ser Ser Thr Gly Pro Pro Gly

20	25	30
Leu Pro Gly Leu Pro Gly Thr Arg Thr Gly Ser Val Ala Met Gly Ser		
35	40	45
Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly Gly Pro Pro		
50	55	60
Ala Ala Gly Ala Gly Ala Val His Ala Leu Thr Thr Ser Pro		
65	70	75
Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala Ser Thr Thr		
85 90 95		
Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe His His His		
100	105	110
Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys Arg Arg Arg		
115	120	125
Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys Thr His Cys		
130	135	140
Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu Ala Arg Arg		
145	150	155
160		
Gly Pro Lys Ala Arg Lys Lys Ser Asp Gln Pro Gly Gln Pro Pro Pro		
165	170	175
Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln Met Pro Pro		
180	185	190
Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Ala Leu Gln Pro Phe Ala		
195	200	205
Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val Glu Pro Leu		
210	215	220
Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp Leu Pro Gly		
225 230 235 240		
Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp Ile His Leu		
245	250	255
Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg Val Ser Lys		
260	265	270
Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu Thr Pro Leu		

275	280	285
Val Tyr Glu Pro Ala Leu Arg Asp Val	Leu Ala Tyr Ile Phe Ser Gln	
290	295	300
305	310	315
Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln Leu Thr Pro		
320		
Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu Ser Trp Ala		
325	330	335
Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser Arg Leu Ala		
340	345	350
Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val Cys Ala Glu		
355	360	365
Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly Glu Ser Val		
370	375	380
Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His Gln His Leu		
385	390	395
390	395	400
Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala Ile Arg Tyr		
405	410	415
Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys Tyr Ser Trp		
420	425	430
425	430	435
His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met Gln Leu His		
435	440	445
Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala Glu Phe Arg		
450	455	460
455	460	465
Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser Ala Ala Ile		
465	470	475
470	475	480
Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe Asp Ala Gly		
485	490	495
485	490	495
Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe Leu Ser Thr		
500	505	510
500	505	510
Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe Asn Ala Asn		
515	520	525

Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Leu Glu Ile Arg Val
 530 535 540
 Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met Pro Pro Asn
 545 550 555 560
 His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser Leu Tyr Val
 565 570 575
 Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu Gln Ser Cys
 580 585 590

Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser Ala Glu Ser
 595 600 605

Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr Phe His Cys
 610 615 620

Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser Tyr Phe Ala
 625 630 635 640
 Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile Val Arg Asp
 645 650 655

Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu Gln Ala Asn

660 665 670
 Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala Ser Leu Leu

675 680 685
 Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr Arg Ala Arg

690 695 700
 Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu Asp Ser Lys

705 710 715 720
 Ala Ser Asp

<210> 13
 <211> 2185
 <212> DNA
 <213> Trichoderma reesei

<400> 13
 atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggcgg ccctccagct 60

gctggcgctg gcgtggcgcc tgtccacgcc ctcaccacct cgcccgagtc tgcctggcc	120
tcgcagcccg gctcgccaaac cgccctcaacc acgcgcgcg agaactcact cgtgtcggt	180
gcaacctcggt tccaccacca tcccagaggg cgtctggta gcagagcctg cgaccgctgc	240
cgccggcgca aggccaagtgcg aggtaccc acgcgtgtcg atagctgcac gcactgcgc	300
gatgcccacg tgcaagtgcac ttgcacccgt cccctggcgac gacgcggccc caaagcgagg	360
agaagagcg accagcccg ccagccgcct cctgatccga gtcgccttc caccgcggct	420
cgaccggcc agatgcccgc gccgctgacc ttctccggcc ccgcagtagc cgccgtcgac	480
cccttcgcct cgtcgctcgct gtcgcccgc gcccgtggg agcccgtcgaa gcccgtcagc	540
attgacaacgc gcctgccccg gcagccgcgt ggcgacactgc cccgccttc caccatccag	600
aacatctcgaa cgcgcgcagcg atggatacac ctggccaacgc ccatgacgct gcgcacacg	660
acgctagagc gctgtcgaa gcatgtatc gacctttct tcgactacact ctacccctc	720
accccccctgg tgtacgagcc ggcctccgg gacgtgctcg catacatctt ctcccagccc	780
ttgcctggcg tcaaccaacc atgcccgtc tcacagctca cgcgcaccc gaccaccggc	840
accacccccc tcaacgcgtgc cgagtcgtgg gccgggtttg gccagcccg cggctcgca	900
accgtcggtca gcaggctggc tccctggcc gactcgaccc tcaccctggt cacggccgtc	960
tgcgcagagg cagcattcat gtcacccaaag gacatccc cccggaggaga atccgtctct	1020
gagatcttgc tcaaggcctc tcgggactgc ctgcaccaggc acctcgaggc cgacactggag	1080
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgcgtcg	1140
gggaagccca agtactcgat gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1200
cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccgcgt	1260
cgctgcttt ggatcctgta cttggcgac aagtcaaggccg ctataactcaa caatcgcccc	1320
atcaccatcc acaagttactg cttcgacgcc ggcattacca cgctataacc gtcgggtatc	1380
gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagcttcat atccggcttc	1440
aacgcacatg tgcggctctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgtcg	1500
caagatcaga tgcgtcgac cttcgaggaccatgc cgaaccatgt gtcgcctcc	1560
gccgacaggc agcatctcgaa ttctctctat gtccgcgttca tcacctgcgtt ggacgtctc	1620
ccgcgttacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680
gccgagtcca agcagttacgt gatacagtgc atcaacgcgtc aggtgacgtt tcactgtctg	1740
cgcgttacc ttacgcgaa attcgaagac ctcttttatt ttgtcctgg cggtgagcag	1800
gctgtatctca gaaagtccgaa gattgtgcga gacatgtcgaa gggtgatgaa cgaggcgccc	1860
ttttggggcc tgcaggccaa tggcgagcga aacgtgagtc gtttccttgc ctcttcctt	1920

ttctgcacac cctttcttc gacgacccccc cctctcttt tataccctg cgatatgtat	1980		
tatcatcaag cctcggcaact tggtgctaat ctgtcctgat tatgttgtct ggatgctgca	2040		
ggttgaaaag attccctta tcggagctag tttgctggcc atcatccatc gcaaccagga	2100		
ttcacccttg gctacgcgag ccaggagcga ctttccgtg ctttgata ttctcacgcg	2160		
gctggactcg aaggcgtcg actaa	2185		
<210> 14			
<211> 678			
<212> PRT			
<213> Trichoderma reesei			
<400> 14			
Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Ala Gly			
1	5	10	15
Gly Pro Pro Ala Ala Gly Ala Gly Ala Val His Ala Leu Thr			
20	25	30	
Thr Ser Pro Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala			
35	40	45	
Ser Thr Thr Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe			
50	55	60	
His His His Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys			
65	70	75	80
Arg Arg Arg Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys			
85	90	95	
Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu			
100	105	110	
Ala Arg Arg Gly Pro Lys Ala Arg Lys Lys Ser Asp Gln Pro Gly Gln			
115	120	125	
Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln			
130	135	140	
Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Ala Leu Gln			
145	150	155	160
Pro Phe Ala Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val			
165	170	175	

Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp

180 185 190

Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp

195 200 205

Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg

210 215 220

Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu

225 230 235 240

Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile

245 250 255

Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln

260 265 270

Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu

275 280 285

Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser

290 295 300

Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val

305 310 315 320

Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly

325 330 335

Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His

340 345 350

Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala

355 360 365

Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys

370 375 380

Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met

385 390 395 400

Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala

405 410 415

Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser

420	425	430
Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe		
435	440	445
Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe		
450	455	460
Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe		
465	470	475
Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Leu Glu		
485	490	495
Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met		
500	505	510
Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser		
515	520	525
Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu		
530	535	540
Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser		
545	550	555
Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr		
565	570	575
Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser		
580	585	590
Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile		
595	600	605
Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu		
610	615	620
Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala		
625	630	635
Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr		
645	650	655
Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu		
660	665	670
Asp Ser Lys Ala Ser Asp		

675
<210> 15
<211> 930

<212> DNA
<213> Artificial Sequence
<220><223> promoter sequence
<400> 15 60
agcagtggct tatagcaata tcgtgcttgt ctctgcctc gctgaagcta tccctccctc
gtgtcccccct tcgtggagaa tgactgcagt aaaggagacg atacgcttt ggaaagggtcc 120
tgaagggggt ggctgttggg attgcaacct ctggcatttg tcaaatggcg ctttatgcgt
cgcgctcgcg agtgatgtt agaggaaagt caacgttgcg acggcttac aatttgcct 180
ttcattaagg cttattgccct gctaaattac tataaggaag gtccggcagct gggattacgc
atgtgttgc 300
agacacaatc tcgaacgaga cgctatcaga ggacaagtt tctgacaatc 360

420
tgtcgtggta ttgaacgctt ttcgtgtttg taaaccagca ttcatgaggt tcgagggccc
acgcattcat tgggatctca tcaatgaagc ggaatagttac aagaagacgc agcatccatt
acagacctca tgcagctaag gaaacggatc acccttaaag acgtggatgt agttccgtt 480
gctcgccac attagaaggc tggtaactata ggtgagggaa ggcgcattgtt gttttttatc
cgccatgata agaggagtgt aagccggctg catggacttt ttccataccc tgcattgcct
gtccatctt 540
aaattggcaa gctccaccc tccgttaaaa tcgcaggctt aaggcttct
aataaccact cagcatttgc tcaaccatgt ctgtgcctt ctcacctgac ctgcattctcc 660
720
840
ttttagccctc caggagttgg cgaattgtatc ccgatttga cgtcaaatcg cattgagata
actcacgata ttcgttttac aggctaaaga attgcctcca gcgttccagt tcccgctgta
ggccgagacg acactgcctc actaccagcc 900
930

<210> 16
<211> 930
<212> DNA
<213> Artificial Sequence
<220><223> promoter sequence
<400> 16 60
accccaatct gaggttcagg tttgggttcc ttccggatctt ccaagttctc tgaatcgat
gtccgcattt 120
attggatcg aagttgtga ttaatctcgaa gaatgtgcattt cttcgtca

cctcaacata cacggaatcc agcctttca tgaggaatcc ttactccttc ataccccaag	180
tgcccgggca gataattgta cccattcaca aatgaactta gactgtactc cgcaacttctt	240
tcaggctctt ctctcctcac gatgcccaa tgcttgccac tccacaaggt acatgttagta	300
atctgcagtt acicccccgc ttttgctac ttagccggca agaatgagat gcaagtttgt	360
tcctgtcggt agtattggct aaatggaaac acaagtagaa gaagagaaac agaaaaggtc	420
caataccggc tattcacaac ggtatctgcgt ctgtgtctga tagaactagt aaaagtccgc	480
agtatccgct gtcataaga tcatatacta aaacgtaaac taaacgcaat ggcttgaaa	540

aggggattga gacagaagat acaccaacga ccacccctgt tagtgacaga gatggcagtc	600
attcgactag cggtggcaa ttgggtgtcg ccttgtatc aggctaaata tctcgaggtg	660
ccggaaatc atggtgagga ggagttggca tgtgtggagc cgtgtatgc gtcggaccac	720
accaagaagc tgggttagca tttccgtccc gtatggggtt aacatggggt aacatgtgg	780
aattgcaa at gatgcattgg ggaagaatgc aacatcgta tcgtcataacc gctcaattta	840
aatattggc tttccgggg atcagatgga agaaggcaac agagagagag acaaggaaga	900
ccgtgagcca ttgaaggaca gccggacgca	930

<210> 17

<211> 930

<212> DNA

<213> Artificial Sequence

<220><223> promoter sequence

<400> 17

tctatggcca ctctgtcgcc tggatcatctg tatcagcacc aaagccatga ccgtgtatgt	60
ttgtatgtgcc atcttggcgt tctttcgca acacaccggg cttaggttt gtcactccct	120
ctccgcactt gagcacaacg ccggctgcct cgtgtccag aatacagtct ctttcgaa	180
cgagggagcc gattctccca gacttccaga aatggacatc cgatcttgca ctctgtcagc	240
ctgcagttga atgcgagttt gttgaagacc ttcaagactc acccgcatat gccagtagct	300

ttgatttgca gcaagacttg tccgcgtct ggttataaca ctgggtgttc gaccagttc	360
aggttgttgtt cggcagttgc ttgttagtgtat ggattgggtc gtggagctt aatgatccta	420
tgctcaagat cccgagctac agaaacgttgc ccgtggca tgggtgtga ggccatgggtt	480
taggtcgagg caatcgacag aagctcgaaa gtagtgaagt ttgggtgtaa tagaatgtcg	540
acggtgactt tgcgaatgcc caaatgtcaa gtgtaaagta ccaaactact gtcggcagg	600
tgcaagaaac gcaatattgc tggaaattat taaatcagaa ggtagacaa taagacgtga	660

tggcctctga ggagctaaa tgatcgctt ccaagaggtg gctaaaactt cccgattcg	720
caatcctcac ggccctctcg atgcggatgc aataacctct attagcctat atctgcggaa	780
gctatagacc gaggatgc gttgtcata tcttcacat catttcctg atcctttct	840
ttttattata tagaagctga agttgtacgg cacaactgtc aacacaccgt tgccgcggaa	900
ttctcccta tacttgacca atccatcaag	930
<210> 18	
<211> 930	
<212> DNA	
<213> Artificial Sequence	
<220><223> promoter sequence	
<400> 18	
tcaacatatg gttggactag gctcgctca cgtccagttt cttaagtca tatgttccag	60
gaccggaaag atgtctgcaa attggatagt ttttcttcg tagctgcctt cgatcaacct	120
gggacgcgat acgcctcagt tcaagtactt gatctgaagt accgtatgtt taggctgata	180
tccgatggaa tgccccaaa tacaaggctc gtcaacctga ttatgagaca attcttcac	240
ctacgtcaat ctgggtcat tctttgcaag aaacacgatc agtccatctc tcaggagcct	300
gtcgtgttt tctgctgaga ccattggct tcccctcgac agtatgtgc cccttggcta	360
actctcctcc acgcctaccc ggcctacctg acctaccagg gaggcggct tcttacctac	420
ctatcaaagg gcctccacgt ggctgcgtatc ctctaatcaa caaattcctt cctcgcaagc	480
ttcttcttc tgactgtctt ccaggttctg ctggccttgt tcctcgccaa ccagagcgat	540
cggcgcggaa agcagcgaga agcgagctct caggagaata agcgagtcga tactccatcg	600
gaagtctctc ggcgcgtgcag ctggcgttcc cgaggccact cgctttgc cgaatgagcg	660
tcacgacggc cttacagata taccgatagc gttgccctgc cacctgatca agttgtcgcc	720
tactatatgt ctgcggcat gaccgtatcc aaagctgcag ctggattct cacctctca	780
gagccgaccc ttggatagca ggggacacat ttggcctctt cagcaccgct ctggcgcgg	840
tctcctctgc gtttgatttg ttcttccact tctctttcc ctgatatctc ttcttgaact	900
gccggctgcg ttgtcgaact cgtgctcacg	930
<210> 19	
<211> 930	
<212> DNA	
<213> Artificial Sequence	

<220><223> promoter sequence

<400> 19

gcgattagca cgcacgcgt ctggaccgt gagctgacg ggctcaacag catggcg	60
tactgaggga taticgtcac atcggtgaa gaggcgaca tgagaaacga ctgtatact	120
gtattattta ctgtacatg tacggcagtc actaataacct gtctcaatca atgaagattg	180
atgtaaataa gcttaatcaa actgtctaaa atatgttac atttaaatgc tttgtctcg	240

gtttgtgcta ataaatgcgt gaggtgacat cgtccggaa tagtgagacg aagtggatc	300
accaagtgga gggactcatt attgttcatg cacgtccat gtacaaggcc taaatccgc	360
agctattccc cgtatgacct gtcaatactc gatacgatc agtagctgtt tggacaagca	420
agtactactg gtatgggtgt aacccagct gctacaaacc ctggaaagca gactctcaca	480
acctattact aggtactctc gttcggttca tcaaagttt gggcatctcg catgagagca	540
agacgttgcc ggcacatgcct cagctactta tgatgccatc tcatcaagcc catgagacca	600
cctatgcact gtcacaagag aacaaacttc gcctttgga acccgaaaa acaggggaga	660

tgcccggtgg tcttgcttg tcttaatcc gcttgatctg tgtccacagg aacggggaa	720
aaaacaactc gggttcgat ttgcgggtgt gcagcttcca acaggggtat ggaaaatcga	780
ccttcccaca cgacgggtta atatccacaa gccggcggaaa tggtaaattt ataaatcgag	840
catcattccc atctgggtgg atttcaaaca aaagaaaagg aagcgcacatcg tgaaagggt	900
gcaaggctcg gataaagtcc tcttctcgcc	930

<210> 20

<211> 930

<212> DNA

<213> Artificial Sequence

<220><223> promoter sequence

<400> 20

tgcgaactgg caatgagttt tcaaccatca aagtacggc gcagttgtcg tacatgggg	60
tccgagaata aaaaggctac atggtaaatt cgttaaaggca caaggtcaat aggacacag	120
tgtatgttcc ttggacagtgt tgcttacat ccgtatcc cacttgaca tacactcgct	180
ttgagtttgc gatcgtgcca acccgattga tacattacaa gcagaactga gccgctgtta	240
tcgacacact tgccgttcc tcagttagac gcttctaccc cgcttgcgtt gatatcacga	300
gggcattgtat acgtcaacca cttcatatca agtctgattt ctgaatgcct tcaatcatc	360
aggccccacaa atatccctac agcatgctca ctttggaaatg agttttctc tctttcg	420

ccagcaggcc	cgttggcgc	gtatgagcgc	cgaagctgac	actccagcga	cgtgctgatc	480
tagattctgg	acacgaatag	attcctattg	aaggaggatga	cgtccccacg	atgaactccc	540
gtcgagacag	tttgcattgc	atacagccgt	ttaacagcca	gccccgttact	ttatacataag	600
aacaacgtcg	ctgttgtgt	tgcgcact	tggcgatgtt	ggacggacgc	cgactcaacc	660
cctgcattcac	aatccgcca	gaagcgtcac	cgcggataaa	aaggctaaat	ggatggctt	720
caactcttcg	cgacatggaa	cgcgagcagg	tattacagaa	tgtatgaccgg	ttaaatctgg	780
cagtcctccat	ggtactgaag	tcaaagcaag	agggacaatt	cacaatggcc	ttgtatataaa	840
agcgataaca	tgtctccgt	caaggcctgg	atgagcagag	tttagagcaa	actgcctgct	900
tgggaatctt	cctctactt	tctcctcgac				930
<210>	21					
<211>	930					
<212>	DNA					
<213>	Artificial Sequence					
<220><223>	promoter sequence					
<400>	21					
ctggggtaac	agagctgttt	gcggggatga	agatcttgc	aatatgggt	aatggggat	60
ttttcctatt	ggaggaagca	ttgtcgaatc	caagggtatg	cgggctcggg	gaagggcac	120
tctccttagc	tgtcaaggct	ctagatactg	tgatatacg	tgtatataaa	gccaaaccca	180
agcaaagatc	cgcgtgtgt	ctgaacaagc	cgaatgccag	gcccttttg	ccggagcccc	240
tcatcgaaat	ccgctgtgca	acgggttcat	gggttcaaca	tgcggatag	cttcctcgca	300
agtctgtggg	gtagcagcag	cactccggca	gatgccactg	tgcctgtgaa	tgtggggaga	360
ggccagagta	tgggaacca	tgtgggaggc	agcacggaga	cactcgaagg	tcatgccttg	420
gagttgtcg	tactttgttt	ttgcatgcct	tggttccatc	atcgaaagga	gagcagcgg	480
agatagagcc	ttaagcatgg	gtaaaactcc	ttccgacaac	tctataacag	acgttagaa	540
aacaacaagt	tcccagttat	tacccagagc	tttcttacc	catcggcaca	acgtgcaacg	600
gccaagaatc	gctgccgatc	tcctcatctc	cagccctca	aagtctccaa	gctgcattca	660
gcatactccag	gcccccttg	cttgcacaca	ctttaggcata	aaggcttctg	ccctggctcg	720
gcagttggcag	aagcaacctg	ctacttgacg	acatgaaccc	gtttgtaccg	gctcggccat	780
gagatgtatgc	agttttccaa	tggcatagtt	gagtcgaagc	ggtttttca	cccttaatg	840
atgttaacatg	tatataacgg	atgaccatgc	cctccttgc	ttcacttcaa	cggtcttatt	900
atcctgcctc	gagccacaag	cagcgcagtc				930
<210>	22					

<211> 930

<212> DNA

<213> Artificial Sequence

<220><223> promoter sequence

<400> 22

catccacatc atcttgacga ggaaccatcg ggtggtcgga cggactctgc agatctgggg	60
tgcgctgac cgacgagggtg tcaattagcc tacgtgtgag taccatgtt gagaattgtt	120
acctatggat gcaagtgcct ctgaactatt actttatcca taggagcatc cgctaccaat	180
gtgcggtaa aattgcctt cagccaccct gtcctcggtt aattgagtga gacgtgcata	240
acatggccgg cagcatggca tggtagatga tattgggtga acgtgtcaga agaaaaggct	300
agaatattcg agacagcttg ctgatatgtt caaaaacttctt caagatattt gatatgttta	360

gagttactct tggcattata ctgtaatgtg aatgttagagt gtacgctaca gtacctcaat 420
ccaagaacctt ataacatacc tacataccta ctgagctaca ttttaggcag tgcctccgct 480
gagaagcttt gaattctact ttgtcgttt taactcgtga cgcatggact ggcgggtcat 540
cctgatacag aatcaggacc attgcatatg aaaaacagtc tgagacccaaa catccaatcg 600
ggagacaatg ctgccaatgaa agctgattta gcctatttcg acttctcaca actcaaagag 660
atgctttatg ttccggggaa gggatatgaa ctgtgacttag aatgcattcg ggctctgcca 720
aagaatgagt tgactatgag gaggcaaata tctgagtcgtat gtgagttagac cagtaaatgaa 780

cagctggggg taatcactta tgtttcacc ggatatactt catattgata taaaatgcta 840
tgcatctccaa gtcacacgta tactgagatg aacagcatct cttacagtt gctttccaca 900
taaaaattc ctcctgttac aagagccaaa 930

<210> 23

<211> 1059

<212> DNA

<213> Artificial Sequence

<220><223> promoter sequence

<400> 23

ggcaggcact ggctcgacg acatgtttg tatattgggtt gggactgcgg ccgcagcggg 60
ggcgggagct ggtggccgct atatgaattt ccggggcttg ctacaacagg taccactttg 120

accacccatg gctgccgtcg ccctgcttgg agctttcagg tcgcttccgg gcgttggcga 180
ggcaagtgg acggtgggga aatgacgaaa aatggtgtcat cgcctttaa ggttgtgtg 240

agttagtagtt ctactatgag gtacgtatgt agcagaagga tcgagctaga atctgccgc	300
attgcaaagg ttatctggaa agagggaaaag ggcctgaacc ggcataatgga tgcattttc	360
gtacgaacta ctatctgata acagtttagt actgttatcc atacaagag tcttatagaa	420
acactgcac gtaataaaat actcggtac tgctigaata tagtaataag atcaacatcc	480
tttacacctt agtctccgtg gattccagta aaagcgctca attctgactt ccgactctgt	540
tgatccccg tgtctgccc tcgggggtgg cttagacgctg cctcaacgcc catgtaccgg	600
cctgatgggg cccttgggg caccacaagt ccactaaacg aagcaactgg gacgggactc	660
gatagccctg agcagcagcc ggtctcagca gccaaaccagc ccagctggaa gcatcgctaa	720
ggggaggggg gcccaactac tacgtgtact actaggtaa taatgaattt gatgggacc	780
agccagccca acctaacttt ccagcctta tagctgcagc ctgcttcccc gtgcctcagc	840
cttttgctc ctctgctggc cggactcgga cctcttgca cctctgctcg accaacaatc	900
cctcttggtt caccctctcg ctttgctac ctgcacgctc aattcctcgc tgccgcctca	960
cctaaccgcg tgtgcttgc tgcctcacg ctggctcgc ctctgctcc gcgagccctcc	1020
ttttacactt ttcaacagct accccggccag aattcaaca	1059
<210> 24	
<211> 1963	
<212> DNA	
<213> Artificial Sequence	
<220><223> promoter sequence	
<400> 24	
gacggaatct tcacacccat tactcgta tactagatca gatgctggc gagtctgggg	60
aatgctggca ctggatggta taaacacccg actgcttgc gccggatcag cggcgtcga	120
ccgagctgga cgtgtcaata ctacatgtat catgacgcag agctgtttct cctctggcac	180
atgaacccaa agacacaagc gtcacccgt tcctgaggcag atcaggtaca gacttccata	240
tctcactggc cccatcttc gggctaaac cgccctccca cttgcacaa gatgccccaa	300
gccgttgcgc agagacgacg tggactcgac ctgcgtcagcc atcgatgccg tctcagctgc	360
cgcttcggcc aaccaacttg ctgcgaggca aaagtatttgc ctgcattgtat gccagccaa	420
ctccctcc tcgcctcc cattcagccg gcagcaatcc aaccacccac ccaccggcgc	480
agcggccacgc aaggcacagt caggactaac actcttaggc tgctggata catgtagaac	540
ggtctttgg tggtgggt tgctgcacat aaggtatata ctgtacatgc atatagtatg	600
catacagtag acgtgcttg gcgccctgt atgttagggaa tacatcatcg tccatctgt	660

gctccccgc aatccgctat ccggcatacc gagagctag gagccggctg ttgcccgtcc	720
ttgtacagt tacacctcca agtctcccc ggcctccat cacagctgcc tccgcacgg	780
gaccatcccc caagcaagtc gtccttgcc acttggtgga gggtcggccag tggacgagag	840
ctgctgcca tggccatgtat cagcacccgc ctgcgtggc tgactggata cttgcaagta	900
ttaaaaacggt cgacctcgcc ggttgttctc gtcgtcaact ctggccctct gactctgact	960
ctgactctt ccgtatcccc gtcgcategt gttgaaagct gtcctcttg tcttttctc	1020
tgccaaaccc tctcgctact aggtatctca acctttgtat acgagatcg tatcgccggc	1080
acgcagggtc tcatacctgac tcgtcctgtg tcggaatcgt ctcacgtctc actgccaaca	1140
agcagtttgc gatacgcaaa tcaatcgctt gcatcaggtg cattccagat ccggccctcg	1200
gacgtcgta tataatcgcc agcgccttc tcccaagag cccagcctt tggagcagga	1260
cccaggttt gtcatattt tgggtcgta gcctgcctt tgccatcgcc atctgttcc	1320
tcccacccaa cccacgcccc tcatcaatct ccatctcccc cgccaaaggc gactcgatcc	1380
ccgcgtgcac acacaaacca ccccgaaa aacgagacca acgtatatac ctgtggcagc	1440
gcccgcctaa tcccttccc agcgtgcgc ttgaccggg agtttgtat ccggtgagt	1500
cagtcggcgt ctcccgacca tccctccctc tctgtttttt agtggcttgc cccacccggc	1560
aatctgcgcc gcccagcgga gcccagcctg agcccgccct cgagctcgcc gccgttctgc	1620
gaccctatac ttcccgaggg cccaaaggcc gacgtcaaaa ctgctgggt atcctgcgg	1680
ctctggtcgg cctgcgttg ggcttggctt tgaaagaggt tgccatgac gggatcgcag	1740
tcccaagtt caaccctctc tctgccccac caatcatgac cccatgccc gtcgttatgc	1800
cactcatgtc atcgcagtgt gaacattgac ggagcttgac taatgtgaca gagtcattca	1860
gttttgttggaa gagccaagca gggtccttag ctagcgagcc ccgaccggag ggcagatgc	1920
ccgaccatag cacttataag gcggcgcttc cgctgtcgta act	1963
<210> 25	
<211> 1067	
<212> DNA	
<213> Artificial Sequence	
<220><223> promoter sequence	
<400> 25	
gtgctgcagt tgctacgcgt agtctcaact ggcattgcgt cgacaatact aaaatcagca	60
aaggctgagg tccgttgggt gaaaacttca ttctgtactcg tactcagacc gcaaattctg	120
gtatggtttg atgggttgggt tcatagtgtat catccatatac atggaaagcg ccgcctgc	180

cgagaagcaa caagtggcca acagcatcg cactatcg catcatgccc caactctc	240
ctagccgtag cgacccccc tcaaccagac agtgcggaa atccttgac gcactatcg	300
cgagttatg tacactgacc aacatgcgc cccggcgact tccagatgg gccaactca	360
gtcgagacgg cgttgtttt cagttatgg cggcgcttg atatgcgtt aatgaaaggc	420
gttccgctgg aactccctg gcccacgtt gttggatctc atgattgtgg tattttccg	480
tgcctcatca gttctcgagg aacttctgtt caccgaatgc catgtgaggc ccccgctc	540
ccccagctgc agcactccgt accccatata ccaggcaaat tatcgacccc gattgcttc	600
aatcaaaaatc gcttggcgcc gcttggat gcccgttgc aaggaaaagg caggatgtc	660
gtttagttt gtcataatc gtcgtcgac gggaaatgtc ccgttttgtt ccaatatcct	720
gtggcgacaa gggctatcaa aaagttccctc atccgcgtt tgccacaac cgtggcgat	780
gacaaagggtt cttccaaaaa gaggcgatcc ctggatcgcc cccatgtt gcttggatct	840
atcaaaacggc cagctggagc ttctgttctg ttctggatgtt cgtgtatggt ggtgttggg	900
tgtaccttgc ccgttgcgtt gccgtgttc tcttggatgtt cattgcattt aaatatcatc	960
gtgcctgtcg ccagtcttta ccgacttggg gattcttgc agctatttctt ctgcaagcgt	1020
tcaactgccc ccggtgagga acactcggtc agtgcacaaca caacacc	1067
<210> 26	
<211> 1124	
<212> DNA	
<213> Artificial Sequence	
<220><223> promoter sequence	
<400> 26	
cgaggcagat accgagatgt ttagcacgg aagtcatctc gtctagctc ccgtggggaa	60
aaaaaaaaaaa gtaagccgga accgcagggtc caggcgtcg gctgtacaaa ggagcaaca	120
cgatacaac acaataccac cctatggtag catcaataat ccgtaatgtt aatagcatgg	180
gcgaatattt gccctctgg tacggcgctc cgtaataaaa ataacacccc cccaaatctc	240
ggtgccccca attcggtct tttctccac cagttggaa ccctccctcc tatgtactct	300
ctctctctct ctcgttgggtt gggggagagc gagcgagcga gagagagagg cagtggaaagg	360
ctcaggctcg gtcttgctg cgccaaacctt ccgttgcgtt atctccctt gattagcgcc	420
ctcacctcag ttctcgccc ctcgtgacc gttccgcctt ccacaatctg gggagccgtt	480
aattagcaac tggactg agggttagat cccacgcgtc ccgttagtaa gtgtttaagc	540
ccgtgcgaa agggtgcgag caccggagtc gatgtcgaaa tactgtatgtc ctccctggaa	600

cccaattact ccggggggcg actaaagcag cccagccagc caaactagca gcaaggcaca	660
cgctcacact agtaatgccg ggccggcgga tggatgttgc tggatgttgc tttcttgttgc	720
cagggccagg aagccaaatc ttgtctgggg gggcgacata aaggccgct cgtcgccata	780
ccttgagcca cgcaccaca cacctgtccg gtttccctaa ttcccttatac caccaaacga	840
accataacaa caacgcctc ctcccttc atcccttc atcattgtat catcgctata	900
gcattgcgtt gccagctctg ttgtgtctgt cagttcatat aacaaagcct tccgtgtctc	960
tttcttcctc tcacactata tttttttttt ttccaccttc atcacaaaga cccacacctt	1020
ttccggccaca actgctccta ccaaaggcac actcgacacct ttgaatgatc tcctccctgt	1080
ccaaaagact caagtccaga agaagttcac ggagttccaa agcc	1124
<210> 27	
<211> 780	
<212> DNA	
<213> Artificial Sequence	
<220><223> promoter sequence	
<400> 27	
agactagcgg ccggccccct tatccagct gttccacgtt ggcctggccc tcagtttagcg	60
ctcaactcaa tgccctcac tggcgaggcg agggcaagga tggaggggca gcatgcctg	120
agtggagca aagcggccgc catggagca gcgaaccaac ggagggatgc cgtgtttgt	180
cgtggctgct gtggccaatc cggcccttg gttggctcac agagcgttgc tgtgagacca	240
tgagctatta ttgcttagta cagttatagag agaggagaga gagagagaga gagagagggg	300
aaaaaaaggta aggttgaagt gagaaaaaaaaaaaaaaa aaatccaaacc actgacggct	360
gccggctctg ccaccccccct ccctccaccc cagacaacct gcacactcag cgcgcagcat	420
cacctaattct tggctcgct tcccgagct caggttgttt ttttttctc tctccctcg	480
cgaagccgccc ctgttccct tatttatttc cctctccatc ctgtctgcc tttggccat	540
ctgccccctt gtctgcattt ctttgcacg catgcctta tcgtgtctc ttttttact	600
cacgggagct tgacgaagac ctgactcgtg agcctaacct gctgatttct ctccccccct	660
cccgaccggc ttgacttttg tttctccctcc agtaccttat cgcgaagccg gaagaaccct	720
ctttaacccc atcaaacaag ttgtacaaa aaagcaggct cgcggccgc ccccttcacc	780
<210> 28	
<211> 58	

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 28

tcagggttat tgtctcatgg ccattttaggc ctggcaggca ctggctcgga cgacatgt 58

<210> 29

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 29

agagccctgg gccggagctg ctgagccat tggtaattc tggcggtta gctgttga 58

<210> 30

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 30

tcaacagcta cccgcaga attcaacaat gggctcagca gctccggccc agggctct 58

<210> 31

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 31

tcgtaaataa acaagcgtaa ctagctagcg tagttatgc gagcaacatt gcacgaaac 59

<210> 32

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 32

gtttcgta atgttgctcg cataacctac gctagctagt tacgctgtt tatttacga 59

<210> 33
 <211> 58
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 33
 acatgtcgctc cgagccagtg cctgccaggc ctaaatggcc atgagacaat aaccctga 58
 <210> 34
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 34
 aggtgtaaaga cgggggagta gcgcagcatt gttgaattct ggccgggttag ctgttga 57

 <210> 35
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 35
 tcaacagcta cccgcaga attcaacaat gctgcgtac tccccgtct tacacct 57
 <210> 36
 <211> 59
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 36
 tcagggttat tgtctcatgg ccatttaggc ctagactagc ggccggtccc cttatccca 59
 <210> 37
 <211> 55
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 37

agagccctgg gccggagctg ctgagccat ggtgaagggg gcggccgcgg agcct	55
<210> 38	
<211> 55	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 38	
aggctccgcg gccccccct tcaccatggg ctcagcagct ccggccagg gctct	55
<210> 39	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 39	
tggataagg ggaccggccg ctagtctagg cctaaatggc catgagacaa taaccctga	59
<210> 40	
<211> 51	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 40	
tgtaagacgg gggagtagcg cagcatggtg aaggggccgg ccgcggagcc t	51
<210> 41	
<211> 51	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 41	
aggctccgcg gccccccct tcaccatgt gcgtactcc cccgtttac a	51
<210> 42	
<211> 7882	
<212> DNA	
<213> Artificial Sequence	

<220><223> plasmid pYL1

<400> 42

ggcaggcact ggctcgacg acatgtttg tatattggc gggactgcgg ccgcagcggg	60
ggcgggagct ggccggcgcg atatgaattt ccggcgcttg ctacaacagg taccacttg	120
accacccatg gctgccgtcg ccctgcttgg agcttcagg tcgcttccgg gcgttggcga	180
ggcaagttgg acgggtggga aatgacgaaa aatggtgcat cgccttgta ggtgtgtgtg	240
agttagtagtt ctactatgag gtacgtatgt agcagaagga tcgagctaga atctgccgc	300
attgcaaagg ttatctggaa agaggaaaag ggcctgaacc ggcataatgga tgcattttc	360
gtacgaacta ctatctgata acagtttagt actgttatcc atacaagag tcttataagaa	420
acactgcac gtaataaaaat actcggtacg tgcttgaata tagtaataag atcaacatcc	480
tttacacctt agtctccgtg gattccagta aaagcgctca attctgactt ccgactctgt	540
tgtatccccg tgcgtgccta tcgggggtggt ctagacgctg cctcaacgccc catgtaccgg	600
cctgatgggg ccctggggg caccacaagt ccactaaacg aagcactggg gacggactc	660
gatagccctg agcagcagcc ggtctcagca gccaaccagc ccagctggaa gcatcggcta	720
ggggagggggg gccaactac tacgtgtact actaggtaca taatgaattt gatgggaccc	780
agccagccca actaacttt ccagcctta tagtgcagc ctgtttcccc gtgcctcacg	840
cttttgctc ctctgctggc cggactcgga cctcttgcga cctctgctcg accaacaatc	900
cctcttgttgc caccctctcg cttttgcata ctcgacgctc aattccctgc tgccgcctca	960
cctaaccgcg tgcgttgc tgcctcaacg ctgcgtcgcc ctctgctcc gcgagccctcc	1020
ttttacactt ttcaacagct accccgcac aattcaacaa tgggctcagc agctccggcc	1080
cagggtctg tagctgcagc tgcaggcggc cctccagctg ctggcgctgg cgctggcgct	1140
gtccacgccc tcaccaccc gcccggatct gcctcggcct cgcagccgg ctcgccaacc	1200
gcctcaacca cggccggca gaactcactc gtgtcggtcg caacctcggtt ccaccacca	1260
cccagaggcc gtcgtgttag cagagcctgc gaccgtcgcc gccggcgcaaa ggccaagggtc	1320
agtctagccc ctgtgttgc tgcgttgcata ctgtgttgc tgcgttgc tgcgtgtcg	1380
ctgatgtgc tgcgttgc tgcgttgc tgcgttgc tgcgttgc tgcgttgc tgcgttgc	1440
cttcataatgt ctttactgccc cgtgtctctt ctcccgatcc cggccccc tcctccgtc	1500
ctcttctctt cgtgtgtctgt catgcgtaca aagcatacat acaatacatc agcatacatg	1560
gcaagcggtt gtttgcgttgc agagttgtgt gtattgtatt gcactgcctt cacaactcg	1620
tcataactgct gcagcctcac cccaaacaccc acctcgatcc ccatgcgtcg ctactcccc	1680
gtcttacacc tggatactctt ctcccttgcac ccactgacca atgctttcc cggccaaag	1740

tgcgagtacc tcagcgctgt cgatagctgc acgcactgcc gcgatgccca cgtgcagtgc	1800
acttcgacc tgccctggc gcgacgcggc cccaaagcga ggaagaagag cgaccagccc	1860
ggccagccgc ctccgtatcc gagctcgctc tccaccgcgg ctgcacccgg ccagatgccg	1920
ccgcccgtga ccttctccgg ccccgagta gcccgcgtgc agcccttcgc ctcgtcgctg	1980
ctgtcgcccg acgcggcgtg ggagcccgta gagccgctca gcattgacaa cggcctgccc	2040
cggcagccgc tgggcaccc gcccggcctc tccaccatcc agaacatctc gacgcgccag	2100
cgtggatac acctggccaa cgccatgacg ctgcgaaca cgcgcgtaga ggcgcgtctcg	2160
aagcgatgta tcgacctt cttcgactac ctctacccct tcacccccc ggtgtacgag	2220
cgggccctcc gggacgtgct cgccatacattt ttctccagc cttgcctgg cgtcaaccaa	2280
ccatcgccgc tgtcacagct cacgcccagac ccgaccaccc gcaccacccccc cctcaacgct	2340
gccgagtcgt gggccggctt tggccagccc agcggctcgca gaaccgtcg cagcaggctg	2400
gctccctggg ccgactcgac cttcacccctg gtcacggccg tctgcgcaga ggcagcatc	2460
atgctaccca aggcattttt ccccgaaagga gaatccgtct ctgagatctt gctcgaagcc	2520
tctcgggact gcctgcacca gcacctcgag gccgacctgg agaatccgac ggcacactcg	2580
attgccattc gctacttcca ctccaaactgc ctccacgctg cggggaaagcc caagtactcg	2640
tggcacatata ttggcgagggc catccgcgtg ggcgcaggta tgcagctgca cgaggaggct	2700
gccctcgagg ggctcgccc catcgaggca gagttccgcc gtcgcgtctt ttggatccgt	2760
tacttggcg acaagtcaacgc cgctataactc aacaatcgcc ccatcaccat ccacaagttac	2820
tgcttcgacg ccggcatcac cacgtatacc ccgtcggtt tcgaggacga gttcctgagc	2880
acggcgtccg agccgccccg gaagagcttc atatccggct tcaacgc当地 tgcggctc	2940
tggcagtccg cggctgattt gctgcgtggaa atccgcgtgc tgcaagatca gatgtatgcag	3000
cacttcgag ggaccatgcc cccgaaccat gtgcgtccct ccgcgcacag gcagcatctc	3060
gattctctct atgtccgctt catcacctgc ttggacgatc tccgcgtt cctccagtcg	3120
tgcaactctgg cgtggcagc gatggcagaa ggcaacgggt ctgcgcgttc caagcgtac	3180
gtgatacagt gcatcaacct gcaggtgacg tttcaactgtc tgcgcgttgt aattacgcag	3240
aaattcgaag acctcttta ttttgcgttcc ggcgttgcgc aggctgtatct cagaaagtgc	3300
gagatgtgc gagacatgct gagggtgtatg aacgaggcgc cttttgggg cttgcaggcc	3360
aatggcgcagc caaacgtgag tcgtttccct gtctttctc ttttctgcac accctttct	3420
tcgacgcaccc cccctcttc tttatatccc tgcggatatg tatatcatca agcctcgca	3480

cttggatgctg caggttggaa agattgcct	3540
tatcgagct agtttgcgtt ccatcatcca tcgcaaccag gattcacccct tggctacgcg	3600
agccaggaggc gactttccg tgcttttggaa tattctcacg cggctggact cgaaggcgctc	3660
ggaccaactg aggaataacgt ccactaccgt tggtggctaa atgtgtgttgaacaacaaa	3720
aaatgtcaaa gtcgggtgtaa atatggccag gatctttgtt ttattcccccc ttcaagcgttt	3780
ctgggtattt ccccttggtt tactctttc tggttttcc agcactgtt tttccagcag	3840
tggggggaaac aaaaggcggtt tctttccct atgcagggg ttgtccgatt tagcattga	3900
gtgtacatct tccctacatt actaggtact taatgagctt atggagatct cccgtcatc	3960
cggatattca tcacgttgggt gtatatatcc gtgggtggct ttgaaacctg gagttgggtt	4020
gcaatgcagt gacgcctttt gcgaaggacc aaaataagcg aaggatgaag tctgaatagg	4080
atacgaactg gctacctatg ggtgagcatg aaatgaagcg gtcggggaaa tggcggagaa	4140
acgctcgacg taacgctgtt gggtttctcc gtttgcgtca atgttgcgtc cataacctac	4200
gctagctagt tacgcttggtt tatttacgac aagatctaga agattcgaga tagaataata	4260
ataataacaa caatttgcctt ctttttcca cttttcagttt cttactctcc cttctgacat	4320
tgaacgcctc aatcagtcag tcgccttgta cttggcacgg taatcctccg tggttctgat	4380
atcctcaggg gtagcaaagc ctttcatgcc atcgataatg tcatccagag tgaggatggc	4440
aaagatgggg atgcgtact ctttcctcag ctgcctaattt gcactcggtc caggcttggaa	4500
gtcgctgcca tccgcagcgg ggagcttctc catgcggtcc agggccacga cgtatgcggc	4560
gacgatgcgg ccctccttgg tggatcttc aatggcggtcc ctcttggcggt tgccggcggt	4620
gatgacgtcg tcgacaatca ggaccctttt gcccatttggc gaagcggccga cgtatgttgc	4680
gcccctcgccg tggcccttgg cttcatttgcgtt gtcaaaacgag taggagacgc ggttcaggtt	4740
ctggggcgcc agctcgccga gcttgcgtt gatggcgagcacagggta tgcccttgta	4800
ggccggggccg aagacgatgt cgaactctag gccggcccttc tcctggccct cgtatgttgg	4860
ctttgcattttt gggggggcga tggcgccggc gagggcgccgtt gttgtggaaatt cggccggcgtt	4920
gaagaagtag ggggatattcc gtttgcgtt gatggactt gagctcgaag ctgcattttt tgaggacgcc	4980
gccgtcgatg gggattttttt ggaaggatccctt ctgtttttttttt ggcagctggg aggtggtagc	5040
cattttttttt gatggata gtttgcgtt tttttttttt ttttttttttt ttttttttttt ttttttttttt	5100
ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt	5160
ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt	5220

ccccacaagct tctgtcttagg tatgtatgtg gctataaggc acccaaggca gaactcatct 5280
tgaagattgt ctagagtgat tttaccgctg atgaaatgac tggactccct cctccgtc 5340
ttatacgaaa aattgcctga ctctgcaag gttgttgtc ttggaagatg atgtgcccc 5400
ccatcgctct tatctcatac cccggcatct ttctagattc tcatcttcaa caagaggggc 5460
aatccatgtat ctgcgatcca gatgtgttc tggccata ctctgccttc aggttgatgt 5520
tcacttaatt ggtgacgaaat tcagctgatt tgctgcagta tgctttgtgt tggttcttc 5580
caggcttgtg ccagccatga gcgcgtttag agcatgttgt cacctataaa ctcgagtaac 5640

ggccacatata ttttcaactac ttgaatcaca tacctaattt tgatagaatt gacatgttta 5700
aagagctgag gtagctttaa tgcctctgaa gtattgtgac acagcttctc acagagttag 5760
aatgaaaagt tggactccccc ctaatgaagt aaaagttcg tctctgaacg gtgaagagaca 5820
tagatccggc atcaactacc tggctagact acgacgtcaa ttctgcggcc ttttgacctt 5880
tatatatgtc cattaatgca atagattctt ttttttttt ttttttttt tttttttttt 5940
tttttttttgc cccaatttcg cagatcaaag tggacgttat agcatcataa ctaagctcag 6000
ttgctgaggg aagccgtcta ctacccatgc ccatccatcc agctccatac cttgatactt 6060

tagacgtgaa gcaattcaca ctgtacgtct cgcacgtctc cttcccgctc ttgcgtcccc 6120
actggggtcc atggtgctg tattgtcccc tccttaatta aggccattta ggccgttgc 6180
ggcgtttttcataggtctcc gccccctgaa cgagcatcac aaaaatcgac gctcaagtc 6240
gaggtggcga aacccgacag gactataaag ataccaggcg ttcccccctg gaagctccct 6300
cgtcgctct cctgttccga ccctgcccgt taccggatac ctgtccgcct ttctccctc 6360
gggaagcgtg ggccttctc atagctcagc ctgttaggtat ctcaagttcg tgtaggtcgt 6420
tcgtcccaag ctgggtctg tgcacgaacc ccccggttag cccgaccgct ggcgccttatac 6480

cggttaactat cgtcttgagt ccaacccggta aagacacgac ttatcgccac tggcagcagc 6540
cactggtaac aggattagca gagcgaggta tggtagcggt gctacagagt tcttgaagt 6600
gtggcctaactac tacggctaca ctagaaggac agtattttgt atctgcgtc tgctgaagcc 6660
agttaccttc gaaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag 6720
cggtggtttttttgcgatc agcagcagat tacggcggaa aaaaaggat ctcaagaaga 6780
tccttgcatttttctacgg ggtctgacgc tcaatggaaac gaaaactcac gtttaaggcct 6840
gcagggccgat ttttggcat gagattatca aaaaggatct tcaatggat ccttttaat 6900

taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggc tgacagttac 6960
caatgcttaa tcaatggaggc acctatctca gcgatctgc tatttcgttc atccatagg 7020
gcctgactcc cggcgtgtat gataactacg atacgggagg gcttaccatc tggcccccagt 7080

gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcgc aataaaccag	7140
ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct	7200
attaattgtt gccgggaagc tagagtaagt agttgccag ttaatagtt gcgcaacgtt	7260
gttgcatttgc acaggcat cgtggtgtca cgctcgctgt tttgtatggc ttcatcagc	7320
tccggttccc aacgatcaag gcgagttaca tgatccccca tttgtgcaa aaaagcggtt	7380
agctccttcg gtcctccat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg	7440
gttatggcag cactgcataa ttcttact gtcatgccat ccgtaaagatg ctttctgt	7500
actggtgagt actcaaccaa gtcattctga gaatagtgtt tgccgcgacc gagttgtct	7560
tgccggcgt caatacgggtaatacccgccacatagca gaactttaaa agtgcicatc	7620
attgaaaac gttttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt	7680
tcgatgtaac ccactcggtc acccaactga tcttcagcat ctttacttt caccagcg	7740
tctgggttag caaaaacagg aaggcaaaat gccgcaaaaa agggataag ggcgacacgg	7800
aaatgttcaa tactcataact cttcctttt caatattttt gaagcatttta tcagggttat	7860
tgtctcatgg ccatttaggc ct	7882
<210> 43	
<211> 7279	
<212> DNA	
<213> Artificial Sequence	
<220><223> plasmid pYL2	
<400> 43	
ggcaggcact ggctcgacg acatgttttgc tatattgggtt gggactgcgg ccgcagcggg	60
ggcgggagct ggtggcggcg atatgaattt ccggcgttg ctacaacagg taccacttg	120
accacccatg gctgccgtcg ccctgcttgg agcttcagg tcgtttccgg gcgttggcga	180
ggcaagtgg acgggtggaa aatgacgaaa aatggtgcat cgccttgcgtt ggtgtgtgt	240
agttagttagt ctactatgag gtacgtatgt agcagaagga tcgagctaga atctgcggc	300
attgcaaagg ttatctggaa agaggaaaag ggcctgaacc ggcataatggaa tgcattttc	360
gtacgaacta ctatctgata acagtttagt actgttatcc atacaagag tcttataagaa	420
acactgcattc gtaataaaat actcggttagc tgcttgaata tagtaataag atcaacatcc	480
tttacacttct agtctccgtg gattccagta aaagcgctca attctgactt ccgactctgt	540
tgtatccccg tgcgtgcctt tcgggggtggt ctagacgctg cctcaacgcc catgtaccgg	600
cctgatgggg cccttgggg caccacaagt ccactaaacg aagcactgg gacgggactc	660

gatagccctg	agcagcagcc	ggtctcagca	gccaaccagc	ccagctggaa	gcatggcta	720
ggggaggggg	gcccaactac	tacggtact	actaggta	aatgaattt	gatgggacc	780
agccagccca	acctaacttt	ccagcctta	tagctgcagc	ctgcttcccc	gtgcctcagc	840
cttttigtc	ctctgctggc	cggactcgga	cctcttgca	ccitctgtcg	accaacaatc	900
cctcttgg	caccctctcg	cttttgctac	ctcgacgctc	aattccctcg	tgccgcctca	960
cctaaccgcg	tgtgcttgac	tgccctcagc	ctcggtcgc	ctcctgctcc	gcgagccctcc	1020
ttttacactt	ttcaacagct	accccgccag	aattcaacaa	tgctgcgcta	ctccccgtc	1080
ttacacctgg	atactctctc	cttgccacca	ctgaccaatg	ctcttccccg	cccaaagtgc	1140
gagtaacctca	gchgctgtcg	tagctgcacg	cactgcccgg	atgcccacgt	gcagtgcact	1200
ttcgacctgc	ccctggcgcg	acggggcccc	aaagcgagga	agaagagcga	ccagccggc	1260
cagccgcctc	ctgatccgag	ctcgctctcc	accggggctc	gaccggcca	gatggcccg	1320
ccgctgacct	tctccggccc	cgcagtagcc	gchgctgcagc	ccttcgcctc	gtcgctgctg	1380
tcgcccgacg	cggcctggga	gcccgtcgag	ccgctcagca	ttgacaacgg	cctggccgg	1440
cagccgtgg	gchgacctgc	ccggctctcc	accatccaga	acatctcgac	gchgccagcg	1500
tggatacacc	tggccaacgc	catgacgctg	cgcaacacga	cgctagagcg	cgtctcgaag	1560
cgtatgtatcg	acctttttt	cgactaccc	tacccctca	ccccctgg	gtacgagccg	1620
gccctccggg	acgtgctcg	atacatcttc	tcccagecct	tgcctggcg	caaccaacca	1680
tcgcccgtgt	cacagctcac	gccagacccg	accacggca	ccacccccc	caacgctgcc	1740
gagtcgtgg	ccggctttgg	ccagccagc	ggctcgcgaa	ccgtcggcag	caggctggct	1800
ccctggggccg	actcgacctt	caccctggtc	acggccgtct	gchgagaggc	agcattcatg	1860
ctacccaagg	acatttccc	cgaaggagaa	tccgtctcg	agatctgtct	cgaaggctct	1920
cgggactgcc	tgcaccagca	cctcgaggcc	gacctggaga	atccgacggc	caactcgatt	1980
gccattcgct	acttccactc	caactgcctc	cacgctgccc	ggaagccaa	gtactcgigg	2040
cacatatttgc	cgaggccat	ccgcctggcg	caggctatgc	agtcgacgca	ggaggctgcc	2100
ctcgaggggc	tctgtccccat	cgaggcagag	ttccggcg	gctgcttttgc	gatcctgtac	2160
ttggcgcaca	agttagccgc	tatactcaac	aatcgcccc	tcaccatcca	caagtactgc	2220
ttcgacgccc	gcatcaccac	gctatacccg	tcgggtatcg	aggacgagtt	cctgagcacg	2280
gcgtccgagc	cggccggaa	gagcttcata	tccggcttca	acgcaaatgt	gcccgtctgg	2340
cagtccgcgg	ctgatttgct	gctggaaatc	cgcgtgtcg	aagatcagat	gatgcagcac	2400
tttcgaggga	ccatgcccc	gaaccatgtg	ctgcctccg	ccgacaggca	gcatctcgat	2460
tctctctatg	tccgcttcat	cacctgcttg	gacgatctcc	ccgggtacct	ccagtcgtgc	2520

actctggcga tggcagcgat ggcagaaggc aacgggtctg ccgagtccaa gcagtagtg	2580
atacagtgca tcaacctgca ggtgacgtt cactgtctgc gcatggtaat tacgcagaaa	2640
ttcgaagacc tctcttattt tgctcctggc gttgaggcagg ctgatctcg aaagtccgag	2700
attgtgcgag acatgctgag ggtgatgaac gaggcgccct tttggggcct gcaggccaaat	2760
ggcgagccaa acgtgagtcg ttccttgc tcttctctt tctgcacacc ctttctcg	2820
acgacccccc ctctctctt atatccctgc ggatatgtat atcatcaagc ctcggcactt	2880
gttctaattc tgtcctgatt atgttgcag gttgaaaaga ttccgcctat	2940
cggagctagt ttgctggcca tcatccatcg caaccaggat tcacccttgg ctacgcgagc	3000
caggagcgac ttccctgtc ttttggatat tctcacgcgg ctggactcga aggctcgga	3060
ccaactgagg aatacgtcca ctaccgttg tggctaaatg tgtgttgaa caacaaaaaa	3120
tgtcaaagtc ggtgtaaata tggccaggat ctttgcgttta ttcccccttc agcgttgctg	3180
gttattttcc ctttgcgttac tctttctgt ttttccagc acttgcgtttt ccagcagtgg	3240
ggggaaacaaa aggctttct ttcccctatg ccaggggttg tccgatttag catttgcgtg	3300
tacatcttcc ctacattact aggtactaa ttagcttgcgat gagatctccc gtcattccgg	3360
atattcatca cgttgggtta tatatccgtg gttggcttg aaacctggag ttgggttgca	3420
atgcagtgac gcctttgcg aaggacacaaa ataagcgaag gatgaagtct gaataggata	3480
cgaactggct acctatgggt gaggcatgaaa tgaaggcgtc gggaaatgg cgagaaacg	3540
ctcgacgtaa cgctgttggc tttctccgtt tcgtcaatg ttgcgtcgat aacctacgct	3600
agcttagttac gcttgcgttac ttacgacaag atctagaaga ttgcgatag aataataata	3660
ataacaacaa ttgccttctt cttccaccc tttcgttactt actctccctt ctgacattga	3720
acgcctcaat cagtcgtcg ctttgcgttactt ggacggtaa tccctccgtt tcttgcgtatc	3780
ctcagggtta gcaaaggccct tcatgccatc gataatgtca tccagagtga ggttggcaaa	3840
gatggggatg ccgtactcct tccctcgtcg tccatggca ctggcccgagc gcttggatc	3900
gtcgccatcc gcagcgggaa gtttctccat gcggtccagg gccacgacga tgccggcgac	3960
gatgccgccc tccttggta tcttctcaat ggcgtccctc ttggcggtgc cggcggtgat	4020
gacgtcgatcg acaatcggaa ccctttgcctt cttgagcgaa ggcggacga ttttgccccc	4080
ctcggccgtgg tccttggccct ctttgcgtcg aaacgagttag gagacgcgtt ccaggttctg	4140
ggggcccgacg tccggcgttgc ttttgcgttgc ggcggacac agcggatgc ctttgcgttgc	4200
cggccgaaacg acgtatgtcgaa actctaggcc ggccttctcc tggccctcgaa ttttgcgttgc	4260

gccttaactac ggctacacta gaaggacagt atttggtata tgcgctctgc tgaagccagt	6060
taccccgaa aaaagagttg gtagctcttg atccggcaaa caaaccaccc ctggtagccgg	6120
tggttttttt gtttgcagc agcagattac ggcgcagaaaa aaaggatctc aagaagatcc	6180
tttgcattttt tctacggggt ctgacgcica gtggacgaa aactcacgtt aaggccgtca	6240
ggcccgattt tggtcatgag attatcaaaa aggatctca cctagatctt tttaaattaa	6300
aatgaagtt ttaaatcaat ctaaagtata tatgagtaaa ctggctgta cagttaccaa	6360
tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcata catagttgcc	6420

tgactccccg tcgtgttagat aactacgata cgggagggct taccatctgg ccccaagtgt	6480
gcaatgatac cgcgagaccc acgctcacgg gctccagatt tatacgaat aaaccagcca	6540
gccggaaggg ccgagcgcag aagtggctt gcaactttat ccgcctccat ccagtctatt	6600
aattgttgcg gggaaagctag agtaagtagt tcgcccgtt atagtttgcg caacgttgg	6660
gccattgcta caggcatcgt ggtgtcacgc tcgtcggtt gtagggcttcc attcagctcc	6720
ggttcccaac gatcaaggcg agttacatga tccccatgt tgtcaaaaaa agcggttagc	6780
tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatac actcatggtt	6840

atggcagcac tgccataattc tcttactgtc atgcccattcg taagatgttt ttctgtgact	6900
ggtgagtagt caaccaagtc attctgagaa tagtgtatgc ggcgaccggat ttgtcttgc	6960
ccggcgtcaa tacggataa taccgcgcca catagcagaa cttaaaagt gctcatcatt	7020
ggaaaacgtt ctcggggcg aaaactctca aggatcttac cgctgttgcg atccagttcg	7080
atgttaaccca ctcgtgcacc caactgtatct tcagcatctt ttacttgc cagcgttct	7140
gggtgagcaa aaacacggaaag gcaaaatgcc gcaaaaaagg gaataaggc gacacggaaa	7200
tgttgaatac tcatactttt ctttttcaa tattattgaa gcatttatca gggttattgt	7260

ctcatggcca tttaggcct	7279
----------------------	------

<210> 44
<211> 7603
<212> DNA
<213> Artificial Sequence
<220><223> plasmid pYL3
<400> 44

agactagcgg ccggccccct tatcccaget gttccacgtt ggcctgcccc tcagttacgc	60
ctcaactcaa tggccctcac tggcgaggcg agggcaagga tggagggca gcatcgccgt	120
agttggagca aagcggccgc catggagca gcgaaccaac ggagggatgc cgtgtttgt	180

cgtggctgct gtggccaatc cgggcccttg gttggctcac agagcggtgc tgtgagacca	240
tgagctatta ttgcttagta cagtatagag agaggagaga gagagagaga gagagagggg	300
aaaaaaaggta aggttgaagt gagaaaaaaa aaaaaaaaaa aaatccaacc actgacggct	360
gccggctctg ccaccccccct ccctccaccc cagacaacct gcacactcag cgcgcagcat	420
cacctaattct tggctcgct tcccgagct cagggtgtt ttttttctc tctccctcg	480
cgaagccgcc ctgttccct tatttatttc cctctccatc ctgtctgccc tttggtccat	540
ctgcccctt gtctgcatct ctttgcacg catgcctta tcgtcgctc ttttttact	600
cacgggagct tgacgaagac ctgactcgta agcctaacct gctgatttct ctccccccct	660
cccgaccggc ttgacttttg tttctccctc agtaccttat cgcaagccg gaagaaccct	720
ctttaacccc atcaaacaag ttgtacaaa aaaggcaggct ccgcggccgc ccccttacc	780
atgggctcag cagctccggc ccagggtct gtagctgcag ctgcaggcgg ccctccagct	840
gctggcgctg ggcgtggcgc tgcgtccaccc ctcaccaccc cgcccgagtc tgcctcgcc	900
tcgcagcccg gctcgccaaac cgcctcaacc acgcggccgc agaactcact cgtgtcgct	960
gcaacctcgat tccaccacca tcccgagggc cgtctggta gcagaggctg cgaccgtgc	1020
cgccggcgca aggccaaggta cagtctagcc ctttgctgt tgcttgcatc tctgttgtca	1080
ttgctccctcc tcctgctgct gctgatgctg ctgctccctcc tcctccctct cctcccccgtc	1140
tcctggtccc tggccctgc tcttcatatg tccttactgc ccgtgtctcc tctcccccgtt	1200
cccggtcccc ctccctccgt cctcttctcc tgcgtgtctg tcatgcgtac aaagcataca	1260
tacaatacat cagcatacat ggcaagcggt gtgttgtt gagagtgtg tgtattgtat	1320
tgcactgcct tcacaactcg ttcatactgc tgcagccctca cccaaacacc gacctcgct	1380
tccatgctgc gctactcccc cgtttacac ctggataactc tctccctgccc accactgacc	1440
aatgctttc cccgccccaa gtgcgagttac cttagcgctg tcgatagctg cacgcactgc	1500
cgcgatgccc acgtgcagtg cacttcgac ctggccctgg cgcgacgcgg ccccaaagcg	1560
aggaagaaga gcgaccagcc cggccagecg cctctgtatc cgagctcgct ctccacccg	1620
gctcgaccgg gccagatgcc gcccggctg accttctccg gccccggcgt agccgcgt	1680
cagcccttcg cctcgctgct gctgtcgccc gacgcccgtt gggagccgt cgagccgt	1740
agcattgaca acggcctgcc cggcagccg ctggccgacc tgccggccct ctccaccatc	1800
cagaacatct cgacgcgcca gcgatggata cacctggcca acggcatgac gctgcgcaac	1860
acgacgctag agcgcgtctc gaagcgatgt atcgacccct tttcgacta cctctacccc	1920

ctcacccccc tggtgtacga gccggccctc cgggacgtgc tcgcatacat cttctccag	1980
cccttgctg gcgtaacca accatcgccg ctgtcacagc tcacgccaga cccgaccacc	2040
ggcaccaccc ccctcaacgc tgccgagtcg tggccggct ttggccagcc cagcggtcg	2100
cgaaccgtcg gcagcaggct ggctccctgg gccgactcga ccttcacccct ggtcacggcc	2160
gtctgcgcag aggcagcatt catgtaccc aaggacattt tccccgaagg agaatccgtc	2220
tctgagatct tgctcgaagc ctctcggac tgcctgcacc agcacctcga ggccgacctg	2280
gagaatccga cggccaactc gattgccatt cgctacttcc actccaactg cctccacgct	2340
gcggggaagc ccaagtactc gtggcacata tttggcgagg ccatccgcct ggcgcaggtc	2400
atgcagctgc acgaggaggc tgccctcgag gggctcgcc ccatcgaggc agagttccgc	2460
cgtcgctgct ttggatct gtacttggc gacaagttag cgcgtataact caacaatcg	2520
cccatcacca tccacaagta ctgcttcgac gccggcatca ccacgtata cccgtcggt	2580
atcgaggacg agttccttag cacggcgtcc gagccgcccc ggaagagctt catabccgc	2640
ttcaacgcaa atgtcggtc ctggcagtcc gcccgtgatt tgctgtgga aatccgcgt	2700
ctgcaagatc agatgtgca gcactttcga gggaccatgc cccgaaacca tgtgtgccc	2760
tccggccgaca ggcagcatct cgattctctc tatgtccgct tcatcacctg cttggacgat	2820
ctccggccgt acctccagtc gtgcactctg gcatggcag cgatggcaga aggcaacggg	2880
tctggcgagt ccaaggacta cgtgatacag tgcataacc tgcaggtgac gtttactgt	2940
ctgcatgg taattacgca gaaattcga gacccctctt atttgtcc tggcggttag	3000
caggctgtac tcagaaagtc ggagatttg cgagacatgc tgagggtgat gaacgaggcg	3060
ccctttggg gcctgcagggc caatggcag ccaaacgtga gtcgttctt tgtctttct	3120
ctttctgca caccctttc ttgcacgacc cccctctct ctttatatcc ctgcggat	3180
gtatatcatc aagcctcgac acttgggtct aatctgtcct gattatgtt tctggatgt	3240
gcaggttggaa aagattcgcc ttatcggtc tagtttgctg gccatcatcc atcgcaacca	3300
ggattcaccc ttggctacgc gagccaggag cgactttcc gtgtttgg atattctcac	3360
gcggctggac tgcaggcgt cgaccaact gaggaatacg tccactaccg ttgttgct	3420
aatgtgtgtt ggaacaacaa aaaatgtcaa agtcgggtta aatatggcca ggatctttgt	3480
gttattcccc ctgcgtt gctgggtatt tccctttgt ttactcttt ctgtttttc	3540
cagcacttgt tttccagca gtggggggaa caaaaggcgt ttctttcccc tatgccaggg	3600
gttgccgat ttagcatttgc agtgcatac ttccctacat tactaggtac ttaatgagct	3660
tatggagatc tcccgatcatt ccggatattc atcagttgg tgcataatc cgtgggttgc	3720
tttggaaacct ggagttgggt tgcaatgcag tgacgcctt tgcgaaggac caaaataagc	3780

gaaggatgaa gtctgaatag gatacgaact ggctacccat gggtagcat gaaatgaagc	3840
gttcgggaa atggcggaga aacgctcgac gtaacgctgt tggtttctc cgttcgtgc	3900
aatgttgcgc gcataaccta cgctagctag ttacgctgtt ttatcacga caagatctag	3960
aagatcggat atagaataat aataataaca acaatttgcc tcttcttcc acctttcag	4020
tcttactctc ccttctgaca ttgaacgcct caatcagtc gtcgccttgt acttggcacg	4080
gtaatccccc gtgttcttga tttccctcagg ggttagcaag cccttcatgc catcgataat	4140
gtcatccaga gtgaggatgg caaagatggg gatgccgtac tccttcctca gctcgccat	4200
ggcactcggt ccaggcttgg agtcgtcgcc atccgcagcg gggagcttct ccatcggtc	4260
cagggccacg acgatgccgg cgacgatgcc gccccttgc gtatcttct caatggcgtc	4320
cctcttggcg gtgcggcggt tgatgacgtc gtcgacaatc aggaccctct tgccctttag	4380
cgaagcgcgg acgatgttgc cgcgcctcgcc gtggccttgc gcttccttgc ggtcaaaacga	4440
gtaggagacg cggcccggt tctggggcgc cagctcgccg agcttgcgtt tgatggcgga	4500
gcacagcggg atgccttgcgtt aggccgggcc gaagacgatg tcgaactcta gcccggcctt	4560
ctcctggcc tcgatgttgc tcttcgttgc ggcggaggcg atggcggccgg cgaggcgcc	4620
cgtgtggat tcgcccgcgt tgaagaagta gggggatatc cgcttgcgtt tgagctcgaa	4680
gctgccaac ttgaggacgc cgcgtcgat ggcggatttgc aggaagtcct gcttgcgtt	4740
aggcagctgg gaggtggtag ccattctgtt ggatttggat agtgcctta ttctctgatt	4800
tgaacagtag atcaggacga gtgagaggga tgcagaggtt ggattggagt gtttgcgtt	4860
taaaatttag aggccgcggc tatcgagttt tcacatggaa gtcaaagcgt acagtgcgag	4920
cttgcgtt ggtcttagta tcccaacaagc ttctgtctgt gtatgtatgt ggctataagt	4980
cacccaaggc agaactcata ttgaagatttgc tctagatgtt tttaccgtt gatgaaatgt	5040
ctggactccc tcctcctgtt cttatacgaa aaattgcctt actctgcataaa ggttgcgtt	5100
cttgcgtt gatgtgcggccccc cccatcgctt ttatctcata cccgcgtt tttcttagatt	5160
ctcatcttca acaagagggg caatccatgtt tctgcgttcc agatgtgcattt ctggcctcat	5220
actctgcctt cagggttgcgtt ttacttaat tgggtgcgtt ttgcgtgcgtt	5280
atgcgttgcgtt tgggttctt ccaggcttgcgtt gcccgcgtt ggcgcgtt ggcgcgtt	5340
tcacccatataa actcgagtttgc cggccacata ttgttgcgtt cttgcgtt gatgttgcgtt	5400
ttgtatgttgcgtt aaagagctgtt ggttagtttgcgtt atgcgttgcgtt agtattgtgtt	5460
cacagcttgcgtt cacaggtgtt gatgttgcgtt ttggactccc cctaatgtt gaaatgtt	5520
gtctctgtt ggtgttgcgtt atagatccgg catcaactac ctggcttagac tacgacgtca	5580
attctgcggc ctttgcgtt ccattatgtt ccattatgtt gatgttgcgtt ttgttgcgtt	5640

ttttttttt ttttttttt ttttttttt gcccaatttc gcagatcaa gtggacgta	5700
tagcatcata actaagctca gttgctgagg gaagccgtct actaccttag cccatccatc	5760
cagctccata cctgatact ttagacgtga agcaattcac actgtacgtc tcgcagctct	5820
cctcccgct ctgtttccc cactgggtc catggcgt gtatgtccc ctccttaatt	5880
aaggccattt aggcgttgc tggcggtttt ccataggctc cgccccctg acgagcatca	5940
caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc	6000
gtttccccc ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata	6060
cctgtccgccc ttctccctt cgggaagcgt ggccgtttct catagctcac gctgttagta	6120
tctcagttcg gttaggtcg ttgcgtccaa gctgggctgt gtgcacgaac ccccggtca	6180
gccccgaccgc tgccgccttat ccggtaacta tcgtcttgc tccaacccgg taagacacga	6240
cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg	6300
tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtattgg	6360
tatctgcgt ctgtgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg	6420
caaacaacc accgctggta gcgggtgttt ttttgggtc aagcagcaga ttacgcccag	6480
aaaaaaagga tctcaagaag atcctttgat ctttctacg gggctgacg ctcagtggaa	6540
cggaaaactca cgtaaggcc tgcaggccg atttggtca tgagattatc aaaaaggatc	6600
ttcacctaga tcctttaaa taaaaatga agtttaat caatctaaag tatatatgag	6660
taaacttggt ctgacagtta ccaatgctta atcgtgagg cacctatctc agcgatctgt	6720
ctatttcgtt catccatagt tgcctgactc cccgtcggt agataactac gatacggag	6780
ggcttaccat ctggcccccag tgctgcaatg ataccgcgag acccacgctc accggctcca	6840
gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact	6900
ttatccgcct ccatccagtc tattatgt tgccggaaag ctagagtaag tagttcgcca	6960
gttaatagtt tgcgcaacgt tttggcatt gctacaggca tcgtgggtc acgctcgctg	7020
tttggatgg ctccattcag ctccgggtcc caacgatcaa ggcgagttac atgatcccc	7080
atgttgtca aaaaagcggt tagtccttc ggtcctccga tcgttgtcag aagtaagttg	7140
gccgcagtgt tatactcat ggttatggca gcactgcata attctttac tgtcatgcca	7200
tccgttaagat gctttctgt gactggtgag tactcaacca agtcattctg agaatagtgt	7260
atgcggcgcac cgagtgcgtc ttggccggcg tcaatacggg ataataccgc gccacatagc	7320
agaactttaa aagtgctcat cattggaaaa cgttctcgg ggccaaaact ctcaaggatc	7380

ttaccgctgt tgagatccag ttcgatgtaa cccactcgta cacccaactg atcttcagca	7440
tctttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa	7500
aaggaaataa gggcacacg gaaatgtta atactcatac tttcccttt tcaatattat	7560
tgaagcattt atcagggtta ttgtctcatg gccatttagg cct	7603
<210> 45	
<211> 7000	
<212> DNA	
<213> Artificial Sequence	
<220><223> plasmid pYL4	
<400> 45	
agactagcgg ccggccccct tatccagct gttccacgtt ggccctggcc tcagtttagcg	60
ctcaactcaa tgccctcac tggcgaggcg agggcaagga tggagggca gcatgcctg	120
agtggagca aagcgccgc catggagca gcgaaccaac ggagggatgc cgtgtttgt	180
cgtggctgct gtggccaatc cggcccttg gttggctcac agagcgttgc tgtgagacca	240
tgagctatta ttgcttagta cagtagatag agaggagaga gagagagaga gagagagggg	300
aaaaaaaggtg aggttgaagt gaaaaaaa aaaaaaaaaaa aaatccaacc actgacggct	360
gccggctctg ccacccccc ccctccaccc cagacaacct gcacactcag cgccgagcat	420
cacctaattt tggctcgct tcccgagct cagttgttt ttttttctc tctccctcg	480
cgaagccgcc ctgttccct tatttatttc cctctccatc ctgtctgcc tttggccat	540
ctgccccctt gtctgcattt ctttgcacg cattgcattt tctgtctc ttttttact	600
cacgggagct tgacgaagac ctgactcgta agcctaacct gctgatttct ctccccccct	660
cccgaccggc ttgacttttg tttctccctt agtaccttat cgcgaagccg gaagaacct	720
cttaacccc atcaaacaag ttgtacaaa aaagcaggct ccggccgc ccccttacc	780
atgctgcgtt actccccgt ttacacccgt gatactctt cttgccacc actgaccaat	840
gctttcccc gccaaagtg cgagtaccc agcgctgtcg atagctgcac gcactgccgc	900
gatccccacg tgcagtgcac ttgcacccgtt cccctggcgcc gacgcccccaaaagcgagg	960
aagaagagcg accagcccgccagccctt cctgtccga gctgcgtctc caccgcggct	1020
cgaccggcc agatgccgc gccgctgacc ttctccggcc ccgcaggtagc cgccgtcgag	1080
cccttcgcctt cgtcgctgtt gtcggccac gccgcgtggg agccgcgtga gccgcgtcagc	1140
attgacaacg gcctgccccg gcagccgtg ggcgacccgtc ccggcccttc caccatccag	1200
aacatctcga cgccgcagcg atggatacac ctggccaaacg ccatgacgct gcgcaacacg	1260

acgctagagc gcgtctcgaa gcgtatgtatc gacctttct tcgactacct ctacccctc	1320
accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	1380
ttgcctggcg tcaaccaacc atcgccgtcg tcacagctca cgccagaccc gaccacggc	1440
accacccccc tcaacgctgc cgagtcgtgg gccggcttg gccagccag cggctcgca	1500
accgtcgcca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	1560
tgcgcagagg cagcattcat gctacccaag gacatttcc cccaaggaga atccgtctct	1620
gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	1680
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctgca	1740
gggaagccca agtactcgta gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1800
cagctgcacg aggaggctgc cctcgagggg ctgcgtcccc tcgaggcaga gttccgcgt	1860
cgctgctttt ggatcctgta ctgggcgac aagttagccg ctataactcaa caatcgcccc	1920
atcaccatcc acaagtactg cttcgacgcc ggcacatcacca cgctataaccc gtcgggtatc	1980
gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagcttcat atccggcttc	2040
aacgc当地 tgcggctctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgtc	2100
caagatcaga tgcggctctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgtc	2160
gccgacaggc agcatctcgat ttctctctat gtccgcgttca tcacctgctt ggacgatctc	2220
ccgcgttacc tccagtcgtcg cactctggcg atggcagcga tggcagaagg caacgggtct	2280
gccgagtcca agcagtagctg gatacgtgc atcaacctgc aggtgacgtt tcactgtctg	2340
cgc当地 ttacgc当地 attcgaagac ctcttttattt ttgtcctgg cggtgagcag	2400
gctgatctca gaaagtccgaa gattgtgcgaa gacatgctga gggtgatgaa cgaggccccc	2460
ttttggggcc tgcaggccaa tggcggccaa aacgtgagtc gttccctgt ctcttcttt	2520
ttctgcacac cttttcttc gacgacccccc cttctcttt tatatccctg cgatgtatgta	2580
tatcatcaag cctcggact tggctaat ctgtcctgat tatgttgtct ggatgctgca	2640
gggtgaaaag attcgcctta tggagctag tttgtggcc atcatccatc gcaaccagga	2700
ttcacccctt gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg	2760
gctgactcg aaggcgtcg accaactgag gaatacgtcc actaccgtt ttggctaaat	2820
gtgtgttgaa acaacaaaaa atgtcaaagt cgggtaaat atggccagga tctttgttt	2880
attccccctt cagcgttgct gggatttcc cttttgttta ctctttctg tttttccag	2940
cactgtttt tccagcgttg ggggaacaa aaggcgttcc tttccctat gccaggggtt	3000

agcttctcac agagtgagaa tgaaaagtt gactccccct aatgaagtaa aagttcgctc	4920
tctgaacggta gaagagcata gatccggcat caactacctg gctagactac gacgtcaatt	4980
ctgcggcctt ttgacctta tatatgtca ttaatgcaat agattcttt tttttttt	5040
ttttttttt tttttttt ttttttttgc caatitcgca gatcaaagtg gacgtatag	5100
catcataact aagctcagtt gctgagggaa gccgtctact accttagccc atccatccag	5160
ctccataacct tgatactta gacgtgaagc aattcacact gtacgtctcg cagctctcct	5220
tcccgctt gctcccccac tggggtccat ggtgcgtta tcgtccctc ctttaattaag	5280
gccatttagg cggttgctgg cgttttcca taggctccgc cccctgacg agcatcaca	5340
aaatcgacgc tcaagtcaaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt	5400
tccccctgga agtccctcg tgcgctctcc tggccgacc ctggcgctta ccggataacct	5460
gtccgcctt ctcccttcgg gaagcgtggc gctttctat agtcacgct gtaggtatct	5520
cagttcggtg taggtcggtc gctccaagct gggctgtgt cacgaacccc ccgttcagcc	5580
cgaccgtgc gccttatccg gtaactatcg tctttagtcc aaccggtaa gacacgactt	5640
atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcgtgc	5700
tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggat	5760
ctgcgtctg ctgaagccag ttaccttcgg aaaaagagtt ggtagcttt gatccggcaa	5820
acaaaccacc gctggtagcg gtggttttt tgtttgcag cagcagatta cgcgcagaaa	5880
aaaaggatct caagaagatc ctttgatctt ttctacggg tctgacgctc agtggaaacga	5940
aaactcacgt taaggcctgc agggccgatt ttggcatga gattatcaaa aaggatctc	6000
acctagatcc tttaaattttaaaaatgaagt tttaaatcaa tctaaagtat atatgagtaa	6060
acttggctcg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta	6120
tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgt acgggaggc	6180
ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc ggctccagat	6240
ttatcagcaa taaaccagcc agccggaagg gccgagcgcga gaagtggtcc tgcaactta	6300
tccgcctcca tccagtttat taatttgtgc cggagaacta gagtaagtag ttgcgcagtt	6360
aatagttgc gcaacgttgc tgccattgtc acaggcatcg tgggtcactc ctcgtcgttt	6420
ggtagtggctt cattcagctc cggttccaa cgatcaaggc gagttacatg atccccatg	6480
tttgtcaaaa aagcggttag ctccctcggt cctccgatcg ttgtcagaag taagtggcc	6540
gcagtggtat cactcatggt tatggcagca ctgcataatt ctcttactgt catccatcc	6600
gtaagatgtt tttctgtgac tggtgagttac tcaaccaagt cattctgaga atagtgtatg	6660
cggcgaccga gttgtcttg cccggcgtca atacggata ataccgcgc acatagcaga	6720

actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctt 6780

ccgctgttga gatccagttc gatgtAACCC actcgac ccaactgatc ttcagcatct 6840

tttactttca ccagcgttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 6900

ggaataaggg cgacacggaa atgttgaata ctcatactt tccttttca atattattga 6960

agcatttac agggattttg tctcatggcc atttaggcct 7000

<210> 46

<211> 940

<212> PRT

<213> Trichoderma reesei

<400> 46

Met Leu Ser Asn Pro Leu Arg Arg Tyr Ser Ala Tyr Pro Asp Ile Ser

1	5	10	15
---	---	----	----

Ser Ala Ser Phe Asp Pro Asn Tyr His Gly Ser Gln Ser His Leu His

20	25	30
----	----	----

Ser Ile Asn Val Asn Thr Phe Gly Asn Ser His Pro Tyr Pro Met Gln

35	40	45
----	----	----

His Leu Ala Gln His Ala Glu Leu Ser Ser Ser Arg Met Ile Arg Ala

50	55	60
----	----	----

Ser Pro Val Gln Pro Lys Gln Arg Gln Gly Ser Leu Ile Ala Ala Arg

65	70	75	80
----	----	----	----

Lys Asn Ser Thr Gly Thr Ala Gly Pro Ile Arg Arg Arg Ile Ser Arg

85	90	95
----	----	----

Ala Cys Asp Gln Cys Asn Gln Leu Arg Thr Lys Cys Asp Gly Leu His

100	105	110
-----	-----	-----

Pro Cys Ala His Cys Ile Glu Phe Gly Leu Gly Cys Glu Tyr Val Arg

115	120	125
-----	-----	-----

Glu Arg Lys Lys Arg Gly Lys Ala Ser Arg Lys Asp Ile Ala Ala Gln

130	135	140
-----	-----	-----

Gln Ala Ala Ala Ala Ala Gln His Ser Gly Gln Val Gln Asp Gly

145	150	155	160
-----	-----	-----	-----

Pro Glu Asp Gln His Arg Lys Leu Ser Arg Gln Gln Ser Glu Ser Ser

165	170	175
Arg Gly Ser Ala Glu Leu Ala Gln Pro Ala His Asp Pro Pro His Gly		
180	185	190
His Ile Glu Gly Ser Val Ser Ser Phe Ser Asp Asn Gly Leu Ser Gln		
195	200	205
His Ala Ala Met Gly Gly Met Asp Gly Leu Glu Asp His His Gly His		
210	215	220
Val Gly Val Asp Pro Ala Leu Gly Arg Thr Gln Leu Glu Ala Ser Ser		
225	230	235
Ala Met Gly Leu Gly Ala Tyr Gly Glu Val His Pro Gly Tyr Glu Ser		
245	250	255
Pro Gly Met Asn Gly His Val Met Val Pro Pro Ser Tyr Gly Ala Gln		
260	265	270
Thr Thr Met Ala Gly Tyr Ser Gly Ile Ser Tyr Ala Ala Gln Ala Pro		
275	280	285
Ser Pro Ala Thr Tyr Ser Ser Asp Gly Asn Phe Arg Leu Thr Gly His		
290	295	300
Ile His Asp Tyr Pro Leu Ala Asn Gly Ser Ser Pro Ser Trp Gly Val		
305	310	315
Ser Leu Ala Ser Pro Ser Asn Gln Phe Gln Leu Gln Leu Ser Gln Pro		
325	330	335
Ile Phe Lys Gln Ser Asp Leu Arg Tyr Pro Val Leu Glu Pro Leu Leu		
340	345	350
Pro His Leu Gly Asn Ile Leu Pro Val Ser Leu Ala Cys Asp Leu Ile		
355	360	365
Asp Leu Tyr Phe Ser Ser Ser Ser Ala Gln Met His Pro Met Ser		
370	375	380
Pro Tyr Val Leu Gly Phe Val Phe Arg Lys Arg Ser Phe Leu His Pro		
385	390	395
Thr Asn Pro Arg Arg Cys Gln Pro Ala Leu Leu Ala Ser Met Leu Trp		
405	410	415
Val Ala Ala Gln Thr Ser Glu Ala Ser Phe Leu Thr Ser Leu Pro Ser		

420	425	430
Ala Arg Ser Lys Val Cys Gln Lys Leu Leu Glu Leu Thr Val Gly Leu		
435	440	445
Leu Gln Pro Leu Ile His Thr Gly Thr Asn Ser Pro Ser Pro Lys Thr		
450	455	460
Ser Pro Val Val Gly Ala Ala Ala Leu Gly Val Leu Gly Val Ala Met		
465	470	475
Pro Gly Ser Leu Asn Met Asp Ser Leu Ala Gly Glu Thr Gly Ala Phe		
485	490	495
Gly Ala Ile Gly Ser Leu Asp Asp Val Ile Thr Tyr Val His Leu Ala		
500	505	510
Thr Val Val Ser Ala Ser Glu Tyr Lys Gly Ala Ser Leu Arg Trp Trp		
515	520	525
Gly Ala Ala Trp Ser Leu Ala Arg Glu Leu Lys Leu Gly Arg Glu Leu		
530	535	540
Pro Pro Gly Asn Pro Pro Ala Asn Gln Glu Asp Gly Glu Gly Leu Ser		
545	550	555
Glu Asp Val Asp Glu His Asp Leu Asn Arg Asn Asn Thr Arg Phe Val		
565	570	575
Thr Glu Glu Glu Arg Glu Arg Arg Arg Ala Trp Trp Leu Val Tyr		
580	585	590
Ile Val Asp Arg His Leu Ala Leu Cys Tyr Asn Arg Pro Leu Phe Leu		
595	600	605
Leu Asp Ser Glu Cys Ser Asp Leu Tyr His Pro Met Asp Asp Ile Lys		
610	615	620
Trp Gln Ala Gly Lys Phe Arg Ser His Asp Ala Gly Asn Ser Ser Ile		
625	630	635
Asn Ile Asp Ser Ser Met Thr Asp Glu Phe Gly Asp Ser Pro Arg Ala		
645	650	655
Ala Arg Gly Ala His Tyr Glu Cys Arg Gly Arg Ser Ile Phe Gly Tyr		
660	665	670

Phe Leu Ser Leu Met Thr Ile Leu Gly Glu Ile Val Asp Val His His
 675 680 685
 Ala Lys Ser His Pro Arg Phe Gly Val Gly Phe Arg Ser Ala Arg Asp
 690 695 700
 Trp Asp Glu Gln Val Ala Glu Ile Thr Arg His Leu Asp Met Tyr Glu
 705 710 715 720
 Glu Ser Leu Lys Arg Phe Val Ala Lys His Leu Pro Leu Ser Ser Lys
 725 730 735

 Asp Lys Glu Gln His Glu Met His Asp Ser Gly Ala Val Thr Asp Met
 740 745 750
 Gln Ser Pro Leu Ser Val Arg Thr Asn Ala Ser Ser Arg Met Thr Glu
 755 760 765
 Ser Glu Ile Gln Ala Ser Ile Val Val Ala Tyr Ser Thr His Val Met
 770 775 780
 His Val Leu His Ile Leu Leu Ala Asp Lys Trp Asp Pro Ile Asn Leu
 785 790 795 800
 Leu Asp Asp Asp Asp Leu Trp Ile Ser Ser Glu Gly Phe Val Thr Ala

 805 810 815
 Thr Ser His Ala Val Ser Ala Val Glu Ala Ile Ser Gln Ile Leu Glu
 820 825 830
 Phe Asp Pro Gly Leu Glu Phe Met Pro Phe Phe Tyr Gly Val Tyr Leu
 835 840 845
 Leu Gln Gly Ser Phe Leu Leu Leu Ile Ala Asp Lys Leu Gln Ala
 850 855 860
 Glu Ala Ser Pro Ser Val Ile Lys Ala Cys Glu Thr Ile Val Arg Ala
 865 870 875 880

 His Glu Ala Cys Val Val Thr Leu Ser Thr Glu Tyr Gln Arg Asn Phe
 885 890 895
 Ser Lys Val Met Arg Ser Ala Leu Ala Leu Ile Arg Gly Arg Val Pro
 900 905 910
 Glu Asp Leu Ala Glu Gln Gln Arg Arg Glu Leu Leu Ala Leu
 915 920 925

Tyr Arg Trp Thr Gly Asn Gly Thr Gly Leu Ala Leu

930 935 940

<210> 47

<211> 341

<212> PRT

<213> Trichoderma reesei

<400> 47

Met Asp Leu Arg Gln Ala Cys Asp Arg Cys His Asp Lys Lys Leu Arg

1 5 10 15

Cys Pro Arg Ile Ser Gly Ser Pro Cys Cys Ser Arg Cys Ala Lys Ala

20 25 30

Asn Val Ala Cys Val Phe Ser Pro Pro Ser Arg Pro Phe Arg Pro His

35 40 45

Glu Pro Leu Asn His Ser His Glu His Ser His Ser His Asn

50 55 60

His Asn Gly Val Gly Val Ser Phe Asp Trp Leu Asp Leu Met Ser Leu

65 70 75 80

Glu Gln Gln Gln Glu Gln Gln Gln Gly Gln Pro Gln His Pro Pro Pro

85 90 95

Pro Val Gln Thr Leu Ser Glu Arg Leu Ala Ala Leu Leu Cys Ala Leu

100 105 110

Asp Arg Met Leu Gln Ala Val Pro Ser Ser Leu Asp Met His His Val

115 120 125

Ser Arg Gln Gln Leu Arg Glu Tyr Ala Asp Thr Val Gly Thr Gly Phe

130 135 140

Asp Leu Gln Ser Thr Leu Asp Ser Leu Leu His His Ala Gln Asp Leu

145 150 155 160

Ala Ser Leu Tyr Ser Glu Ala Val Pro Ala Ser Phe Asn Lys Arg Thr

165 170 175

Thr Ala Ala Glu Ala Asp Ala Leu Cys Ala Val Pro Asp Cys Val His

180 185 190

Gln Asp Arg Thr Ser Leu His Thr Thr Pro Leu Pro Lys Leu Asp His

195

200

205

Ala Leu Leu Asn Leu Val Met Ala Cys His Ile Arg Leu Leu Asp Val

210

215

220

Met Asp Thr Leu Ala Glu His Gly Arg Met Cys Ala Phe Met Val Ala

225 230 235 240

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

245

250

255

Val Gly Thr Phe Val Ala Pro Thr Asp Thr Ala Ala Ser Met Leu Leu

260

265

270

Ser Val Val Val Glu Leu Gln Thr Val Leu Val Ala Arg Val Lys Asp

275

280

285

Leu Val Ala Met Val Asp Gln Val Lys Asp Asp Ala Arg Ala Ala Arg

290

295

300

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

305 310 315 320

Glu Ser Thr Leu Gly Glu Trp Ser Arg Phe Lys Asp Gly Leu Val Ser

325

330

335

Ala Arg Leu Leu Lys

340

<210> 48

<211> 940

<212> PRT

<213> Trichoderma reesei

<400> 48

Met Leu Ser Asn Pro Leu Arg Arg Tyr Ser Ala Tyr Pro Asp Ile Ser

1 5 10 15

Ser Ala Ser Phe Asp Pro Asn Tyr His Gly Ser Gln Ser His Leu His

20 25 30

Ser Ile Asn Val Asn Thr Phe Gly Asn Ser His Pro Tyr Pro Met Gln

35 40 45

His Leu Ala Gln His Ala Glu Leu Ser Ser Ser Arg Met Ile Arg Ala

50 55 60

Ser Pro Val Gln Pro Lys Gln Arg Gln Gly Ser Leu Ile Ala Ala Arg

65 70 75 80

Lys Asn Ser Thr Gly Thr Ala Gly Pro Ile Arg Arg Arg Ile Ser Arg

85 90 95

Ala Cys Asp Gln Cys Asn Gln Leu Arg Thr Lys Cys Asp Gly Leu His

100 105 110

Pro Cys Ala His Cys Ile Glu Phe Gly Leu Gly Cys Glu Tyr Val Arg

115 120 125

Glu Arg Lys Lys Arg Gly Lys Ala Ser Arg Lys Asp Ile Ala Ala Gln

130 135 140

Gln Ala Ala Ala Ala Ala Gln His Ser Gly Gln Val Gln Asp Gly

145 150 155 160

Pro Glu Asp Gln His Arg Lys Leu Ser Arg Gln Gln Ser Glu Ser Ser

165 170 175

Arg Gly Ser Ala Glu Leu Ala Gln Pro Ala His Asp Pro Pro His Gly

180 185 190

His Ile Glu Gly Ser Val Ser Ser Phe Ser Asp Asn Gly Leu Ser Gln

195 200 205

His Ala Ala Met Gly Gly Met Asp Gly Leu Glu Asp His His Gly His

210 215 220

Val Gly Val Asp Pro Ala Leu Gly Arg Thr Gln Leu Glu Ala Ser Ser

225 230 235 240

Ala Met Gly Leu Gly Ala Tyr Gly Glu Val His Pro Gly Tyr Glu Ser

245 250 255

Pro Gly Met Asn Gly His Val Met Val Pro Pro Ser Tyr Gly Ala Gln

260 265 270

Thr Thr Met Ala Gly Tyr Ser Gly Ile Ser Tyr Ala Ala Gln Ala Pro

275 280 285

Ser Pro Ala Thr Tyr Ser Ser Asp Gly Asn Phe Arg Leu Thr Gly His

290 295 300

Ile His Asp Tyr Pro Leu Ala Asn Gly Ser Ser Pro Ser Trp Gly Val
 305 310 315 320
 Ser Leu Ala Ser Pro Ser Asn Gln Phe Gln Leu Gln Leu Ser Gln Pro
 325 330 335
 Ile Phe Lys Gln Ser Asp Leu Arg Tyr Pro Val Leu Glu Pro Leu Leu
 340 345 350

 Pro His Leu Gly Asn Ile Leu Pro Val Ser Leu Ala Cys Asp Leu Ile
 355 360 365
 Asp Leu Tyr Phe Ser Ser Ser Ser Ala Gln Met His Pro Met Ser
 370 375 380
 Pro Tyr Val Leu Gly Phe Val Phe Arg Lys Arg Ser Phe Leu His Pro
 385 390 395 400
 Thr Asn Pro Arg Arg Cys Gln Pro Ala Leu Leu Ala Ser Met Leu Trp
 405 410 415
 Val Ala Ala Gln Thr Ser Glu Ala Ser Phe Leu Thr Ser Leu Pro Ser

 420 425 430
 Ala Arg Ser Lys Val Cys Gln Lys Leu Leu Glu Leu Thr Val Gly Leu
 435 440 445
 Leu Gln Pro Leu Ile His Thr Gly Thr Asn Ser Pro Ser Pro Lys Thr
 450 455 460
 Ser Pro Val Val Gly Ala Ala Ala Leu Gly Val Leu Gly Val Ala Met
 465 470 475 480
 Pro Gly Ser Leu Asn Met Asp Ser Leu Ala Gly Glu Thr Gly Ala Phe
 485 490 495

 Gly Ala Ile Gly Ser Leu Asp Asp Val Ile Thr Tyr Val His Leu Ala
 500 505 510
 Thr Val Val Ser Ala Ser Glu Tyr Lys Gly Ala Ser Leu Arg Trp Trp
 515 520 525
 Gly Ala Ala Trp Ser Leu Ala Arg Glu Leu Lys Leu Gly Arg Glu Leu
 530 535 540
 Pro Pro Gly Asn Pro Pro Ala Asn Gln Glu Asp Gly Glu Gly Leu Ser
 545 550 555 560

Glu Asp Val Asp Glu His Asp Leu Asn Arg Asn Asn Thr Arg Phe Val

565 570 575

Thr Glu Glu Glu Arg Glu Arg Arg Ala Trp Trp Leu Val Tyr

580 585 590

Ile Val Asp Arg His Leu Ala Leu Cys Tyr Asn Arg Pro Leu Phe Leu

595 600 605

Leu Asp Ser Glu Cys Ser Asp Leu Tyr His Pro Met Asp Asp Ile Lys

610 615 620

Trp Gln Ala Gly Lys Phe Arg Ser His Asp Ala Gly Asn Ser Ser Ile

625 630 635 640

Asn Ile Asp Ser Ser Met Thr Asp Glu Phe Gly Asp Ser Pro Arg Ala

645 650 655

Ala Arg Gly Ala His Tyr Glu Cys Arg Gly Arg Ser Ile Phe Gly Tyr

660 665 670

Phe Leu Ser Leu Met Thr Ile Leu Gly Glu Ile Val Asp Val His His

675 680 685

Ala Lys Ser His Pro Arg Phe Gly Val Gly Phe Arg Ser Ala Arg Asp

690 695 700

Trp Asp Glu Gln Val Ala Glu Ile Thr Arg His Leu Asp Met Tyr Glu

705 710 715 720

Glu Ser Leu Lys Arg Phe Val Ala Lys His Leu Pro Leu Ser Ser Lys

725 730 735

Asp Lys Glu Gln His Glu Met His Asp Ser Gly Ala Val Thr Asp Met

740 745 750

Gln Ser Pro Leu Ser Val Arg Thr Asn Ala Ser Ser Arg Met Thr Glu

755 760 765

Ser Glu Ile Gln Ala Ser Ile Val Val Ala Tyr Ser Thr His Val Met

770 775 780

His Val Leu His Ile Leu Leu Ala Asp Lys Trp Asp Pro Ile Asn Leu

785 790 795 800

Leu Asp Asp Asp Asp Leu Trp Ile Ser Ser Glu Gly Phe Val Thr Ala

805	810	815	
Thr Ser His Ala Val Ser Ala Ala Glu Ala Ile Ser Gln Ile Leu Glu			
820	825	830	
Phe Asp Pro Gly Leu Glu Phe Met Pro Phe Phe Tyr Gly Val Tyr Leu			
835	840	845	
Leu Gln Gly Ser Phe Leu Leu Leu Ile Ala Asp Lys Leu Gln Ala			
850	855	860	
Glu Ala Ser Pro Ser Val Ile Lys Ala Cys Glu Thr Ile Val Arg Ala			
865	870	875	880
His Glu Ala Cys Val Val Thr Leu Ser Thr Glu Tyr Gln Arg Asn Phe			
885	890	895	
Ser Lys Val Met Arg Ser Ala Leu Ala Leu Ile Arg Gly Arg Val Pro			
900	905	910	
Glu Asp Leu Ala Glu Gln Gln Arg Arg Arg Glu Leu Leu Ala Leu			
915	920	925	
Tyr Arg Trp Thr Gly Asn Gly Thr Gly Leu Ala Leu			
930	935	940	
<210> 49			
<211> 65			
<212> DNA			
<213> Artificial Sequence			
<220><223> primer			
<400> 49			
gtaacgccag gttttccca gtcacgacgg tttaaactcc atacgcagca aacatggct	60		
tgggc	65		
<210> 50			
<211> 64			
<212> DNA			
<213> Artificial Sequence			
<220><223> primer			
<400> 50			
gtacgagtagac tagtgtgaa gattccgtca agcttggcgc gaatgaagga ggtatgtgtga	60		

gagg	64
<210> 51	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 51	
cacacatcct ctttattcc gcccaagctt gacggaatct tcacacctag tactcgac	59
<210> 52	
<211> 57	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 52	
tgacatttt ttttttcca acacagcatg ctttagtccga cgccttcgag tccagcc	57
<210> 53	
<211> 54	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 53	
ctggactcga aggcgtcgga ctaagcatgc tgtgttgaa caacaaaaaa tgtc	54
<210> 54	
<211> 54	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 54	
gcagagcgc agtagtcgat gctattaatt aagtaggtt tgcgagcaac attg	54
<210> 55	
<211> 58	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	

<400> 55
 ctcagccctc ctcagccctca tcagccgcgg ccgctgaatc ggcaagggtt agtactag 58

<210> 56
 <211>
 > 62
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 56
 gcggataaca attcacaca ggaaacacg tttaaaccac atgccaggt tcgatgcga 60
 ag 62
 <210> 57
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 57
 gtacctcagc gctgtcgata gctgcacgca ctggcgcat gcccacgtgc agtgcac 57
 <210> 58
 <211> 59
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 58
 gtgcactgca cgtggcata cgccgcgtgc gtgcagctat cgacagcgct gaggtactc 59
 <210> 59
 <211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 59
 gcggcgcttc cgctgtcgta actatgtgc gctactcccc cgtttac 48
 <210> 60
 <211> 58

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 60

gtaagacggg ggagtagcgc agcatagttt cgacagcgga agcgccgcct tataagtg 58

<210> 61

<211>

> 53

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 61

ggcggcgctt ccgctgtcgt aactatggc tcagcagctc cggcccgagg 53

<210> 62

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 62

gccctggcc ggagctgctg agcccatagt tacgacagcg gaagcgccgc 58

<210> 63

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 63

ggcggcgctt ccgctgtcgt aactatggcc acagcgcccg cggcagcagc 53

<210> 64

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 64

cagctgtgc cgcggccgt gtggccatag ttacgacagc ggaagcgccg 59

<210>	65	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	65	
gctggaagct gctgagcaga tc		22
<210>	66	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	66	
gtgccagcat tccccagact cg		22
<210>	67	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	67	
ttaggcgacc tcttttcca		20
<210>	68	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	68	
gccgctcagg catacgagcg ac		22
<210>	69	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	69	

ctctggtcgg cctgccgttg	20
<210> 70	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 70	
tgagtatagc ggctgacttg tcg	23
<210> 71	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 71	
tctagtagt acgagacta ggtgtgaaga ttccgtcatt tcctcgacat gcgaatgcg	59
<210> 72	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 72	
tgcctatgaa accccgcatt cgcatgtcga ggaaatgacg gaatcttcac acctagtag	59
<210> 73	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 73	
tgcagctaca gagccctggg ccggagctgc tgagccata gttacgacag cgaaagcgc	59
<210> 74	
<211> 57	
<212> DNA	
<213> Artificial Sequence	

<220><223> primer

<400> 74

atagcactta taaggcggcg cttccgctgt cgtaactatg ggctcagcag ctccggc 57

<210> 75

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 75

taaacaaata ggggtccgc gcacatttc ccgaaaagtgc accctggat agactagcat 60

ctgagccatt gcagc 75

<210> 76

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 76

agtggcacgg agtcgggtgt gctttttt ctatcgagag cattggtcag tggtggcaag 60

60

<210> 77

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 77

accaatatac aaaacatgtc gtccgagcca gtgcctgccat tttcctcgac atgcgaatgc 60

60

<210> 78

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 78

gttgcgcatgc aaacccgca ttgcgcatgc gagggaaatgg cagggactgg ctcggacgac	60
	60
<210> 79	
<211> 60	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 79	
agctacagag ccctggccg gagctgctga gcccattgtt gaattctggc gggtagctg	60
	60
<210> 80	
<211> 60	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 80	
cttttacact ttcaacagc taccggcca gaattcaaca atgggctcag cagctccggc	60
	60
<210> 81	
<211> 81	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 81	
tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg gtgttttt ttctatcag	60
	60
atgttctgga tggtgagag g	81
<210> 82	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 82	
agtctatcgc agccttgct tagctaatgt tt	32

<210> 83
<211> 32
<212> DNA
<213> Artificial Sequence
<220><223> primer
<400> 83
tctaaaacat tagctaaggc aaggctgcga ta 32
<210> 84
<211> 32
<212> DNA
<213> Artificial Sequence

<220><223> primer
<400> 84
agtctatcg cagagtgcgc tcttccgggt tt 32
<210> 85
<211> 32
<212> DNA
<213> Artificial Sequence
<220><223> primer
<400> 85
tctaaaaccc ggaagacgcg actctgccga ta 32
<210> 86
<211> 33
<212> DNA
<213> Artificial Sequence
<220><223> primer
<400> 86
agtctatcga atgagtgttag gtacgagtag ttt 33
<210> 87
<211>
> 33
<212> DNA
<213> Artificial Sequence
<220><223> primer

<400>	87	
tctaaaacta ctcgtaccta cactcattcg ata		33
<210>	88	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	88	
agtctatcg cgccaatgc ttcctaattt tt		32
<210>	89	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	89	
tctaaaacat taggaagcta ttgcggccga ta		32
<210>	90	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	90	
agtctatcgc agcgcaatca gtgcagtggt tt		32
<210>	91	
<211>	32	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	91	
tctaaaacca ctgcactgat tgcgctgcga ta		32
<210>	92	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	

<220><223>	primer	
<400>	92	
tggagagact cggagaggat agg		23
<210>	93	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	93	
agcgtggagg cagtggagt gg		22
<210>	94	
<211>	24	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	94	
tggacaaagc ctgggtcctg ctcc		24
<210>	95	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	95	
atcctgactc gtccctgtgtc gg		22
<210>	96	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	primer	
<400>	96	
agtgcgtcggt ttagtggact tg		22
<210>	97	
<211>	22	

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 97

ctcgtagct gcttgaatat ag 22

<210> 98

<211> 45

<212> PRT

<213> Artificial Sequence

<220><223> EL form N-terminal extension

<400> 98

Met Ala Thr Ala Ala Ala Ala Ala Gly Gly Ala Ala Val Ala Ala

1 5 10 15

Gly Ala Asp Thr Gly Ala Ala Gly Ser Ser Ser Thr Gly Pro Pro Gly

20 25 30

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

35 40 45

<210> 99

<211> 1857

<212> DNA

<213> Trichoderma reesei

<400> 99

atgctgcgct actccccgt cttacacctg gatactctct cttgccacc actgaccaat 60

gctttcccc gcccaaagtg cgagtacctc agcgctgtcg atagctgcac gcactgccgc 120

gatgccacg tgcagtgcac tttcgaccctg cccctggcgc gacgcggccc caaagcgagg 180

aagaagagcg accagcccg ccagccgcct cctgatccga gctcgctctc caccgcggct 240

cgaccggcc agatgccgcc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag 300

cccttcgcct cgtcgctcgct gtcgcccgc acggcctggg agcccgtcga gccgctcagc 360

attgacaacg gcctgccccg gcagccgctg ggcgacactgc ccggctctc caccatccag 420

aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacacg 480

acgctagagc gcgctcgaa gcgatgtatc gacctttct tcgactacct ctacccctc 540

accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc 600

ttgcctggcg tcaaccaacc atcggcgctg tcacagctca cgccagaccc gaccaccggc	660
accacccccc tcaacgctgc cgagtctgg gccggcttg gccagccag cggctcgca	720
accgtcgca gcaggctggc tccctggcc gactgacct tcaccctggt cacggcgctc	780
tgcgccaggagg cagcattcat gctacccaag gacatccc ccgaaggaga atccgtctc	840
gagatcttgc tcgaaggcctc tcgggactgc ctgcaccaggc acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctg	960
gggaagccca agtactcgat gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1020
cagtcgcacg aggaggctgc cctcgagggg ctctccca tcgaggcaga gttccgcgt	1080
cgctgcgttt ggatcctgtt cttggcgcac aagtcaagccg ctataactcaa caatcgcc	1140
atcaccatcc acaagttactg cttcgacgccc ggcacatccca cgctataacc gtcgggtatc	1200
gaggacgagt tcctgagcac ggcgtccgag ccgccccgga agagcttcat atccggctc	1260
aacgcaaatg tgcggctctg gcagtcgcgc gctgatttgc tgctggaaat ccgcgtctg	1320
caagatcaga tcatgcacca ctttcgaggg accatgcccc cgaaccatgt gctgccctcc	1380
ggccacaggc agcatctcgat ttctcttat gtccgcttca tcacctgctt ggacgatctc	1440
ccggcgtacc tccagtcgt cacttcggcg atggcagcga tggcagaagg caacgggtct	1500
ggcgagtcca agcagttacgt gatacgtgc atcaacctgc aggtgacgtt tcactgtctg	1560
cgcatggtaa ttacgcagaa attcgaagac ctctttattt ttgctctgg cgttgagcag	1620
gctgatctca gaaagtccgaa gattgtgcga gacatgctga gggtgatgaa cgaggcgccc	1680
ttttggggcc tgcaggccaa tggcagccaa aacgttgaaa agattcgct tatcgagact	1740
agtttgctgg ccatcatcca tcgcaaccag gattcacct tggctacgcg agccaggagc	1800
gactttccg tgcttttggatatttcacg cggtggact cgaaggcgctc ggactaa	1857
<210> 100	
<211> 2070	
<212> DNA	
<213> Trichoderma reesei	
<400> 100	
atgggcttag cagctccggc ccagggtct gtagctgcag ctgcaggcgcc ccctccagct	60
gctggcgctg ggcgtggcgc tgcacgcctt ctcaccaccc cgcccgagtc tgccctggcc	120
tcgcagcccg gctcgccaaac cgcctcaacc acgcccggcga agaactcaact cgtgtcggt	180
gcaacccttgtt tccaccacca tcccaaggcgc cgtctggta gcagagcctg cgaccgctgc	240
cgccggcgca aggccaaatgc cgagtacctc agcgctgtcg atagctgcac gcactgccc	300

gatccccacg tgcaagtgcac ttgcacccgt cccctggcgc gacgcggccc caaagcgagg	360
aagaagagcg accagcccg ccagccgcct cctgatccga gctcgctctc caccgcggct	420
cgaccggcc agatgccgcc gccgctgacc ttctccggcc ccgcagtagc cgccgtgcag	480
cccttcgcct cgtcgctcgct gtcgcccac gcggccctggg agcccgtcga gccgcctcag	540
attgacaacg gcctgccccg gcagccgctg ggccgacctgc ccggccctctc caccatccag	600
aacatctcga cgccgcagcg atggatacac ctggccaacg ccatgacgct gcgcacacg	660
acgctagagc gcgtctcgaa gcgtatgtatc gaccccttct tcgactaccc ctacccctc	720
accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccgccc	780
ttgcctggcg tcaaccaacc atcgccgcgt tcacagctca cgccagaccc gaccaccggc	840
accacccccc tcaacgctgc cgagtcgtgg gccggctttg gccagcccgag cgctcgca	900
accgtggca gcaggctggc tccctggcc gactcgacct tcaccctggt cacggccgtc	960
tgcgcagagg cagcattcat gctacccaag gacatttcc ccgaaggaga atccgtctct	1020
gagatcttgc tcaagccctc tcggactgc ctgcaccagg acctcgaggc cgacctggag	1080
aatccgacgg ccaactcgat tgccattcg tacttccact ccaactgcct ccacgcgtcg	1140
gggaagccca agtactcgat gcacatattt ggccgaggcca tccgcctggc gcaggtcatg	1200
cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccgcgt	1260
cgctgtttt ggatcctgta cttggcgac aagttagccg ctataactcaa caatcgcccc	1320
atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgctataaccc gtcgggtatc	1380
gaggacgact tcccgaccc ggcgtccgag ccggcccgga agagcttcat atccggcttc	1440
aacgcaaatg tgccgctctg gcagtcggcg gctgatttgc tgctggaaat ccgcgtgtc	1500
caagatcaga tggatgcagca ctttcgaggaccatggcccgaaaccatgt gtcgcctcc	1560
ggccgacaggc agcatctcgat ttctctctat gtccgcgttca tcacctgtttt ggacgtatctc	1620
ccggcgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680
ggccgacttca agcagtagctg gatacgtgc atcaacctgc aggtgacgtt tcactgtctg	1740
cgcatggtaa ttacgcgaaatttcaagac ctcttttattttgtccctgg cggtgagcag	1800
gctgatctca gaaagtggaa gattgtgcga gacatgtga gggtgatgaa cgaggccccc	1860
ttttggggcc tgcaggccaa tggcgagccaa aacgttggaa agattccgcct tatcgagct	1920
agtttgctgg ccatcatcca tcgcaaccag gattcacccct tggctacgcg agccaggagc	1980
gactttccg tgctttggatatttcacgcg ccgtggact cgaaggcgctc ggaccaactg	2040
aggaataacgt ccactaccgt tggtggctaa	2070

<211> 2172

<212> DNA

<213> Trichoderma reesei

<400> 101

atggccacag	cgccgcggc	agcagctggc	ggcgccgg	ttgtgcgg	tgcagacaca	60
ggcgctgcag	gctccagctc	tacaggccct	ccaggccctc	cagggcttcc	aggcacccgg	120
acaggctccg	tggcgatggg	ctcagcagct	ccggcccagg	gctctgttagc	tgcagctgca	180
ggcgccctc	cagctgctgg	cgctggcgct	ggcgctgtcc	acgcctcac	cacctcgccc	240
gagtcgtgcct	cggctcgca	gcccggctcg	ccaaccgcct	caaccacgcc	gccgcagaac	300
tcactcggt	cggctgcaac	ctcggttccac	caccatccca	gaggccgtct	ggtgagcaga	360
gcctgctgacc	gctggcgccg	gctcaaggcc	aagtgcgagt	acctcagcgc	tgtcgatagc	420
tgcacgcact	gccgcgatgc	ccacgtgcag	tgcactttcg	acctgcctct	ggcgcgacgc	480
ggccccaag	cgaggaagaa	gagcgcaccag	cccgcccaggc	cgccctctga	tccgagctcg	540
ctctccaccg	cggctcgacc	cggccagatg	ccggcccccgc	tgaccttctc	cggccccgca	600
gtagccgcgc	tgcagccctt	cgcctcgctg	tcgctgtcgc	ccgacgcggc	ctgggagccc	660
gtcgagccgc	tcagcattga	caacggctg	ccccggcagc	cgctggcga	cctgcccggc	720
ctctccacca	tccagaacat	ctcgacgcgc	cagcgatgga	tacacctggc	caacgcctatg	780
acgctgcgca	acacgacgct	agagcgcgtc	tcgaagcgat	gtatcgacct	cttcttcgac	840
tacctctacc	ccctcacccc	cctgggtgtac	gagccggccc	tccggacgt	gctcgatatac	900
atcttctccc	agcccttgcc	tggcgtcaac	caaccatcgc	cgctgtcaca	gctcaegcca	960
gaccgcacca	ccggcaccac	ccccctcaac	gctgcccagg	cgtggccgg	ctttggccag	1020
cccgagggct	cgcgaaccgt	cggcagcagg	ctggctccct	ggccgcactc	gaccttcacc	1080
ctggtcacgg	ccgtctgcgc	agaggcagca	ttcatgctac	ccaaggacat	tttccccgaa	1140
ggagaatccg	tctctgagat	cttgctcgaa	gcctctcggt	actgcctgca	ccagcacctc	1200
gaggccgacc	tggagaatcc	gacggccaac	tcgattgcca	ttcgctactt	ccactccaac	1260
tcgcctccacg	ctgcggggaa	gcccaagtac	tcgtggcaca	tatttggcga	ggccatccgc	1320
ctggcgcagg	tcatgcagct	gcacgaggag	gctgcctcg	aggggctcg	ccccatcgag	1380
gcagagttcc	gccgtcgctg	ctttggatc	ctgtacttgg	gcgacaagtc	agccgctata	1440
ctcaacaatc	ggcccatcac	catccacaag	tactgcttcg	acgcggcat	caccacgcta	1500
tacccgtcg	gtatcgagga	cgagttcctg	agcacggcgt	ccgagccgccc	ccggaagagc	1560
ttcatatccg	gctcaacgc	aatgtgcgg	ctctggcagt	ccgcggctga	tttgctgctg	1620

gaaatcccg	tgctgcaaga	tcatatgtat	cagcaatttc	gagggaccat	gccccgaac	1680
catgtgtgc	cctccgcccga	caggcagcat	ctcgattctc	tctatgtccg	cttcatcacc	1740
tgcttggacg	atctccgccc	gtacctccag	tcgtcactc	tggcgatggc	agcgatggca	1800
gaaggcaacg	ggtctgccga	gtccaaggcag	tacgtatac	agtgcataa	cctgcagggt	1860

acgtttact	gtctgcgc	ggttaattacg	cagaattcg	aagacctctc	ttat	ttgt	1920	
cctggcg	ttg	agcaggctga	tctcagaaag	tcggagat	tg	cgagacat	gctgagggt	1980
atgaacgagg	cgc	cccttttgc	gggcctgc	gccaatggcg	agccaaacgt	tgaaaagatt	2040	
cgccttata	cg	cg	gagctagttt	gctggccatc	atccatcgca	accaggattc	acccttgct	2100
acgcgagcca	ggagcgactt	ttccgtgtt	ttggatattc	tcacgcggct	ggactcgaag	2160		
cggtcggact	aa						2172	

<210> 102

<211> 2037

<212> DNA

<213> Trichoderma reesei

<400> 102

atgggctcag	cagtcggc	ccagggct	ctgtacgtc	ctgcaggc	ccctcc	cagct	60
gctggcg	ctg	gctggc	tgtccacgc	ctcaccac	cgc	ccgg	120
tcgc	agcc	ccgc	ccac	acgc	ccgc	agaact	180
gcaac	ctcg	tcc	accac	tcc	cg	actc	240
cgccgg	cgca	aggc	caatgt	cgat	ac	gtgtcg	300
gatgcccac	g	tg	ca	tt	cc	tcac	360

aagaagagcg	accagcccg	ccagccgc	cctgatccga	gctcg	ctc	cacc	cggt	420
cgacccggcc	agat	gccc	gccc	gtgacc	ttct	cccg	cgat	480
ccctcg	cgt	cg	cc	ccg	cc	ccg	ctcg	540
attgacaac	gc	c	cc	cccg	cc	cc	ctc	600
aacatctcga	cgc	cc	cc	cc	cc	cc	cc	660
acgttagagc	gc	gt	tc	cc	cc	cc	cc	720
acccccc	tgt	ac	cc	cc	cc	cc	cc	780

ttgcctggcg	tcaaccaacc	atcgccg	ct	ac	ag	ct	ca	cc	gg	cc	840
accacccccc	tcaac	cg	tc	at	gg	cc	cc	cc	gg	cc	900
accgtcg	cg	cc	cc	cc	cc	cc	cc	cc	cc	cc	960

tgccgcagg cagcattcat gctacccaag gacat tttcc ccgaaggaga atccgtctct	1020
gagatcttgc tcgaagcctc tcggactgc ctgcaccagc acctcgaggc cgacctggag	1080
aatccgacgg ccaactcgat tgccattgc tacttccact ccaactgcct ccacgctgcg	1140
gggaagccca agtactcgta gcacatattt ggcgaggcca tccgcgttgc gcaggatcatg	1200
cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccgcgt	1260
cgctgcttt ggatcctgta ctggcgac aagttagccg ctataactcaa caatcgcccc	1320
atcaccatcc acaagtactg cttcgacgcc ggcacatcacca cgctatacccg tcgggtatc	1380
gaggacgagt tcctgagcac ggcgtccgag ccgcggcgga agagcttcat atccggcttc	1440
aacgcaaatg tgccgtctg gcagtcgcg gctgatttc tgctggaaat ccgcgtgtcg	1500
caagatcaga tgcgtcagca cttcgaggg accatgcccc cgaaccatgt gtcgcctcc	1560
ccgcacaggc agcatctcgat ttctcttat gtccgcgttca tcacctgctt ggacgatctc	1620
ccgcccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680
gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcactgtctg	1740
cgcatggtaa ttacgcagaa attcgaagac ctctttatt ttgctcctgg cggtgagcag	1800
gctgatctca gaaagtggaa gattgtgcga gacatgctga gggtgatgaa cgaggcgccc	1860
ttttggggcc tgcaggccaa tggcgagcga aacgttggaa agattgcct tatcgagct	1920
agtttgctgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1980
gactttccg tgctttggaa tattctcagc cggctggact cgaaggcgtc ggactaa	2037