



- (19) 국제특허분류(Int. Cl.)
C12N 9/42 (2006.01) *C12N 15/56* (2006.01)
C12P 19/00 (2006.01)

(21) 출원번호 10-2013-7027127

(22) 출원일자(국제) 2012년03월16일
 심사청구일자 없음

(85) 번역문제출일자 2013년10월15일

(86) 국제출원번호 PCT/US2012/029498

(87) 국제공개번호 WO 2012/125951
 국제공개일자 2012년09월20일

(30) 우선권주장
 61/453,918 2011년03월17일 미국(US)

(71) 출원인
 다니스코 유에스 인크.
 미합중국 캘리포니아 (우편번호 94304) 팔로 알토
 페이지 밀 로드 925

(72) 발명자
 카페르 티스
 미국 94304 캘리포니아주 팔로 알토 페이지 밀 로
 드 925 다니스코 유에스 인크.
 니콜라에프 이고르
 미국 94304 캘리포니아주 팔로 알토 페이지 밀 로
 드 925 다니스코 유에스 인크.
 (뒷면에 계속)

(74) 대리인
 특허법인코리아나

전체 청구항 수 : 총 32 항

- (54) 발명의 명칭 **셀룰라제 조성물 및 리그노셀룰로스를 바이오매스의 발효성 당으로의 전환을 향상시키기 위해 상기 조성물을 사용하는 방법**

(57) 요약

본 발명은 β -글루코시다제 활성을 갖는 폴리펩티드를 포함하는 조성물과 같은 바이오매스를 가수분해하는데 사용될 수 있는 조성물, 바이오매스 물질을 가수분해하는 방법, 및 그러한 β -글루코시다제 폴리펩티드 및/또는 활성을 포함하는 조성물의 안정성 및 당화 효능을 향상시키는 방법에 관한 것이다.

대표도 - 도1a

[illegible]

(72) 발명자

랜츠 수잔

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로드 925 다니스코 유에스 인크.

푸즈달라 메러디스 케이

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로드 925 다니스코 유에스 인크.

시 메건 와이

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로드 925 다니스코 유에스 인크.

특허청구의 범위

청구항 1

- a) 서열 번호 135에 대하여 적어도 약 70% 동일성을 갖는 아미노산 서열; 또는
- b) N-말단 서열 및 C-말단 서열 - 여기서, N-말단 서열은 제1 β -글루코시다제로부터 유래되는 제1 아미노산 서열을 포함하고, 길이가 적어도 200개의 잔기로 되어 있으며, 서열 번호 164 내지 169 중 하나 이상 또는 모두를 포함하고, C-말단 서열은 제2 β -글루코시다제로부터 유래되는 제2 아미노산 서열을 포함하며, 길이가 적어도 50개의 잔기로 되어 있고, 서열 번호 170을 포함함 - 을 포함하며,
- β -글루코시다제 활성을 갖는 단리된 폴리펩티드.

청구항 2

제1항에 있어서, 서열 번호 135에 대하여 적어도 약 80% 동일성을 갖는 아미노산 서열을 포함하는 단리된 폴리펩티드.

청구항 3

제1항 또는 제2항에 있어서, 서열 번호 135에 대하여 적어도 약 90% 동일성을 갖는 아미노산 서열을 포함하는 단리된 폴리펩티드.

청구항 4

제1항에 있어서, 제1 β -글루코시다제로부터 유래되는 N-말단 서열 및 제2 β -글루코시다제로부터 유래되는 C-말단 서열을 포함하며, 여기서 제1 β -글루코시다제 및 제2 β -글루코시다제는 서로 상이한 단리된 폴리펩티드.

청구항 5

제1항 또는 제4항에 있어서, N-말단 서열 및 C-말단 서열은 직접적으로 연결되지 않으나, 링커 도메인을 통해 기능적으로 연결되는 단리된 폴리펩티드.

청구항 6

제5항에 있어서, N-말단 서열, C-말단 서열, 또는 링커 도메인은 서열 번호 171 또는 172의 아미노산 서열을 포함하는 길이가 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 영역 서열을 포함하는 단리된 폴리펩티드.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 제1 β -글루코시다제 또는 제2 β -글루코시다제에 비하여 향상된 안정성을 갖는 단리된 폴리펩티드.

청구항 8

제7항에 있어서, 향상된 안정성은 저장 조건 또는 생성 조건 하에서의 단백질 가수분해에 의한 절단 저항성 증가인 단리된 폴리펩티드.

청구항 9

제4항 내지 제8항 중 어느 한 항에 있어서, N-말단 서열은 동일한 길이의 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 또는 79의 서열에 대하여 적어도 90% 서열 동일성을 갖는 아미노산 서열을 포함하고, C-말단 서열은 서열 번호 170의 서열 모티프를 포함하는 단리된 폴리펩티드.

청구항 10

제4항 내지 제8항 중 어느 한 항에 있어서, N-말단 서열은 서열 번호 164 내지 169의 서열 모티프 중 하나 이상 또는 모두를 포함하며, C-말단 서열은 동일한 길이의 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74,

76, 78 또는 79의 서열에 대하여 적어도 90% 서열 동일성을 갖는 아미노산 서열을 포함하는 단리된 폴리펩티드.

청구항 11

제9항 또는 제10항에 있어서, N-말단 서열은 서열 번호 136 내지 148 중 3개 이상, 4개 이상, 5개 이상의 서열 모티프에 이어지며, C-말단 서열은 서열 번호 149 내지 156 중 2개 이상, 3개 이상, 또는 4개 이상의 서열 모티프에 이어지는 단리된 폴리펩티드.

청구항 12

제1항 내지 제11항 중 어느 한 항의 단리된 폴리펩티드를 포함하는 조성물.

청구항 13

제12항에 있어서, 하나 이상의 셀룰라제를 추가로 포함하는 조성물.

청구항 14

제13항에 있어서, 하나 이상의 셀룰라제는 엔도글루카나제, GH61/엔도글루카나제, 셀로비오하이드롤라제 및 다른 베타-글루코시다제로부터 선택되는 조성물.

청구항 15

제12항 내지 제14항 중 어느 한 항에 있어서, 하나 이상의 헤미셀룰라제를 추가로 포함하는 조성물.

청구항 16

제15항에 있어서, 하나 이상의 헤미셀룰라제는 자일라나제, β -자일로시다제, 또는 L- α -아라비노푸라노시다제로부터 선택되는 조성물.

청구항 17

제12항 내지 제16항 중 어느 한 항에 있어서, β -글루코시다제는 조성물 중의 단백질의 총량에 대하여 1 wt.% 내지 75 wt.%의 양으로 존재하는 조성물.

청구항 18

제12항 내지 제17항 중 어느 한 항에 있어서, 배양 혼합물 또는 발효 브로쓰(fermentation broth)인 조성물.

청구항 19

제18항에 있어서, 전체 브로쓰 제제(whole broth formulation)인 조성물.

청구항 20

- a) 서열 번호 83에 대하여 적어도 70% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하거나;
- b) 높은 엄격성 조건 하에서 서열 번호 83 또는 그의 상보체와 혼성화할 수 있는 뉴클레오티드 서열을 포함하거나;
- c) 서열 번호 135에 대하여 적어도 약 70% 동일성을 갖는 아미노산 서열을 포함하는, β -글루코시다제 활성을 갖는 단리된 폴리펩티드, 또는 N-말단 서열 및 C-말단 서열 - 여기서, N-말단 서열은 제1 β -글루코시다제로부터 유래되는 제1 아미노산 서열을 포함하고, 길이가 적어도 200개의 잔기로 되어 있으며, 서열 번호 164 내지 169 중 하나 이상 또는 모두를 포함하고, C-말단 서열은 제2 β -글루코시다제로부터 유래되는 제2 아미노산 서열을 포함하며, 길이가 적어도 50개의 잔기로 되어 있고, 서열 번호 170을 포함함 - 을 포함하는, β -글루코시다제 활성을 갖는 단리된 폴리펩티드를 암호화하는 단리된 폴리뉴클레오티드.

청구항 21

제20항에 있어서, 서열 번호 83에 대하여 적어도 90% 동일성을 갖는 뉴클레오티드 서열을 포함하는 단리된 폴리뉴클레오티드.

청구항 22

제20항 또는 제21항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 23

제20항 또는 제21항의 폴리뉴클레오티드를 발현하도록 조작된 재조합 숙주 세포.

청구항 24

제23항에 있어서, 박테리아 또는 진균 세포인 재조합 숙주 세포.

청구항 25

제24항에 있어서, 바실루스(*Bacillus*) 또는 에스케리키아 콜라이(*E. coli*)로부터 선택되는 재조합 숙주 세포.

청구항 26

제24항에 있어서, 트리코데르마(*Trichoderma*), 아스페르길루스(*Aspergillus*), 크리소스포리움(*Chrysosporium*), 또는 효모 세포로부터 선택되는 재조합 숙주 세포.

청구항 27

제23항 내지 제26항 중 어느 한 항의 재조합 숙주 세포를 발효시켜 제조된 발효 브로쓰 또는 배양 혼합 조성물.

청구항 28

셀룰로스에 바이오매스 물질을 제1항 내지 제11항 중 어느 한 항의 폴리펩티드, 제12항 내지 제19항 중 어느 한 항의 조성물, 또는 제27항의 발효 브로쓰 또는 배양 혼합 조성물과 접촉시키는 것을 포함하는, 셀룰로스에 바이오매스 물질을 가수분해하는 방법.

청구항 29

제28항에 있어서, 바이오매스 물질은 종자, 낱알, 덩이줄기, 식품 가공 또는 생산 가공의 식물 폐기물 또는 부산물, 줄기, 옥수수 속대, 옥수수 대, 잎, 목초, 다년생 줄기, 목재, 종이, 펄프 및 재생지, 감자, 대두, 보리, 호밀, 귀리, 밀, 비트(beet) 및 사탕수수 바가스(sugar cane bagasse)로부터 선택되는 방법.

청구항 30

제28항 또는 제29항에 있어서, 바이오매스 물질을 전처리하는 방법.

청구항 31

제30항에 있어서, 전처리는 산 전처리 또는 염기 전처리, 또는 산 전처리 및 염기 전처리의 조합을 포함하는 방법.

청구항 32

상업적 환경(commercial setting) 또는 산업 환경에서의 제1항 내지 제11항 중 어느 한 항의 폴리펩티드, 제12항 내지 제19항 중 어느 한 항의 조성물, 또는 제27항의 발효 브로쓰 또는 배양 혼합 조성물을 적용하는 방법, 또는 제28항 내지 제31항 중 어느 한 항의 가수분해 방법 - 여기서, 상기 방법은 머천트(merchant) 효소 공급 모델 전략 또는 온-사이트 바이오리파이너리(on-site biorefinery) 모델 전략을 취함 - .

명세서

기술 분야

관련 출원과의 상호 참조

본 출원은 본 명세서에 그 전문이 참고로 포함되는 미국 가출원 제61/453,918호(출원일: 2011년 3월 17일)에 대한 우선권을 주장한다.

[0003] 본 개시내용은 일반적으로 특정 β -글루코시다제 효소, 및 조작된 β -글루코시다제 효소 조성물, β -글루코시다제 발효 브로쓰(fermentation broth) 조성물, 및 이러한 β -글루코시다제를 포함하는 다른 조성물, 및 예를 들어, 헤미셀룰로스 및 임의로 셀룰로스를 포함하는 바이오매스 물질의 발효성 당으로의 당화 또는 전환을 위한 연구, 산업 또는 상업적 환경에서의 상기 효소 및 조성물의 사용 또는 제조 방법에 관한 것이다.

배경 기술

[0004] 액체 연료의 대체품으로서의 알코올(예를 들어, 에탄올)을 생성하도록 후속적으로 발효시키는 발효성 당으로의 재생가능한 리그노셀룰로스로부터 바이오매스의 생물전환(bioconversion)은 석유 위기가 일어났던 1970년대 이후로 연구자의 집중적인 주목을 끌어왔다(문헌[Bungay, H. R., "Energy: the biomass options". NY: Wiley; 1981]; 문헌[Olsson L, Hahn-Hagerdal B. Enzyme Microb Technol 1996,18:312-31]; 문헌[Zaldivar, J *et al.*, Appl Microbiol Biotechnol 2001, 56: 17-34]; 문헌[Galbe, M *et al.*, Appl Microbiol Biotechnol 2002, 59:618-28]). 에탄올은 지난 수십년간 미국에서는 휘발유에 대한 10% 블렌드로, 또는 브라질에서는 차량용 순수한(neat) 연료로서 사용되어 왔다. 연료용 바이오에탄올의 중요성은 유가 상승 및 이의 공급원의 점진적인 고갈과 병행하여 증가할 것이다. 게다가, 발효성 당은 플라스틱, 폴리머 및 기타 바이오 제품을 제조하는데 점점 더 사용이 증가되고 있다. 따라서, 석유계 연료 공급원료 대신에 사용될 수 있는 풍부한 저가 발효성 당에 대한 수요가 빠르게 증가하고 있다.

[0005] 유용한 재생가능한 바이오매스 물질 중 주된 것은 셀룰로스 및 헤미셀룰로스(자일란)인데, 이는 발효성 당으로 전환될 수 있다. 이들 다당류의 가용성 당, 예를 들어, 글루코스, 자일로스, 아라비노스, 갈락토스, 만노스 및/또는 다른 6탄당 및 5탄당으로의 효소에 의한 전환은 다양한 효소의 조합된 작용에 기인하여 발생한다. 예를 들어, 엔도-1,4- β -글루카나제(EG) 및 엑소-셀로비오하이드롤라제(CBH)는 불용성 셀룰로스의 셀로올리고당(예를 들어, 셀로비오스가 주요 산물임)으로의 가수분해를 촉매하는 한편, β -글루코시다제(BGL)는 올리고당을 글루코스로 전환시킨다. 다른 부속 단백질과 함께 자일라나제(헤미셀룰라제; 이의 비제한적인 예는 L- α -아라비노푸라노시다제, 페룰로일 및 아세틸자일란 에스테라제, 글루쿠로니다제 및 β -자일로시다제를 포함함)는 헤미셀룰로스의 가수분해를 촉매한다.

[0006] 식물의 세포벽은 공유 및 비공유 수단을 통해 상호작용하는 복합 다당류의 불균질 혼합물로 이루어진다. 고등 식물 세포벽의 복합 다당류는 예를 들어, 셀룰로스(β -1,4 글루칸)를 포함하며, 이는 일반적으로 세포벽 성분에서 관찰되는 탄소의 35 내지 50%를 구성한다. 셀룰로스 폴리머는 수소 결합, 반 데르 발스 상호작용 및 소수성 상호작용을 통해 자가 회합하여, 반결정질 셀룰로스 미세섬유(microfibril)를 형성한다. 이들 미세섬유는 또한 일반적으로 무정형 셀룰로스로 알려져 있는 비결정질 영역을 포함한다. 셀룰로스 미세섬유는 헤미셀룰로스(예를 들어, 자일란, 아라비난 및 만난 포함), 펙틴(예를 들어, 갈락투로난 및 갈락탄) 및 다양한 다른 β -1,3 및 β -1,4 글루칸으로 형성된 매트릭스 내에 매립된다. 이들 매트릭스 폴리머는 종종 예를 들어, 아라비노스, 갈락토스 및/또는 자일로스 잔기로 치환되어, 고도의 복합 아라비노자일란, 아라비노갈락탄, 갈락토만난 및 자일로글루칸을 제공한다. 헤미셀룰로스 매트릭스는 결국 폴리페놀계 리그닌으로 둘러싸인다.

[0007] 바이오매스 물질로부터 유용한 발효성 당을 수득하기 위하여, 리그닌은 전형적으로 투과화되고, 헤미셀룰로스는 셀룰로스-가수분해 효소에 의한 접근이 가능하도록 파괴된다. 효소 활성의 컨소시엄(consortium)은 발효성 당을 수득할 수 있기 전에, 바이오매스 물질의 복합 매트릭스를 파괴하는데 필요할 수 있다.

[0008] 셀룰로스로부터 공급원료의 유형과 상관없이, 효소의 비용 및 가수분해 효율은 바이오매스 생물전환 공정의 상용화를 제한하는 주요 인자이다. 미생물에 의해 생성되는 효소의 생성 비용은 효소 생성 균주의 생산성 및 발효 브로쓰의 최종 활성 수율과 관련이 있다. 다효소(multienzyme) 복합체의 가수분해 효율은 다수의 인자, 예를 들어, 개별 효소의 특성, 그들 간의 상승작용 및 다효소 블렌드에서의 그들의 비율에 좌우될 수 있다.

[0009] 식물 및/또는 다른 셀룰로스로부터 또는 헤미셀룰로스로부터 물질을, 충분하거나 향상된 효능, 향상된 발효성 당 수율 및/또는 매우 다양한 셀룰로스로부터 또는 헤미셀룰로스로부터 물질에서 작용하는 향상된 능력으로, 발효성 당으로 전환시킬 수 있는 효소 및/또는 효소 조성물을 동정하는 것이 당업계에 필요하다. 본 명세서에 기재된 향상된 방법 및 조성물은 저 비용으로 재생가능한 공급원료로부터 발효성 당을 수득할 수 있는 그러한 효소 조성물을 제공한다.

[0010] 특허, 특허 출원, 문헌, 뉴클레오타이드/단백질 서열 데이터베이스 수탁 번호 및 본 명세서에 인용된 논문은 본 명세서에 그 전문이 참고로 포함된다.

[0011] 발명의 요약

[0012] 본 명세서에는 변이체, 돌연변이체, 하이브리드/키메라/융합 효소를 비롯한 다수의 β -글루코시다제 폴리펩티드, 이러한 폴리펩티드를 암호화하는 핵산, 이러한 폴리펩티드를 포함하는 조성물 및 이러한 조성물을 사용하는 방법이 제공된다. 본 명세서의 조성물은 일부 태양에서, 비천연 셀룰라제 조성물이다. 상기 조성물은 하나 이상의 헤미셀룰라제를 추가로 포함할 수 있으며, 이를 테면 헤미셀룰라제 조성물이다. 일부 태양에서, 상기 조성물은 다양한 바이오매스 물질을 발효성 당으로 전환시키는 당화 공정에 사용될 수 있다. 일부 태양에서, 본 명세서의 조성물은 향상된 당화 효능 또는 효율 및 기타 이점을 제공한다. 본 명세서에는 세포, 예를 들어, 재조합에 의해 조작된(recombinantly engineered) 숙주 세포, 이들 세포로부터 유래된 발효 브로쓰, 및 이들 세포 또는 발효 브로쓰를 사용하는 방법 또는 공정도 제공된다. 또한 이러한 폴리펩티드, 이러한 폴리펩티드를 암호화하는 핵산, 및 이러한 폴리펩티드를 포함하는 조성물을 사용하는 비즈니스 방법도 본 발명에 기재되어 고려된다.

[0013] 특정 태양에서, 본 개시내용은 적어도 2개의 β -글루코시다제 서열로 된 키메라(또는 하이브리드, 또는 융합, 이들 용어는 동일한 개념을 나타내도록 본 명세서에서 교호적으로 사용됨)인 β -글루코시다제 폴리펩티드를 포함하는 비천연 셀룰라제 조성물을 제공한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함할 수 있다. 따라서, 조성물은 헤미셀룰라제 조성물일 수 있다. 비천연 셀룰라제/헤미셀룰라제 조성물은 적어도 2개의 상이한 공급원으로부터 유래되는 성분을 포함한다. 일부 태양에서, 비천연 셀룰라제/헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 포함한다. 조성물 중의 β -글루코시다제 폴리펩티드는 하나 이상의 글리코실화 부위를 추가로 포함할 수 있다. 일부 태양에서, β -글루코시다제 폴리펩티드는 N-말단 서열 및 C-말단 서열을 포함하며, N-말단 서열 또는 C-말단 서열 각각은 상이한 β -글루코시다제로부터 유래되는 하나 이상의 하위서열을 포함한다. 특정 태양에서, N-말단 및 C-말단 서열은 상이한 공급원으로부터 유래된다. 일부 실시형태에서, N-말단 및 C-말단 서열의 하나 이상의 하위서열 중 적어도 2개는 상이한 공급원으로부터 유래된다. 일부 태양에서, N-말단 서열 또는 C-말단 서열 중 어느 하나는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 영역 서열을 추가로 포함한다. 특정 실시형태에서, N-말단 서열 및 C-말단 서열은 바로 인접해 있거나, 직접 연결되어 있다. 다른 실시형태에서, N-말단 및 C-말단 서열은 바로 인접해 있지 않지만, 이들은 링커 도메인을 통하여 기능적으로 연결되어 있다. 특정 실시형태에서, 링커 도메인은 키메라 폴리펩티드의 중앙에 위치한다(예를 들어, N-말단 또는 C-말단 중 어느 하나에 위치하지 않음). 특정 실시형태에서, 하이브리드 폴리펩티드의 N-말단 서열 또는 C-말단 서열 중 어느 것도 루프 서열을 포함하지 않는다. 대신에, 링커 도메인은 루프 서열을 포함한다. 일부 태양에서, N-말단 서열은 길이가 적어도 약 200개(예를 들어, 약 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개)의 잔기로 된 β -글루코시다제 또는 그의 변이체의 제1 아미노산 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 길이가 적어도 약 50개(예를 들어, 약 50, 75, 100, 125, 150, 175, 또는 200개)의 아미노산 잔기로 된 β -글루코시다제 또는 그의 변이체의 제2 아미노산 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 일부 태양에서, C-말단 또는 N-말단 서열 중 어느 하나는 루프 서열을 포함하며, 루프 서열은 FDRSPG(서열 번호 171) 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성된다. 일부 태양에서, C-말단 또는 N-말단 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, C-말단 서열 및 N-말단 서열은 루프 서열을 포함하는 링커 도메인을 통하여 연결되며, 루프 서열은 FDRSPG(서열 번호 171) 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기를 포함한다. 특정 실시 형태에서, β -글루코시다제 폴리펩티드는 서열 번호 135에 대하여 적어도 약 65%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%)의 동일성을 갖는 서열을 포함한다. 일부 실시형태에서, β -글루코시다제 활성을 갖는 폴리펩티드(즉, β -글루코시다제 폴리펩티드)는 서열 번호 83에 대하여 적어도 약 65%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%)의 동일성을 갖는 뉴클레오티드, 또는 서열 번호 83 또는 그의 상보체와 높은 엄격성 조건 하에서 혼성화할 수 있는 폴리뉴클레오티드에 의해 암호화된다. 일부 태양에서, 비천연 셀룰라제 또는 헤미셀룰라제 조성물 중의 β -글루코시

다제 폴리펩티드(들)는 키메라 폴리펩티드의 C-말단 및/또는 N-말단 서열 각각이 유래되는 임의의 고유 효소에 비해 안정성이 향상되었다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실을 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 30% 미만, 또는 약 20% 미만, 더욱 바람직하게는 15% 미만, 또는 10% 미만이다.

[0014] 본 개시내용의 폴리펩티드는 적당히 얻어질 수 있고/있거나 "실질적으로 순수한" 형태로 사용될 수 있다. 예를 들어, 본 개시내용의 폴리펩티드는 소정 조성물 중의 총 단백질의 적어도 약 80 wt.%(예를 들어, 적어도 약 85 wt.%, 90 wt.%, 91 wt.%, 92 wt.%, 93 wt.%, 94 wt.%, 95 wt.%, 96 wt.%, 97 wt.%, 98 wt.%, 또는 99 wt.%)를 구성하며, 또한 완충제 또는 용액과 같은 기타 성분을 포함한다.

[0015] 일부 태양에서, 본 개시내용은 변이체, 돌연변이체 및 하이브리드/융합/키메라 폴리펩티드를 포함하는 β -글루코시다제 폴리펩티드를 암호화하는 핵산을 제공한다. 예를 들어, 본 개시내용은 β -글루코시다제 폴리펩티드를 암호화하는 단리된 핵산을 제공하며, 여기서 핵산은 서열 번호 83에 대하여 적어도 약 65%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%)의 동일성을 갖는 것, 또는 서열 번호 83 또는 그의 상보체와 높은 엄격성 조건 하에서 혼성화할 수 있는 것이다. 본 개시내용은 또한 이러한 핵산 분자를 포함하는 숙주 세포를 제공한다. 일부 실시형태에서, 본 개시내용은 추가로, 핵산 분자 및 숙주 세포와 함께 사용하기에 적합한 프로모터 및 벡터를 제공한다. 특정 태양에서, 본 개시내용은 셀룰라제 조성물 또는 헤미셀룰라제 조성물을 비롯한, 숙주 세포를 발효하여 제조된 조성물을 제공한다. 이와 같이, 본 개시내용은 발효 브로쓰 조성물을 제공한다.

[0016] 일부 태양에서, 본 개시내용은 바이오매스 기질/물질의 당화를 달성하기 위해 본 명세서의 폴리펩티드를 암호화하는 핵산, 세포, 폴리펩티드, 또는 조성물을 사용하는 방법을 제공한다. 특정 실시형태에서, 바이오매스 기질/물질은 적절히 전처리되거나, 적절한 전처리 방법에 처해진다. 일부 실시형태에서, 본 개시내용은 또한 본 명세서에 기재된 조성물, 폴리펩티드, 세포, 또는 핵산과 관련된 특정 상업적 또는 비즈니스 방법을 제공한다.

도면의 간단한 설명

[0017] 하기의 도면 및 표는 본 명세서의 개시내용 또는 특허청구범위의 범주와 내용을 제한하지 않고 예시하고자 한다.

<도 1>

도 1은 다양한 효소 및 이들 효소 중 일부를 암호화하는 뉴클레오타이드의 본 개시내용에 사용되는 서열 아이덴티파이어(sequence identifier)의 요약을 제공한다.

<도 2>

도 2는 -1 하위부위(subsite)(단백질 데이터뱅크 수탁 번호 pdb:2X41의 결정 구조)에서 글루코스와 복합체화된 씨모토가 네아폴리타나(*T. neapolitana*) Bgl3B의 결정 구조에 기초하여 예측되는, 특정 β -글루코시다제(예를 들어, Fv3C) 상동체 중의 보존된 잔기를 제공한다.

<도 3>

도 3은 트리코테르마 리세이(*T. reesei*) 통합 균주 H3A에 의해 생성되는 발효 브로쓰의 효소 조성물을 제공한다.

<도 4a 내지 4e>

도 4a는 실시예 2의 각 시료에 개별적으로 첨가한 효소(정제되거나 정제하지 않은) 및 이들 효소의 스톡(stock) 단백질 농도를 열거한 것이다. 도 4b는 실시예 2에 따라 트리코테르마 리세이 통합 균주 H3A에 첨가한 도 4a의 다양한 정제되거나 정제되지 않은 효소를 포함하는 효소 조성물을 첨가함으로써 희석 암모니아로 전처리된 옥수수 속대의 당화 후의 글루코스 방출량을 도시한 것이다. 도 4c는 실시예 2에 따라 트리코테르마 리세이 통합 균주 H3A에 첨가한 도 4a의 다양한 정제되거나 정제되지 않은 효소를 포함하는 효소 조성물을 첨가함으로써 희석 암모니아로 전처리된 옥수수 속대의 당화 후의 셀로비오스 방출량을 도시한 것이다. 도 4d는 실시예 2에 따라 트리코테르마 리세이 통합 균주 H3A에 첨가한 도 4a의 다양한 정제되거나 정제되지 않은 효소를 포함하는 효소 조성물을 첨가함으로써 희석 암모니아로 전처리된 옥수수 속대의 당화 후의 자일로비오스 방출량을 도시한 것이다. 도 4e는 실시예 2에 따라 트리코테르마 리세이 통합 균주 H3A에 첨가한 도 4a의 다양한 정제되거나 정

제되지 않은 효소를 포함하는 효소 조성물을 첨가함으로써 희석 암모니아로 전처리된 옥수수 속대의 당화 후의 자일로스 방출량을 도시킨 것이다.

<도 5a 및 5b>

도 5a는 트리코테르마 리세이 Bgl1(Tr3A), 아스페르길루스 니게르(*A. niger*) Bglu(An3A), Fv3C, Fv3D 및 Pa3C를 비롯한 수많은 β -글루코시다제 상동체의 β -글루코시다제 활성을 열거한다. 실시예 4에 따라, 셀로비오스 및 CNPG 기질에 대한 활성을 측정하였으며; 도 5b는 실시예 5A에 따라 셀로비오스 및 CNPG 기질에 대한, 트리코테르마 리세이 Bgl1에 대한 다른 그룹의 β -글루코시다제 상동체의 활성을 비교한 것이다.

<도 6>

도 6은 실시예 5B 내지 5D에서 시험한 효소 혼합물/조성물 중의 효소의 상대 중량을 열거한 것이다.

<도 7>

도 7은 희석 암모니아로 전처리된 옥수수 속대에 대한 효소 조성물의 영향의 비교를 제공한다.

<도 8a 및 8b>

도 8a는 Fv3A 뉴클레오티드 서열(서열 번호 1)을 도시킨 것이다. 도 8b는 Fv3A 아미노산 서열(서열 번호 2)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 9a 및 9b>

도 9a는 Pf43A 뉴클레오티드 서열(서열 번호 3)을 도시킨 것이다. 도 9b는 Pf43A 아미노산 서열(서열 번호 4)을 도시킨 것이다. 예측된 신호 서열에 밑줄이 그어져 있고, 예측된 보존 도메인은 볼드체로 되어 있으며, 예측된 탄수화물 결합 모듈("CBM")은 대문자로 되어 있고, CD와 CBM을 분리하는 예측된 링커는 이탤릭체로 되어 있다.

<도 10a 및 10b>

도 10a는 Fv43E 뉴클레오티드 서열(서열 번호 5)을 도시킨 것이다. 도 10b는 Fv43E 아미노산 서열(서열 번호 6)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 11a 및 11b>

도 11a는 Fv39A 뉴클레오티드 서열(서열 번호 7)을 도시킨 것이다. 도 11b는 Fv39A 아미노산 서열(서열 번호 8)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 12a 및 12b>

도 12a는 Fv43A 뉴클레오티드 서열(서열 번호 9)을 도시킨 것이다. 도 12b는 Fv43A 아미노산 서열(서열 번호 10)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있으며, 예측된 CBM은 대문자로 되어 있고, 보존 도메인과 CBM을 분리하는 예측된 링커는 이탤릭체로 되어 있다.

<도 13a 및 13b>

도 13a는 Fv43B 뉴클레오티드 서열(서열 번호 11)을 도시킨 것이다. 도 13b는 Fv43B 아미노산 서열(서열 번호 12)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 14a 및 14b>

도 14a는 Pa51A 뉴클레오티드 서열(서열 번호 13)을 도시킨 것이다. 도 14b는 Pa51A 아미노산 서열(서열 번호 14)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 볼드체로 되어 있다. 트리코테르마 리세이에서의 발현을 위하여, 게놈 DNA를 코돈 최적화시켰다(도 27c 참조).

<도 15a 및 15b>

도 15a는 Gz43A 뉴클레오티드 서열(서열 번호 15)을 도시킨 것이다. 도 15b는 Gz43A 아미노산 서열(서열 번호 16)을 도시킨 것이다. 예측된 신호 서열은 밑줄이 그어져 있으며, 예측된 보존 도메인은 볼드체로 되어 있다. 트리코테르마 리세이에서의 발현을 위하여, 예측된 신호 서열을 트리코테르마 리세이에서 트리코테르마 리세이

CBH1 신호 서열(MYRKLAVISAFLETARA(서열 번호 159))로 치환하였다.

<도 16a 및 16b>

도 16a는 Fo43A 뉴클레오타이드 서열(서열 번호 17)을 도시한 것이다. 도 16b는 Fo43A 아미노산 서열(서열 번호 18)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다. 트리코테르마 리세이에서의 발현을 위하여, 예측된 신호 서열을 트리코테르마 리세이 CBH1 신호 서열(MYRKLAVISAFLETARA(신호 서열 159))로 치환하였다.

<도 17a 및 17b>

도 17a는 Af43A 뉴클레오타이드 서열(서열 번호 19)을 도시한 것이다. 도 17b는 Af43A 아미노산 서열(서열 번호 20)을 도시한 것이다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 18a 및 18b>

도 18a는 Pf51A 뉴클레오타이드 서열(서열 번호 21)을 도시한 것이다. 도 18b는 Pf51A 아미노산 서열(서열 번호 22)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 볼드체로 되어 있다. 트리코테르마 리세이에서의 발현을 위하여, 예측된 Pf51A 신호 서열을 트리코테르마 리세이 CBH1 신호 서열(MYRKLAVISAFLETARA(서열 번호 159))로 치환하고, Pf51A 뉴클레오타이드 서열을 트리코테르마 리세이에서의 발현을 위해 코돈 최적화시켰다.

<도 19a 및 19b>

도 19a는 AfuXyn2 뉴클레오타이드 서열(서열 번호 23)을 도시한 것이다. 도 19b는 AfuXyn2 아미노산 서열(서열 번호 24)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 GH11 보존 도메인은 볼드체로 되어 있다.

<도 20a 및 20b>

도 20a는 AfuXyn5 뉴클레오타이드 서열(서열 번호 25)을 도시한 것이다. 도 20b는 AfuXyn5 아미노산 서열(서열 번호 26)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 GH11 보존 도메인은 볼드체로 되어 있다.

<도 21a 및 21b>

도 21a는 Fv43D 뉴클레오타이드 서열(서열 번호 27)을 도시한 것이다. 도 21b는 Fv43D 아미노산 서열(서열 번호 28)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 22a 및 22b>

도 22a는 Pf43B 뉴클레오타이드 서열(서열 번호 29)을 도시한 것이다. 도 22b는 Pf43B 아미노산 서열(서열 번호 30)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 23a 및 23b>

도 23a는 뉴클레오타이드 서열(서열 번호 31)을 도시한 것이다. 도 23b는 Fv51A 아미노산 서열(서열 번호 32)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 볼드체로 되어 있다.

<도 24a 및 24b>

도 24a는 트리코테르마 리세이 Xyn3 뉴클레오타이드 서열(서열 번호 41)을 도시한 것이다. 도 24b는 트리코테르마 리세이 Xyn3 아미노산 서열(서열 번호 42)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 25a 및 25b>

도 25a는 트리코테르마 리세이 Xyn2의 아미노산 서열(서열 번호 43)을 도시한 것이다. 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다. 도 25b는 트리코테르마 리세이 Xyn2의 뉴클레오타이드 서열(서열 번호 162)을 도시한 것이다. 암호화 서열은 문헌[Törrönen *et al.* Biotechnology, 1992, 10:1461-65]에서 찾을 수 있다.

<도 26a 및 26b>

도 26a는 트리코테르마 리세이 Bx11의 아미노산 서열(서열 번호 44)을 도시한 것이다. 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다. 도 26b는 트리코테르마 리세이 Bx11의 뉴클레오티드 서열(서열 번호 163)을 도시한 것이다. 암호화 서열은 문헌[Margolles-Clark *et al.* Appl. Environ. Microbiol. 1996, 62(10):3840-46]에서 찾을 수 있다.

<도 27a 내지 27f>

도 27a는 트리코테르마 리세이 Bg11의 아미노산 서열(서열 번호 45)을 도시한 것이다. 신호 서열은 밑줄이 그어져 있다. 암호화 서열은 문헌[Barnett *et al.* Bio-Technology, 1991, 9(6):562-567]에서 찾을 수 있다. 도 27b는 Pa51A에 대한 추정된 cDNA(서열 번호 46)를 도시한 것이다. 도 27c는 Pa51A에 대한 코돈 최적화된 cDNA(서열 번호 47)를 도시한 것이다. 도 27d는 성숙 Gz43A를 암호화하는 게놈 DNA(서열 번호 48)의 CBH1 신호 서열(밑줄) 업스트림을 포함하는 구축물(construct)에 대한 암호화 서열이다. 도 27e는 성숙 Fo43A를 암호화하는 게놈 DNA(서열 번호 49)의 CBH1 신호 서열(밑줄) 업스트림을 포함하는 구축물에 대한 암호화 서열이다. 도 27f는 Pf51A를 암호화하는 코돈 최적화된 DNA(서열 번호 50)의 CBH1 신호 서열(밑줄) 업스트림을 포함하는 구축물에 대한 암호화 서열이다.

<도 28a 및 28b>

도 28a는 트리코테르마 리세이 Eg4의 뉴클레오티드 서열(서열 번호 51)을 도시한 것이다. 도 28b는 트리코테르마 리세이 Eg4의 아미노산 서열(서열 번호 52)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다. 예측된 링커는 이탤릭체로 되어 있다.

<도 29a 및 29b>

도 29a는 Pa3D의 뉴클레오티드 서열(서열 번호 53)을 도시한 것이다. 도 29b는 Pa3D의 아미노산 서열(서열 번호 54)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 30a 및 30b>

도 30a는 Fv3G의 뉴클레오티드 서열(서열 번호 55)을 도시한 것이다. 도 30b는 Fv3G의 아미노산 서열(서열 번호 56)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 31a 및 31b>

도 31a는 Fv3D의 뉴클레오티드 서열(서열 번호 57)을 도시한 것이다. 도 31b는 Fv3D의 아미노산 서열(서열 번호 58)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 32a 및 32b>

도 32a는 Fv3C의 뉴클레오티드 서열(서열 번호 59)을 도시한 것이다. 도 32b는 Fv3C의 아미노산 서열(서열 번호 60)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 33a 및 33b>

도 33a는 Tr3A의 뉴클레오티드 서열(서열 번호 61)을 도시한 것이다. 도 33b는 Tr3A의 아미노산 서열(서열 번호 62)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 34a 및 34b>

도 34a는 Tr3B의 뉴클레오티드 서열(서열 번호 63)을 도시한 것이다. 도 34b는 Tr3B의 아미노산 서열(서열 번호 64)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 35a 및 35b>

도 35a는 Te3A의 코돈 최적화된 뉴클레오타이드 서열(서열 번호 65)을 도시한 것이다. 도 35b는 Te3A의 아미노산 서열(서열 번호 66)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 36a 및 36b>

도 36a는 An3A의 뉴클레오타이드 서열(서열 번호 67)을 도시한 것이다. 도 36b는 An3A의 아미노산 서열(서열 번호 68)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 37a 및 37b>

도 37a는 Fo3A의 뉴클레오타이드 서열(서열 번호 69)을 도시한 것이다. 도 37b는 Fo3A의 아미노산 서열(서열 번호 70)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 38a 및 38b>

도 38a는 Gz3A의 뉴클레오타이드 서열(서열 번호 71)을 도시한 것이다. 도 38b는 Gz3A의 아미노산 서열(서열 번호 72)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 39a 및 39b>

도 39a는 Nh3A의 뉴클레오타이드 서열(서열 번호 73)을 도시한 것이다. 도 39b는 Nh3A의 아미노산 서열(서열 번호 74)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 40a 및 40b>

도 40a는 Vd3A의 뉴클레오타이드 서열(서열 번호 75)을 도시한 것이다. 도 40b는 Vd3A의 아미노산 서열(서열 번호 76)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 41a 및 41b>

도 41a는 Pa3G의 뉴클레오타이드 서열(서열 번호 77)을 도시한 것이다. 도 41b는 Pa3G의 아미노산 서열(서열 번호 78)을 도시한 것이다. 예측된 신호 서열은 밑줄이 그어져 있다. 예측된 보존 도메인은 볼드체로 되어 있다.

<도 42>

도 42는 Tn3B의 아미노산 서열(서열 번호 79)을 도시한 것이다. 표준 신호 예측 프로그램 SignalP는 예측된 신호 서열을 제공하지 않는다.

<도 43a 및 43b>

도 43a는 특정 β -글루코시다제 상동체의 아미노산 서열 정렬을 도시한 것이다. 도 43b는 β -글루코시다제 상동체의 정렬을 도시한 것으로, 그 일부는 단백질 가수분해 클리핑(clipping)에 대하여 민감하나 다른 것들은 그렇지 않지 않은 것으로 알려져 있다. 제1의 밑줄 그어진 영역은 대략적으로 이러한 효소 분류의 중앙에 위치하는 루프 서열 내에 있는 잔기를 포함한다. 제1의 밑줄 그어진 영역의 제 2의 밑줄 그어진 영역 다운스트림은 흔히 초기 단백질 가수분해 소화 또는 클리핑에 민감한 잔기를 포함한다.

<도 44>

도 44는 Fv3C 오픈 리딩 프레임이 있는 pENTR/D-TOPO 벡터를 도시한 것이다.

<도 45a 및 45b>

도 45a는 pTrex6g 벡터를 도시한 것이다. 도 45b는 발현 구축물 pTrex6g/Fv3C를 도시한 것이다.

<도 46a 내지 46c>

도 46a는 Fv3C 게놈 DNA 서열의 예측된 암호화 영역을 도시한 것이다. 도 46b는 Fv3C의 N-말단 아미노산 서열을 도시한 것이다. 화살표는 추정된 신호 펩티드 절단 부위를 보여준다. 성숙 단백질의 시작에는 밑줄이 그어 있다. 도 46c는 주석이 달린(annotated) (1) 및 대안적인 (2) 출발 코돈으로부터 Fv3C를 발현하는 트리코테르마 리세이 형질전환체의 SDS-PAGE 겔을 도시한 것이다.

<도 47>

도 47은 50℃에서 인산 팽윤된 셀룰로스의 당화에서의 다수의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여, 0.7% 셀룰로스, pH 5.0으로 인산 팽윤된 셀룰로스를 가수분해시켰다. 도면에서 백그라운드(background)로 표시된 시료는 β-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 2시간 동안 마이크로타이터 플레이트에서 수행하였다. 시료를 3별로 시험하였다. 이는 실시예 5A에 따른 것이다.

<도 48>

도 48은 50℃에서 산으로 전처리된 옥수수 대(PCS)의 당화에서의 다수의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여 13% 고형분, pH 5.0로 PCS를 가수분해시켰다. 도면에서 백그라운드로 표시된 시료는 β-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 48시간 동안 마이크로타이터 플레이트에서 수행하였다. 시료를 3별로 시험하였다. 실험 상세사항은 실시예 5B에 기재되어 있다.

<도 49>

도 49는 50℃에서 희석 암모니아로 전처리된 옥수수 속대의 당화에서의 다수의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 8 mg/g 헤미셀룰라제 및 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여 20% 고형분, pH 5.0으로 희석 암모니아로 전처리된 옥수수 속대를 가수분해시켰다. 도면에서 백그라운드로 표시된 시료는 β-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 + 8 mg/g 헤미셀룰로스 혼합물 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 48시간 동안 마이크로타이터 플레이트에서 수행하였다. 시료를 3별로 시험하였다. 실험 상세사항은 실시예 5C에 기재되어 있다.

<도 50>

도 50은 50℃에서 수산화나트륨(NaOH)으로 전처리된 옥수수 속대의 당화에서의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여 17% 고형분, pH 5.0으로 NaOH로 전처리된 옥수수 속대를 가수분해시켰다. 도면에서 백그라운드로 표시된 시료는 β-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 혼합물 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 48시간 동안 마이크로타이터 플레이트에서 수행하였다. 각 시료를 4별로 실시하였다. 이는 실시예 5D에 따른 것이다.

<도 51>

도 51은 50℃에서 희석 암모니아로 전처리된 스위치그래스의 당화에서의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여 17% 고형분, pH 5.0으로 스위치그래스를 가수분해시켰다. 도면에서 백그라운드로 표시된 시료는 β-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 혼합물 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 48시간 동안 마이크로타이터 플레이트에서 수행하였다. 각 시료를 4별로 실시하였다. 실험 상세사항은 실시예 5E에 기재되어 있다.

<도 52>

도 52는 50℃에서 AFEX 옥수수 대의 당화에서의 전체 셀룰라제와 β-글루코시다제의 혼합물의 성능을 비교한 것이다. 이 실험에서, 10 mg(단백질)/g(셀룰로스)의 전체 셀룰라제를 5 mg/g β-글루코시다제와 배합하여, 효소 혼합물을 사용하여 14% 고형분, pH 5.0으로 AFEX 옥수수 대를 가수분해시켰다. 도면에서 백그라운드로 표시된 시료는 베타-글루코시다제가 첨가되지 않은 10 mg/g 전체 셀룰라제 혼합물 단독으로부터 수득된 전환을 나타내었다. 반응을 50℃에서 48시간 동안 마이크로타이터 플레이트에서 수행하였다. 각 시료를 4별로 실시하였다.

실험 상세사항은 실시예 5F에 기재되어 있다.

<도 53a 내지 53c>

도 53a 내지 53c는 0 내지 50%의 다양한 β -글루코시다제 대 전체 셀룰라제의 비율로 20% 고형분에서의 희석 암모니아로 전처리된 옥수수 대로부터의 글루칸 전환율을 도시한 것이다. 효소 용량을 각 실험에서 일정하게 유지하였다. 도 53a는 트리코테르마 리세이 Bgl1을 사용하여 행한 실험을 도시한 것이다. 도 53b는 Fv3C를 사용하여 행한 실험을 도시한 것이다. 도 53c는 아스페르길루스 니게르 Bglu(An3A)를 사용하여 행한 실험을 도시한 것이다.

<도 54>

도 54는 실시예 7에 따라, 2.5 내지 40 mg/g 글루칸의 수준으로 투여되는 세가지 상이한 효소 조성물에 의한, 20% 고형분에서의 희석 암모니아로 전처리된 옥수수 대로부터의 글루칸 전환율을 도시한 것이다. \triangle 는 아셀러라제(Accellerase) 1500 + 멀티펙트 자일라나제(Multifect Xylanase)로 관찰되는 글루칸 전환을 표시하며, \diamond 는 트리코테르마 리세이 통합 균주 H3A로부터의 전체 셀룰라제로 관찰되는 글루칸 전환을 표시하고, \blacklozenge 는 트리코테르마 리세이 통합 균주 H3A로부터의 75 wt.% 전체 셀룰라제 + 25 wt.% Fv3C를 포함하는 효소 조성물로 관찰되는 글루칸 전환을 표시한다.

<도 55a 내지 55i>

도 55a는 아스페르길루스 니게르에서의 발현을 위해 사용되는 pRAX2-Fv3C 발현 플라스미드의 맵을 도시한 것이다. 도 55b는 pENTR-TOPO-Bgl1-943/942 플라스미드를 도시한 것이다. 도 55c는 pTrex3g 943/942 발현 벡터를 도시한 것이다. 도 55d는 pENTR/트리코테르마 리세이 Xyn3 플라스미드를 도시한 것이다. 도 55e는 pTrex3g/트리코테르마 리세이 Xyn3 발현 벡터를 도시한 것이다. 도 55f는 pENTR-Fv3A 플라스미드를 도시한 것이다. 도 55g는 pTrex6g/Fv3A 발현 벡터를 도시한 것이다. 도 55h는 TOPO Blunt/Peg11-Fv43D 플라스미드를 도시한 것이다. 도 55i는 TOPO Blunt/Peg11-Fv51A 플라스미드를 도시한 것이다.

<도 56>

도 56은 트리코테르마 리세이 β -자일로시다제 Bx11과 Fv3A 간의 아미노산 정렬을 도시한 것이다.

<도 57>

도 57은 특정 GH43 패밀리 가수분해효소의 아미노산 서열 정렬을 도시한 것이다. 패밀리의 구성원 간에 보존된 아미노산 잔기에는 밑줄이 그어져 있고, 볼드체로 되어 있다.

<도 58>

도 58은 특정 GH51 패밀리 효소의 아미노산 서열 정렬을 도시한 것이다. 패밀리의 구성원 간에 보존된 아미노산 잔기에는 밑줄이 그어져 있고, 볼드체로 되어 있다.

<도 59a 및 59b>

다수의 GH10 및 GH11 패밀리 엔도자일라나제의 아미노산 서열 정렬을 도시한 것이다. 도 59a: GH10 패밀리 자일라나제의 정렬. 볼드체의 밑줄이 그어져 있는 잔기는 촉매 친핵체 잔기(정렬 위에 "N"으로 표시)이다. 도 59b: GH11 패밀리 자일라나제의 정렬. 볼드체의 밑줄이 그어져 있는 잔기는 촉매 친핵체 잔기 및 일반 산 염기 잔기(정렬 위에, 각각 "N" 및 "A"로 표시)이다.

<도 60a 내지 60c>

도 60a는 Fv3C/트리코테르마 리세이 Bgl3("FB") 키메라/융합 폴리펩티드를 암호화하는 유전자의 개략도를 도시한 것이다. 도 60b는 융합/키메라 폴리펩티드 Fv3C/트리코테르마 리세이 Bgl3("FB")을 암호화하는 뉴클레오타이드 서열(서열 번호 82)을 도시한 것이다. 도 60c는 융합/키메라 폴리펩티드 Fv3C/트리코테르마 리세이 Bgl3을 암호화하는 아미노산 서열(서열 번호 159)을 도시한 것이다. 볼드체의 서열은 트리코테르마 리세이 Bgl3으로부터의 것이다.

<도 61>

도 61은 pTTT-pyrG13-Fv3C/Bgl3 융합 플라스미드의 맵을 도시한 것이다.

<도 62>

도 62는 희석 암모니아로 전처리된 옥수수 속대의 당화에서의 아스페르길루스 니게르에 생성된 트리코테르마 리세이 Bgl1(닫힌 마름모꼴) 및 Fv3C(열린 마름모꼴)를 비교한 것이다. 이 실험에서, 트리코테르마 리세이 Bgl1 및 Fv3C를 0 내지 10 mg(단백질)/g(셀룰로스)로부터 10 mg/g H3A-5의 일정 수준으로 로딩하고, 이들 혼합물을 사용하여, 5% 셀룰로스, pH 5.0으로 희석 암모니아로 전처리된 옥수수 속대를 가수분해시켰다. 반응을 50℃에서 2일간 마이크로타이터 플레이트에서 수행하였다. 각 시료를 5별로 검정하였다. 실험 상세사항은 실시예 13에 기재되어 있다.

<도 63>

도 63은 50 mM 아세트산나트륨 완충제, pH 5에서 90℃/r 주사 속도(25℃-110℃)로 수집된 β -글루코시다제 트리코테르마 리세이 Bgl1(Tr3A), Fv3C, 및 Fv3C/Te3A/Bgl1("FAB") 키메라 폴리펩티드의 DSC 프로파일이다.

<도 64a 내지 64e>

도 64a: 전체 셀룰라제의 성능: 50℃에서 인산 팽윤된 셀룰로스의 당화에서의 트리코테르마 리세이 Bgl3 혼합물. 도 64b: 37℃에서 인산 팽윤된 셀룰로스의 당화에서의 트리코테르마 리세이 Bgl3 혼합물. 도 64c: 50℃에서 산으로 전처리된 옥수수 대의 당화에서의 트리코테르마 리세이 Bgl3 혼합물. 도 64d: 37℃에서 산으로 전처리된 옥수수 대의 당화에서의 트리코테르마 리세이 Bgl3 혼합물.

<도 65a 및 65b>

도 65a는 인산 팽윤된 셀룰로스 당화에서 트리코테르마 리세이 Bgl1(닫힌 마름모꼴)과 트리코테르마 리세이 Bgl3(열린 마름모꼴)을 비교한 것이다. 도 65b는 인산 팽윤된 셀룰로스의 당화에서 트리코테르마 리세이 Bgl1(좌측 패널)과 트리코테르마 리세이 Bgl3(우측 패널)에 의해 생성된 셀로비오스(블랙 바) 및 글루코스(화이트 바)를 비교한 것이다.

<도 66>

도 66은 다수의 프라이머의 뉴클레오티드 서열을 도시한 것이다.

<도 67a 및 67b>

도 67a는 Fv3C/Te3A/트리코테르마 리세이 Bgl3("FAB")의 전장 아미노산 서열(서열 번호 135)(Te3A는 볼드 이텔릭체의 대문자로 되어 있고, 트리코테르마 리세이 Bgl3은 밑줄 그어진 대문자로 되어 있음)을 도시한 것이다. 도 67b는 Fv3C/Te3A/트리코테르마 리세이 Bgl3("FAB") 키메라를 암호화하는 핵산 서열(서열 번호 83)을 도시한 것이다.

<도 68a 내지 68c>

도 68a는 특정 키메라 β -글루코시다제 폴리펩티드의 N- 및 C- 말단 도메인에 존재하는 구조 모티프를 열거한 표이다. 도 68b는 본 발명의 적절한 β -글루코시다제 폴리펩티드 하이브리드/키메라를 설계하는데 사용되는 특정 아미노산 서열 모티프를 열거한 표이다. 도 68c는 GH61/엔도글루카나제의 아미노산 서열 모티프를 열거한 것이다.

<도 69>

도 69는 Pa3C의 뉴클레오티드 및 단백질 서열(각각, 서열 번호 80 및 81)을 도시한 것이다.

<도 70a 내지 70g>

도 70a는 "삽입 1"의 구조를 보이게 하는 제1 각도로부터 관찰한 Fv3C 및 Te3A, 및 트리코테르마 리세이 Bgl1의 3차원 중첩 구조를 도시한 것이다. 도 70b는 "삽입 2"의 구조를 보이게 하는 제2 각도로부터 관찰한 동일한 중첩 구조를 도시한 것이다. 도 70c는 "삽입 3"의 구조를 보이게 하는 제3 각도로부터 관찰한 동일한 중첩 구조를 도시한 것이다. 도 70d는 "삽입 4"의 구조를 보이게 하는 제4 각도로부터 관찰한 동일한 중첩 구조를 도시한 것이다. 도 70e는 모두 루프-유사 구조인 삽입 1 내지 4로 표시된 트리코테르마 리세이 Bgl1(Q12715_TRI), Te3A(ABG2_T_eme), 및 Fv3C(FV3C)의 서열 정렬이다. 도 70f는 잔기 W59/W33 및 W355/W325(Fv3C/Te3A) 간의 보존된 상호작용을 나타내는, Fv3C(연회색), Te3A(진회색) 및 트리코테르마 리세이 Bgl1(흑색)의 구조의 중첩된 부분을 도시한 것이다. 도 70g는 제1 잔기의 쌍: S57/31 및 N291/261(Fv3C/Te3A) 간의 보존된 상호작용; 및 제2 잔기의 그룹: Y55/29, P775/729 및 A778/732(Fv3C/Te3A) 간의 보존된 상호작용을 나타내는, Fv3C(연회색), Te3A(진회색) 및 트리코테르마 리세이 Bgl1(흑색)의 구조의 중첩된 부분을 도시한 것이다. 도 70h는 "삽입 2"

내에서, K162에서 Fv3C와, V409의 주쇄 산소 원자의 수소 결합 상호작용, Te3A에는 보존되지만, 트리코테르마 리세이 Bgl1에서는 관찰되지 않는 상호작용을 나타내는, 구조 Fv3C(진회색) 및 트리코테르마 리세이 Bgl1(흑색)의 중첩된 부분을 도시한 것이다. 도 70i (a) 및 (b)는 Fv3C, Te3A 및 서열 번호 135의 키메라/하이브리드 β -글루코시다제 중에 공유되는 서열 번호 168 내의 보존된 글리코실화 부위를 도시한 것이며, (a)는 Te3A(진회색) 및 트리코테르마 리세이 Bgl1(흑색)과 중첩되는 동일한 영역을 도시한 것이며; (b)는 서열 번호 135의 키메라/하이브리드 β -글루코시다제(연회색), Te3A(진회색) 및 트리코테르마 리세이 Bgl1(흑색)과 중첩되는 동일한 영역을 도시한 것이다. 흑색 화살표는 글리코실화 글리칸을 매립하는 것으로 보이는 Te3A(서열 번호 135의 하이브리드 β -글루코시다제 내에도 존재) 내의 "삽입 3"의 루프 구조를 나타낸다. 도 70j는 Fv3C 및 Te3A의 "삽입 2"의 W95/68(Fv3C/Te3A)과 상호작용하는 잔기 W386/355 간의 보존된 상호작용을 나타내는, Fv3C(연회색), Te3A(진회색) 및 트리코테르마 리세이 Bgl1(흑색)의 구조의 중첩된 부분을 도시한 것이다. 상호작용은 트리코테르마 리세이 Bgl1에서 없어진다.

<도 71a 내지 71c>

도 71a는 실시예 13에 따라, 44시간의 50℃ 인큐베이션 후에 가용성 분획(상청액) 중의 측정된 비결합 단백질의 양을 도시한 것이다. 도 71b는 실시예 13에 따라, 44시간의 50℃ 인큐베이션 후에 슬러리 중의 총 단백질(결합 및 비결합)을 도시한 것이다. 도 71c는 실시예 13에 따라, 완충제에서의 추가의 30분간의 인큐베이션 후에 슬러리 중의 비결합 단백질을 도시한 것이다.

발명의 상세한 설명

효소는 관례적으로 기질 특이성 및 반응 산물에 의해 분류되어 왔다. 게놈 시대 이전에는 기능이 효소를 비교하기 위한 가장 다루기 쉬운(아마도 가장 유용한) 기초로 간주되었고, 다양한 효소 활성화에 대한 검정법이 수년간 널리 개발되어서, 잘 알려진 EC 분류 체계로 이어졌다. 2개의 탄수화물 부분(또는 탄수화물 및 비-탄수화물 부분 - 니트로페놀-글리코시드 유도체에서 발생한 바와 같음) 사이의 글리코시드 결합 상에서 작용하는 셀룰라제 및 다른 글리코실 하이드롤라제는 이러한 분류 체계 하에서 EC 3.2.1로 지정되며, 마지막 숫자는 절단된 결합의 정확한 유형을 나타낸다. 예를 들어, 이러한 체계에 따라 엔도-작용 셀룰라제(1,4- β -엔도글루카나제)는 EC 3.2.1.4로 지정된다.

널리 보급된 게놈 시퀀싱 프로젝트의 등장으로, 시퀀싱 데이터에 의해 분석과, 관련 유전자 및 단백질의 비교가 용이하게 되었다. 또한, 탄수화물 부분에 작용할 수 있는 증가하는 수의 효소 (즉, 카보하이드라제)를 결정화하고 그들의 3차원 구조를 확인하였다. 그러한 분석으로 관련 서열을 갖는 별개의 효소의 패밀리를 동정하였으며, 이는 이들의 아미노산 서열을 기초로 하여 예측될 수 있는 보존된 3차원 폴드(fold)를 포함한다. 추가로, 동일하거나 유사한 3차원 폴드를 갖는 효소가 가수분해의 동일하거나 유사한 입체특이성을 나타내는 것으로 밝혀졌고, 심지어 상이한 반응을 촉매하는 경우에도 그러하다(문헌[Henrissat *et al.*, FEBS Lett 1998, 425(2): 352-4]; 문헌[Coutinho and Henrissat, Genetics, biochemistry and ecology of cellulose degradation, 1999, T. Kimura. Tokyo, Uni Publishers Co: 15-23]).

이러한 발견은 카보하이드라제 모듈의 서열-기초 분류의 기반을 형성하였고, 이는 www.cazy.org에서 인터넷 데이터베이스, 탄수화물-활성 효소 서버(Carbohydrate-Active enZYme server(CAZy))의 형태로 이용가능하다(문헌[Cantarel *et al.*, 2009, The Carbohydrate-Active EnZymes database(CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37 (Database issue issue):D233-38] 참조).

CAZy는 촉매되는 반응의 유형에 의해 구별될 수 있는 카보하이드라제의 4개의 주요 분류를 정의한다: 글리코실 하이드롤라제(GH's), 글리코실트랜스퍼라제(GT's), 폴리사카라이드 리아제(PL's) 및 탄수화물 에스테라제(CE's). 본 개시내용의 효소는 글리코실 하이드롤라제이다. GH's는 2개의 탄수화물 사이, 또는 탄수화물과 비-탄수화물 부분 사이의 글리코시드 결합을 가수분해하는 효소의 그룹이다. 서열 유사성에 의해 그룹화되는 글리코실 하이드롤라제에 대한 분류 시스템은 120개 초과 상이한 패밀리의 정의를 야기하였다. 이러한 분류는 CAZy 웹 사이트에서 이용가능하다. 본 발명의 효소는 글리코실 하이드롤라제 패밀리 3(GH3)에 속한다.

GH3 효소는 예를 들어, β -글루코시다제(EC:3.2.1.21); β -자일로시다제(EC:3.2.1.37); N-아세틸 β -글루코사미니다제(EC:3.2.1.52); 글루칸 β -1,3-글루코시다제(EC:3.2.1.58); 셀로텍스트리나제(EC:3.2.1.74); 엑소-1,3-1,4-글루카나제(EC:3.2.1); 및 β -갈락토시다제(EC 3.2.1.23)를 포함한다. 예를 들어, GH3 효소는 β -글루코시다제, β -자일로시다제, N-아세틸 β -글루코사미니다제, 글루칸 β -1,3-글루코시다제, 셀로텍스트리나제, 엑소-1,3-1,4-글루카나제 및/또는 β -갈락토시다제 활성을 갖는 것일 수 있다. 일반적으로, GH3 효소는 구형 단백질

이고 둘 이상의 서브도메인(subdomain)으로 이루어질 수 있다. 촉매 잔기는 β -글루코시다제에서, 펩티드의 세 번째 N-말단에 위치하고 아미노산 단편 SDW 내에 위치하는 아스파르테이트 잔기인 것으로 동정되었다(문헌[Li *et al.* 2001, Biochem. J. 355:835-840]). 트리코테르마 리세이로부터의 Bg11에서 상응하는 서열은 T266D267W268이고(출발 위치에서 메티오닌으로부터 카운팅), 촉매 잔기 아스파르테이트는 D267이다. 또한, 하이드록실/아스파르테이트 서열은 시험된 GH3 β -자일로시다제에서 보존된다. 예를 들어, 트리코테르마 리세이 Bx11에서 상응하는 서열은 S310D311이고, Fv3A에서 상응하는 서열은 S290D291이다.

본 발명의 폴리펩티드

셀룰라제

본 개시내용의 조성물은 하나 이상의 셀룰라제를 포함할 수 있다. 셀룰라제는 셀룰로스(β -1,4-글루칸 또는 β -D-글루코시드 결합)를 가수분해하여, 글루코스, 셀로비오스, 셀로올리고당 등의 생성을 야기하는 효소이다. 셀룰라제는 관례적으로 3개의 주요 분류로 기질 특이성 및 반응 산물에 의해 분류되어 왔다: 엔도글루카나제(EC 3.2.1.4)("EG"), 엑소글루카나제 또는 셀로비오하이드롤라제(EC 3.2.1.91)("CBH") 및 β -글루코시다제(β -D-글루코시드 글루코하이드롤라제; EC 3.2.1.21)("BG")(문헌[Knowles *et al.*, 1987, Trends in Biotechnology 5(9):255-261]; 문헌[Shulein, 1988, Methods in Enzymology, 160:234-242]).

본 개시내용의 방법 및 조성물에 따라 사용되는 셀룰라제는 제한 없이 하나 이상의 하기 유기체로부터 수득될 수 있거나 재조합에 의해 생성될 수 있다: 크리소스포리움 룩노웬스(*Chrysosporium lucknowense*), 크리니펠리스 스카펠라(*Crinipellis scapella*), 마크로포미나 파세올리나(*Macrophomina phaseolina*), 마이셀리오프토라 썬모필라(*Myceliophthora thermophila*), 소르다리아 피미콜라(*Sordaria fimicola*), 볼루텔라 콜레토티리코이데스(*Volutella colletotrichoides*), 티엘라비아 테레스트리스(*Thielavia terrestris*), 아크레모니움(*Acremonium*) sp., 엑시디아 글란둘로사(*Exidia glandulosa*), 포메스 포멘타리우스(*Fomes fomentarius*), 스폰기펠리스(*Spongipellis*) sp., 리조플릭티스 로세아(*Rhizophlyctis rosea*), 리조무코르 푸실루스(*Rhizomucor pusillus*), 파이코마이세스 니테우스(*Phycomyces niteus*), 카에토스틸룸 프레세니이(*Chaetostylum fresenii*), 디플로디아 고사이피나(*Diplodia gossypina*), 울로스포라 빌그라미이(*Ulospora bilgramii*), 사코볼루스 딜루텔루스(*Saccobolus dilutellus*), 페니실리움 베루쿨로숨(*Penicillium verruculosum*), 페니실리움 크라이소게눔(*Penicillium chrysogenum*), 썬모마이세스 베루코수스(*Thermomyces verrucosus*), 디아포르테 사인게네시아(*Diaporthe syngenesia*), 콜레토티리쿰 라게나리움(*Colletotrichum lagenarium*), 니그로스포라(*Nigrospora*) sp., 자일라리아 하이폭실론(*Xylaria hypoxylon*), 넥트리아 피네아(*Nectria pinea*), 소르다리아 마크로스포라(*Sordaria macrospora*), 티엘라비아 썬모필라(*Thielavia thermophila*), 카에토미움 모로룸(*Chaetomium mororum*), 카에토미움 비르센스(*Chaetomium virscens*), 카에토미움 브라질리엔시스(*Chaetomium brasiliensis*), 카에토미움 쿠니콜로룸(*Chaetomium cunicolorum*), 사이스파스토스포라 보닌렌시스(*Syspastospora boninensis*), 클라도리눔 포에쿤디시눔(*Cladorrhinum foecundissimum*), 사이탈리디움 썬모필라(*Scytalidium thermophila*), 글리오클라디움 카테눌라툼(*Gliocladium catenulatum*), 푸사리움 옥시스포룸(*Fusarium oxysporum*) ssp. 라이코페르시시(*lycopersici*), 푸사리움 옥시스포룸 ssp. 파시플로라(*passiflora*), 푸사리움 솔라니(*Fusarium solani*), 푸사리움 안구이오이데스(*Fusarium anguoides*), 푸사리움 포아에(*Fusarium poae*), 후미콜라 니그레센스(*Humicola nigrescens*), 후미콜라 그리세아(*Humicola grisea*), 파나에올루스 레티루기스(*Panaeolus retirugis*), 트라메테스 상귀네아(*Trametes sanguinea*), 스킴조필룸 코뮤네(*Schizophyllum commune*), 트리코테슘 로세움(*Trichothecium roseum*), 마이크로스페롭시스(*Microsphaeropsis*) sp., 악소볼루스 스틱토이데우스(*Acsobolus stictoideus*) spej., 포로니아 푼크타타(*Poronia punctata*), 노둘리스포룸(*Nodulisporum*) sp., 트리코테르마(*Trichoderma*) sp.(예를 들어, 트리코테르마 리세이) 및 실린드로카르폰(*Cylindrocarpon*) sp. 셀룰라제는 또한 박테리아로부터 수득되거나 재조합에 의해 생성될 수 있거나, 효모로부터 재조합에 의해 생성될 수 있다.

예를 들어, 본 개시내용의 방법 및/또는 조성물에 사용하기 위한 셀룰라제는 전체 셀룰라제이고/이거나, 칼코플루오르 검정법에 의해 측정된 것으로서, 적어도 0.1(예를 들어, 0.1 내지 0.4) 분율의 생성물을 달성할 수 있다.

β -글루코시다제

β -글루코시다제(들) (또는 본 명세서에서 상호교환적으로 " β -글루코시다제 폴리펩티드(들)")는 글루코스의 방출과 함께 β -D-글루코시드의 말단 비환원성 잔기의 가수분해를 촉매한다. β -글루코시다제 폴리펩티드의 예에는 β -글루코시다제 폴리펩티드의 적어도 하나의 활성을 갖는 폴리펩티드, 폴리펩티드 단편, 펩티드, 및 융합

폴리펩티드가 포함된다. β -글루코시다제 폴리펩티드 및 핵산의 예에는 본 명세서에 기재된 임의의 공급원 유기체로부터의 고유 폴리펩티드(예를 들어, 변이체 포함) 및 핵산, 및 β -글루코시다제 폴리펩티드의 적어도 하나의 활성을 갖는 본 명세서에 기재된 임의의 공급원 유기체로부터의 돌연변이체 폴리펩티드 및 핵산이 포함된다.

본 개시내용의 조성물은 하나 이상의 β -글루코시다제 폴리펩티드를 포함할 수 있다. 본 명세서에 사용되는 용어 " β -글루코시다제"는 EC 3.2.1.21로 분류된 β -D-글루코시드 글루코하이드롤라제, 및/또는 셀로비오스의 가수분해를 촉매하여 β -D-글루코스를 방출하는 GH 패밀리의 3의 구성원을 지칭한다. 본 발명의 GH3 β -글루코시다제는 제한 없이, Fv3C, Pa3D, Fv3G, Fv3D, Tr3A("트리코테르마 리세이 Bgl1" 또는 "트리코테르마 리세이 Bgl1"로도 명명), Tr3B("트리코테르마 리세이 Bgl3"로도 명명), Te3A, An3A("아스페르길루스 니게르 Bglu"로도 명명), Fo3A, Gz3A, Nh3A, Vd3A, Pa3G, 또는 Tn3B 폴리펩티드를 포함한다. 일부 실시형태에서, 본 명세서의 GH3 β -글루코시다제 폴리펩티드는 β -글루코시다제 폴리펩티드의 적어도 하나의 활성을 갖는다.

적절한 β -글루코시다제 폴리펩티드는 다수의 미생물로부터 획득되거나, 재조합 수단에 의해 획득되거나, 상업적 공급원으로부터 구입될 수 있다. 미생물로부터의 β -글루코시다제의 예에는 제한 없이, 박테리아 및 진균으로부터의 것들이 포함된다. 예를 들어, 본 개시내용의 β -글루코시다제는 적절하게 사상 진균으로부터 획득된다.

β -글루코시다제 폴리펩티드는 그 중에서도, 아스페르길루스 아쿨레아투스(*A. aculeatus*) (문헌[Kawaguchi et al. Gene 1996, 173: 287-288]), 아스페르길루스 카와치(*A. kawachi*) (문헌[Iwashita et al. Appl. Environ. Microbiol. 1999, 65: 5546-5553]), 아스페르길루스 오리자에(*A. oryzae*) (국제 특허 공개 제WO 2002/095014호), 셀룰로모나스 비아조테아(*C. biazotea*) (문헌[Wong et al. Gene, 1998, 207:79-86]), 페니실리움 푸니쿨로숨(*P. funiculosum*) (국제 특허 공개 제WO 2004/078919호), 사카로마이코시스 피불리게라(*S. fibuligera*) (문헌[Machida et al. Appl. Environ. Microbiol. 1988, 54: 3147-3155]), 스킨조사카로마이세스 폼베(*S. pombe*) (문헌[Wood et al. Nature 2002, 415: 871-880]), 트리코테르마 리세이(예를 들어, β -글루코시다제 1(미국 특허 제6,022,725호), β -글루코시다제 3 미국 특허 제6,982,159호), β -글루코시다제 4(미국 특허 제7,045,332), β -글루코시다제 5(미국 특허 제7,005,289호), β -글루코시다제 6(미국 특허 공개 제20060258554호), β -글루코시다제 7(미국 특허 공개 제20060258554호)), 포도스포라 안세리나(예를 들어, Pa3D), 푸사리움 베르티실리오이데스(*F. verticillioideus*) (예를 들어, Fv3G, Fv3D, 또는 Fv3C), 트리코테르마 리세이(예를 들어, Tr3A, 또는 Tr3B), 탈라로마이세스 에메르소니이(*T. emersonii*) (예를 들어, Te3A), 아스페르길루스 니게르(예를 들어, An3A), 푸사리움 옥시스포룸(*F. oxysporum*) (예를 들어, Fo3A), 지베렐라 제아에(*G. zeae*) (예를 들어, Gz3A), 넥트리아 헤마토코카(*N. haematococca*) (예를 들어, Nh3A), 버티실리움 달리아에(*V. dahliae*) (예를 들어, Vd3A), 포도스포라 안세리나(예를 들어, Pa3G), 또는 씨모토가 네아폴리타(예를 들어, Tn3B)로부터 획득되거나 재조합에 의해 생성될 수 있다.

β -글루코시다제 폴리펩티드는 β -글루코시다제, 변이체, 하이브리드/키메라/융합, 또는 돌연변이체를 암호화하는 내인성/외인성 유전자를 발현함으로써 생성될 수 있다. 예를 들어, β -글루코시다제 폴리펩티드는 예를 들어, 그람 양성균, 예컨대 바실루스(*Bacillus*) 또는 방선균류(*Actinomycetes*), 또는 진핵생물 숙주, 예컨대 진균(예를 들어, 트리코테르마, 크리소스포리움, 아스페르길루스, 사카로마이세스, 피치아(*Pichia*))에 의해 세포의 공간으로 분비될 수 있다. β -글루코시다제 폴리펩티드는 효모, 예컨대 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*)에서 발현될 수 있다. β -글루코시다제 폴리펩티드는 과발현되거나 저발현될 수 있다.

β -글루코시다제 폴리펩티드는 또한 상업적 공급원으로부터 획득될 수 있다. 본 개시내용에 사용하기에 적합한 시판용 β -글루코시다제 제제의 예에는 예를 들어, 아셀러라제(등록상표) BG(다니스코 유에스 인코포레이티드(Danisco US Inc.)), 제넨코(Genencor))의 트리코테르마 리세이 β -글루코시다제; NOVOZYM™ 188(아스페르길루스 니게르 유래의 β -글루코시다제); 아그로박테리움(*Agrobacterium*) sp. β -글루코시다제, 및 메가자임(Megazyme)(아일랜드 소재의 메가자임 인터내셔널 아일랜드 엘티디.(Megazyme International Ireland Ltd.))의 씨모토가 마리티마(*T. maritima*) β -글루코시다제가 포함된다.

게다가, β -글루코시다제 폴리펩티드는 셀룰라제 조성물, 전체 세포 셀룰라제 조성물, 셀룰라제 발효 브로쓰, 또는 전체 브로쓰 포틀레이션 셀룰라제 조성물의 성분일 수 있다.

β -글루코시다제 활성은 당업계에 공지된 다수의 적절한 수단에 의해 측정될 수 있으며, 비제한적인 예로, 문헌[Chen et al., in *Biochimica et Biophysica Acta* 1992, 121:54-60]에 기재된 검정법을 들 수 있는데, 여기서

1 pNPG는 50℃ 및 pH 4.8에서 10분 내에 4-니트로페닐-β-D-글루코피라노시드로부터 유리되는 1 μmol의 니트로페놀을 나타낸다.

β-글루코시다제 폴리펩티드는 적절하게는 본 발명의 셀룰라제 조성물 중의 효소의 총 중량의 약 0 wt.% 내지 약 75 wt.%를 구성한다. 서로에 대한 임의의 효소 쌍의 비는 본 명세서의 개시내용에 기초하여 용이하게 계산될 수 있다. 본 명세서에 개시된 중량 백분율로부터 유도가능한 임의의 중량비의 효소를 포함하는 셀룰라제 조성물이 고려된다. β-글루코시다제 함량은 하한치가 셀룰라제 조성물 중의 효소의 총 중량의 약 0 wt.%, 1 wt.%, 2 wt.%, 3 wt.%, 4 wt.%, 5 wt.%, 6 wt.%, 7 wt.%, 8 wt.%, 9 wt.%, 10 wt.%, 12 wt.%, 15 wt.%, 17%, 20 wt.%, 25 wt.%, 30 wt.%, 40 wt.%, 45 wt.%, 또는 50 wt.%이고, 상한치가 셀룰라제 조성물 중의 효소의 총 중량의 약 10 wt.%, 12 wt.%, 15 wt.%, 17 wt.%, 20 wt.%, 25 wt.%, 30 wt.%, 35 wt.%, 40 wt.%, 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.%, 또는 70 wt.%인 범위로 될 수 있다. 예를 들어, β-글루코시다제(들)는 적절하게는 셀룰라제 조성물 중의 효소의 총 중량의 약 0.1 wt.% 내지 약 40 wt.%, 약 1 wt.% 내지 약 35 wt.%, 약 2 wt.% 내지 약 30 wt.%, 약 5 wt.% 내지 약 25 wt.%, 약 7 wt.% 내지 약 20 wt.%, 약 9 wt.% 내지 약 17 wt.%, 약 10 wt.% 내지 약 20 wt.%, 또는 약 5 wt.% 내지 약 10 wt.%를 나타낸다.

돌연변이체 β-글루코시다제 폴리펩티드

본 개시내용은 돌연변이체 β-글루코시다제 폴리펩티드를 제공한다. 돌연변이체 β-글루코시다제 폴리펩티드는 하나 이상의 아미노산 잔기가 β-글루코시다제 활성(즉, 글루코스의 방출과 함께 β-D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 능력)을 보유하면서 아미노산 치환을 행한 것을 포함한다. 이와 같이, 돌연변이체 β-글루코시다제 폴리펩티드는 그 용어가 본 명세서에 정의된 바와 같이 특정 유형의 "β-글루코시다제 폴리펩티드"를 구성한다. 돌연변이체 β-글루코시다제 폴리펩티드는 하나 이상의 아미노산을 폴리펩티드의 고유 또는 야생형 아미노산 서열로 치환시킴으로써 형성될 수 있다. 일부 태양에서, 본 발명은 전구체 효소 아미노산 서열에 비해서 변경된 아미노산 서열을 포함하는 폴리펩티드를 포함하며, 여기서, 돌연변이체 효소는 전구체 효소의 특징적인 셀룰로스 분해 성질을 보유하나, 전구체 효소에 비하여 변경된 특성, 일부 특정 태양에서는, 예를 들어, 최적의 pH의 증가 또는 감소; 산화 안정성의 증가 또는 감소; 열 안정성의 증가 또는 감소 및 하나 이상의 기질에 대한 비활성도 수준의 증가 또는 감소를 가질 수 있다. 생물학적 활성에 영향을 미치지 않고, 어떠한 아미노산 잔기가 치환되거나, 삽입되거나, 결실될 수 있는지의 결정에 관한 지침은 당업계에 널리 공지되어 있는 컴퓨터 프로그램, 예를 들어, 레이저진(LASERGENE) 소프트웨어(디엔에이스타(DNASTAR))를 사용하여 관찰될 수 있다. 아미노산 치환은 보존적 또는 비보존적일 수 있고, 이러한 치환된 아미노산 잔기는 유전 암호에 의해 암호화된 것이거나 아닐 수 있다. 아미노산 치환은 폴리펩티드 탄수화물-결합 모듈(CBM), 폴리펩티드 촉매 도메인(CD) 및/또는 CBM 및 CD 둘 모두에 위치할 수 있다. 표준 20개 아미노산 "알파벳"을 그들의 측쇄의 유사성에 기초하여 화학물질 패밀리로 나누었다. 그들 패밀리에는 염기성 측쇄(예를 들어, 라이신, 아르기닌, 히스티딘), 산성 측쇄(예를 들어, 아스파르트산, 글루탐산), 하전되지 않은 극성 측쇄(예를 들어, 글리신, 아스파라긴, 글루타민, 세린, 트레오닌, 티로신, 시스테인), 비극성 측쇄(예를 들어, 알라닌, 발린, 류신, 아이소류신, 프롤린, 페닐알라닌, 메티오닌, 트립토판), 베타-분지형 측쇄((예를 들어, 트레오닌, 발린, 아이소류신) 및 방향족 측쇄(예를 들어, 티로신, 페닐알라닌, 트립토판, 히스티딘)가 있는 아미노산이 포함된다. "보존적 아미노산 치환"은 아미노산 잔기가 화학적으로 유사한 측쇄를 갖는 아미노산 잔기로 대체된 것이다(염기성 측쇄를 갖는 아미노산을 염기성 측쇄를 갖는 다른 아미노산으로 대체). "비보존적 아미노산 치환"은 아미노산 잔기가 화학적으로 상이한 측쇄를 갖는 아미노산 잔기로 대체된 것이다(즉, 염기성 측쇄를 갖는 아미노산을 방향족 측쇄를 갖는 다른 아미노산으로 대체).

키메라 폴리펩티드

본 개시내용은 또한 하나 이상의 융합 세그먼트에 부착된 본 개시내용의 단백질의 도메인을 포함하는 하이브리드/융합/키메라 단백질을 제공하며, 이는 전형적으로 단백질에 대하여 이중이다(즉, 본 개시내용의 단백질과는 상이한 공급원으로부터 유래). 하이브리드/융합/키메라 효소는 또한 고유 또는 야생형 참조 β-글루코시다제와는 상이한 다른 특성을 갖더라도, 야생형 참조 β-글루코시다제와는 서열이 다르나, β-글루코시다제 활성을 보유한다는 점에서, 돌연변이체 β-글루코시다제의 한 유형인 것으로 여겨질 수 있다. 적절한 키메라 세그먼트에는 제한 없이, 단백질 안정성을 향상시키며, 다른 바람직한 생물학적 활성 또는 바람직한 생물학적 활성 수준 증가를 제공하고/하거나, (예를 들어, 친화성 크로마토그래피에 의해) 단백질 정제를 용이하게 할 수 있는 세그먼트가 포함된다. 적절한 키메라 세그먼트는 원하는 기능(예를 들어, 향상된 안정성, 용해도, 작용, 또는 생물학적 활성을 부여하고/하거나; 단백질 정제를 단순화함)을 갖는 임의의 크기로 된 도메인일 수 있다. 본 발명의 키메라 단백질은 2개 이상의 키메라 세그먼트로 구성될 수 있으며, 각각 또는 이들 중 적어도 2개가 상이한

공급원 또는 미생물로부터 유래된다. 키메라 세그먼트는 본 개시내용의 단백질의 도메인(들)의 아미노 및/또는 카르복실 말단에 결합될 수 있다. 키메라 세그먼트는 절단에 민감할 수 있다. 이러한 민감성을 가지는 것이 유리할 수 있는데, 예를 들어 대상으로 하는 단백질을 간단하게 회수할 수 있게 된다. 키메라 단백질은 바람직하게는 단백질 또는 이의 도메인의, 카르복실 또는 아미노 말단 중 어느 하나에 부착된 키메라 세그먼트, 또는 카르복실 및 아미노 말단 둘 다에 부착된 키메라 세그먼트를 포함하는 단백질을 암호화하는 키메라 핵산으로 트랜스펙션된 재조합 세포를 배양함으로써 생성된다.

따라서, 본 개시내용의 β -글루코시다제 폴리펩티드는 또한 유전자 융합(예를 들어, 재조합 단백질의 과발현형, 가용형, 및 활성형), 돌연변이체 유전자(예를 들어, 유전자 전사 및 번역을 향상시키도록 코돈 변형된 유전자), 및 절단(truncated) 유전자(예를 들어, 신호 서열이 제거되거나 이중 신호 서열로 치환된 유전자)의 발현 산물을 포함한다.

불용성 기질을 이용하는 글리코실 하이드롤라제는 보통 모듈러 효소이다. 이들은 통상 하나 이상의 비촉매 탄수화물 결합 모듈(CBM)에 추가된 촉매 모듈을 포함한다. 사실상, CBM은 글리코실 하이드롤라제와 이의 표적 기질 다당류와의 상호작용을 촉진시키는 것으로 여겨진다. 따라서, 본 개시내용은 "스플라이스-인(spliced-in)" 이중 CBM의 결과로서 다수의 기질을 갖는 키메라 효소를 비롯하여, 기질 특이성이 변화된 키메라 효소를 제공한다. 본 개시내용의 키메라 효소의 이중 CBM은 촉매 모듈 또는 촉매 도메인(예를 들어, 활성 부위의 "CD")에 추가되도록 모듈화되게 디자인될 수 있으며, 마찬가지로 글리코실 하이드롤라제에 대하여 이중 또는 동종일 수 있다.

그리하여, 본 개시내용은 CBM/CD 모듈로 구성되거나 이를 포함하는 펩티드 및 폴리펩티드를 제공하며, 상기 모듈은 상동적으로 쌍을 이루거나 상동적으로 결합되어 키메라(이중) CBM/CD 쌍을 형성할 수 있다. 따라서, 이러한 키메라 폴리펩티드/펩티드는 대상으로 하는 효소의 성능을 향상시키거나 변경시키는데 사용될 수 있다. 따라서, 일부 태양에서, 본 개시내용은 예를 들어 이용가능한 경우, 서열 번호 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79의 효소의 적어도 하나의 CBM을 포함하는 키메라 효소를 제공한다. 본 개시내용의 폴리펩티드는 예를 들어, 서열 번호 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79의 폴리펩티드 서열의 CD 및/또는 CBM을 포함하는 아미노산 서열을 포함한다. 그리하여, 본 개시내용의 폴리펩티드는 적절하게는 2개 이상의 상이한 단백질로부터의 기능성 도메인(예를 들어, 하나의 단백질로부터의 CD에 연결된 또 하나의 단백질로부터의 CBM)을 포함하는 융합 단백질일 수 있다.

본 개시내용은 또한 적어도 2개의 β -글루코시다제 서열로 된 키메라인 β -글루코시다제 폴리펩티드를 포함하는 비천연 셀룰라제 조성물을 제공한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 상기 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함할 수 있다. 따라서, 상기 조성물은 헤미셀룰라제 조성물이다. 일부 태양에서, 비천연 셀룰라제/헤미셀룰라제 조성물은 적어도 2개의 상이한 공급원으로부터 유래되는 효소 성분 또는 폴리펩티드를 포함한다. 일부 태양에서, 비천연 셀룰라제/헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 포함한다.

일부 태양에서, 조성물 중의 β -글루코시다제 폴리펩티드는 하나 이상의 글리코실화 부위를 추가로 포함한다. 일부 태양에서, β -글루코시다제 폴리펩티드는 N-말단 서열 및 C-말단 서열을 포함하며, N-말단 서열 또는 C-말단 서열 각각은 상이한 β -글루코시다제로부터 유래되는 하나 이상의 하위서열을 포함할 수 있다. 특정 태양에서, N-말단 및 C-말단 서열은 상이한 공급원으로부터 유래된다. 일부 실시형태에서, N-말단 및 C-말단 서열의 하나 이상의 하위서열 중 적어도 2개는 상이한 공급원으로부터 유래된다. 일부 태양에서, N-말단 서열 또는 C-말단 서열 중 어느 하나는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 영역 서열을 추가로 포함한다. 특정 실시형태에서, N-말단 서열 및 C-말단 서열은 바로 인접해 있거나, 직접 연결되어 있다. 다른 실시형태에서, N-말단 및 C-말단 서열은 바로 인접해 있지 않지만, 이들은 링커 도메인을 통하여 기능적으로 연결되어 있다. 링커 도메인은 키메라 폴리펩티드의 중앙에 위치할 수 있다(예를 들어, N-말단 또는 C-말단 중 어느 하나에 위치하지 않음). 특정 실시형태에서, 하이브리드 폴리펩티드의 N-말단 서열 또는 C-말단 서열 중 어느 것도 루프 서열을 포함하지 않는다. 대신에, 링커 도메인은 루프 서열을 포함한다. 일부 태양에서, N-말단 서열은 길이가 적어도 약 200개(예를 들어, 약 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개)의 잔기로 된 β -글루코시다제 또는 그의 변이체의 제1 아미노산 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 길이가 적어도 약 50개(예를 들어, 약 50, 75, 100, 125, 150, 175, 또는

200개)의 아미노산 잔기로 된 β -글루코시다제 또는 그의 변이체의 제2 아미노산 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 일부 태양에서, C-말단 또는 N-말단 서열 중 어느 하나는 루프 서열을 포함하며, 루프 서열은 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기, 및 FDRRSPG(서열 번호 171), 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함한다. 일부 태양에서, C-말단 또는 N-말단 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, C-말단 서열 및 N-말단 서열은 루프 서열을 포함하는 링커 도메인을 통하여 연결되며, 루프 서열은 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기, 및 FDRRSPG(서열 번호 171), 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함한다. 일부 태양에서, 비천원 셀룰라제 또는 헤미셀룰라제 조성물 중의 β -글루코시다제 폴리펩티드(들)는 키메라 폴리펩티드의 C-말단 및/또는 N-말단 서열 각각이 유래되는 임의의 고유 효소에 비해 안정성이 향상된다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 30% 미만, 또는 약 20% 미만, 더욱 바람직하게는 15% 미만, 또는 10% 미만이다.

본 개시내용의 폴리펩티드는 적당히 수득될 수 있고/있거나 "실질적으로 순수한" 형태로 사용될 수 있다. 예를 들어, 본 개시내용의 폴리펩티드는 소정 조성물 중의 총 단백질의 적어도 약 80 wt.%(예를 들어, 적어도 약 85 wt.%, 90 wt.%, 91 wt.%, 92 wt.%, 93 wt.%, 94 wt.%, 95 wt.%, 96 wt.%, 97 wt.%, 98 wt.%, 또는 99 wt.%)를 구성하며, 또한 완충제 또는 용액과 같은 기타 성분을 포함한다.

발효 브로쓰

또한, 본 개시내용의 폴리펩티드는 적당히 수득될 수 있고/있거나, 발효 브로쓰(예를 들어, 사상진균 배양 브로쓰)에서 사용될 수 있다. 발효 브로쓰는 조작된 효소 조성물일 수 있으며, 예를 들어 발효 브로쓰는 대상으로 하는 이종 폴리펩티드를 발현하도록 조작된 재조합 숙주 세포, 또는 본 개시내용의 내인성 폴리펩티드를 발현하도록 조작된 재조합 숙주 세포에 의해, 내인성 발현 수준보다 크거나 적은 양으로(예를 들어, 내인성 발현 수준의 약 1-, 2-, 3-, 4-, 5배 이상 또는 내인성 발현 수준 미만인 양으로) 생성될 수 있다. 본 발명의 발효 브로쓰는 본 개시내용의 다수의 폴리펩티드를 원하는 비율로 발현하도록 조작된 특정 "통합" 숙주 세포주에 의해 생성될 수도 있다. 대상으로 하는 폴리펩티드를 암호화하는 하나 이상의 또는 모든 유전자는 예를 들어, 숙주 세포주의 유전 물질로 통합될 수 있다.

Fv3C

Fv3C의 아미노산 서열(서열 번호 60)은 도 32b 및 43에 나타나 있다. 서열 번호 60은 미성숙 Fv3C의 서열이다. Fv3C는 서열 번호 60의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 60의 위치 20 내지 899에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 32b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Fv3C 잔기 E536 및 D307은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포름(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Fv3C 폴리펩티드"는 일부 태양에서, 서열 번호 60의 잔기 20 내지 899 중에, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 또는 800개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv3C 폴리펩티드는 바람직하게는 잔기 E536 및 D307이 고유 Fv3C와 비교하여, 변경되지 않는다. Fv3C 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리의 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv3C 폴리펩티드는 적절하게는 도 32b에 나타난 고유 Fv3C의 예측된 전체 보존

도메인을 포함한다. 예시적인 Fv3C 폴리펩티드는 도 32b에 나타난 성숙 Fv3C 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv3C 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Fv3C 폴리펩티드는 적절하게는 서열 번호 60의 아미노산 서열에 대하여 또는 서열 번호 60의 잔기 (i) 20 내지 327, (ii) 22 내지 600, (iii) 20 내지 899, (iv) 428 내지 899, 또는 (v) 428 내지 660에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Fv3C 폴리펩티드"는 돌연변이체 Fv3C 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성 및/또는 안정성을 향상시키도록 Fv3C 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Fv3C 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Fv3C의 능력을 향상시키는 아미노산 치환이 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Fv3C 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Fv3C 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fv3C 폴리펩티드 CD에 존재한다. 혹은 하나 이상의 아미노산 치환은 Fv3C 폴리펩티드 CBM에 존재한다. 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재할 수 있다. 일부 태양에서, Fv3C 폴리펩티드 아미노산 치환은 아미노산 E536 및/또는 D307에서 일어날 수 있다. 일부 태양에서, Fv3C 폴리펩티드 아미노산 치환은 아미노산 D119, R125, L168, R183, K216, H217, R227, M272, Y275, D307, W308, S477, 및/또는 E536 중 하나 이상 또는 모두에서 일어날 수 있다. 돌연변이체 Fv3C 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Fv3C 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드 또는 키메라 구축물을 포함하며, 여기서 제1 서열은 제1 β -글루코시다제로부터 유래되고, 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Fv3C 서열(서열 번호 60)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 동일성을 포함하고, 제2 서열은 제2 β -글루코시다제로부터 유래되고, 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 동일성을 포함하거나 서열 번호 170의 아미노산 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 60의 적어도 약 200개의 연속 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 아미노산 서열 모티프를 포함한다.

특정 태양에서, Fv3C 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이브리드/융합 또는 키메라 구축물일 수 있으며, 여기서 제1 서열은 제1 β -글루코시다제로부터 유래되고, 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 동일성을 포함하거나, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하고, 제2 서열은 제2 β -글루코시다제로부터 유래되고, 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Fv3C 서열(서열 번호 60)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79의 적어도 200개의 연속 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 60의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 일부 실시형태에서, 제1, 제2 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또

는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Fv3C 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 하나 이상의 또는 모든 폴리펩티드 서열 모티프를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드/키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내, N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Fv3C를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실을 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실을 또는 그 손실 정도는 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, β -글루코시다제 폴리펩티드는 트리코데르마 리세이 Bgl3의 서열에 작동가능하게 연결된 Fv3C 폴리펩티드의 서열을 포함하는 키메라 또는 융합 효소이다. 특정 실시형태에서, β -글루코시다제 폴리펩티드는 Fv3C 폴리펩티드로부터 유래되는 N-말단 서열, 및 트리코데르마 리세이 Bgl3 폴리펩티드로부터 유래되는 C-말단 서열을 포함한다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함할 수 있다.

Pa3D

Pa3D의 아미노산 서열(서열 번호 54)은 도 29b 및 도 43에 나타나 있다. 서열 번호 54는 미성숙 Pa3D의 서열이다. Pa3D는 서열 번호 2의 잔기 1 내지 17에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 54의 잔기 18 내지 733에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 본 개시내용의 이러한 폴리펩티드 및 다른 폴리펩티드에 대한 신호 서열 예측은 SignalP-NN 알고리즘(www.cbs.dtu.dk)을 사용하여 행해졌다. 예측된 보존 도메인은 도 29b에서 볼드체로 되어 있다. 본 개시내용의 이러한 폴리펩티드 및 다른 폴리펩티드에 대한 도메인 예측은 Pfam, SMART 또는 NCBI 데이터베이스에 기초하여 행해졌다. Pa3D 잔기 E463 및 D262는 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니(수탁 번호 AAL69548), 트리코데르마 리세이(수탁 번호 AAP57755), 트리코데르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 다수의 GH3 패밀리의 β -글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Pa3D 폴리펩티드"는 일부 태양에서, 서열 번호 54의 잔기 18 내지 733 중에, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Pa3D 폴리

펩티드는 바람직하게는 잔기 E463 및 D262가 고유 Pa3D와 비교하여, 변경되지 않는다. Pa3D 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Pa3D 폴리펩티드는 적절하게는 도 29b에 나타난 고유 Pa3D의 예측된 전체 보존 도메인을 포함한다. 예시적인 Pa3D 폴리펩티드는 도 29b에 나타난 성숙 Pa3D 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Pa3D 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Pa3D 폴리펩티드는 적절하게는 서열 번호 54의 아미노산 서열에 대하여, 또는 서열 번호 54의 잔기 (i) 18 내지 282, (ii) 18 내지 601, (iii) 18 내지 733, (iv) 356 내지 601, 또는 (v) 356 내지 733에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

본 발명의 "Pa3D 폴리펩티드"는 또한 돌연변이체 Pa3D 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 β -글루코시다제 활성 및/또는 다른 특성을 향상시키도록 Pa3D 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Pa3D 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Pa3D의 능력을 향상시키는 아미노산 치환이 도입될 수 있다. 일부 태양에서, 돌연변이체 Pa3D 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 혹은 돌연변이체 Pa3D 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함할 수 있다. 일부 태양에서, 하나 이상의 아미노산 치환은 Pa3D 폴리펩티드 CD에 존재한다. 혹은, 하나 이상의 아미노산 치환은 Pa3D 폴리펩티드 CBM에 존재한다. 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재할 수 있다. 일부 태양에서, Pa3D 폴리펩티드 아미노산 치환은 아미노산 E463 및/또는 D262에서 일어날 수 있다. Pa3D 폴리펩티드 아미노산 치환은 아미노산 D87, R93, L136, R151, K184, H185, R195, M227, Y230, D262, W263, S406 및/또는 E463 중 하나 이상 또는 모두에서 일어날 수 있다. 돌연변이체 Pa3D 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Pa3D 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하ybrid/융합일 수 있으며, 여기서 제1 서열은 제1 β -글루코시다제로부터 유래되고, 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Pa3D 서열(서열 번호 54)에 대하여 약 60%(예를 들어, 약 60%, 65%, 70%, 75%, 또는 80%) 이상의 동일성을 포함하고, 제2 서열은 제2 β -글루코시다제로부터 유래되고, 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 70%, 75%, 80% 또는 그 이상의 동일성을 갖거나, 서열 번호 170의 아미노산 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54의 적어도 약 200개의 연속 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 아미노산 서열 모티프를 포함한다.

일부 태양에서, 본 발명의 Pa3D 폴리펩티드는 β -글루코시다제 서열로 된 키메라/하ybrid/융합 또는 키메라 구축물을 포함하며, 여기서 제1 서열은 제1 β -글루코시다제로부터 유래되고, 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%(예를 들어, 60%, 65%, 70%, 75%, 또는 80%) 이상의 동일성을 갖거나 서열 번호 164 내지 169의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하고, 제2 서열은 제2 β -글루코시다제로부터 유래되고, 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Pa3D의 서열(서열 번호 54)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 동일성을 갖는다. 예를 들어, 제1 β -글루코시다제 서열은 서열 번호 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79의 적어도 200개의 연속 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열

(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Pa3D 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 서열 모티프 중 하나 이상 또는 모두를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Pa3D를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Fv3G

Fv3G의 아미노산 서열(서열 번호 56)은 도 30b 및 43에 나타나 있다. 서열 번호 56은 미성숙 Fv3G의 서열이다. Fv3G는 서열 번호 56의 위치 1 내지 21에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 56의 위치 22 내지 780에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 상기에 기재된 바와 같이, 그것이 본 명세서의 개시내용의 다른 폴리펩티드에 대해서 행해진 것처럼 SignalP-NN 알고리즘(<http://www.cbs.dtu.dk>)으로 행해졌다. 예측된 보존 도메인은 도 30b에서 볼드체로 되어 있다. 도메인 예측은 본 명세서에서 본 발명의 다른 폴리펩티드를 사용하여 행해진 것처럼, Pfam, SMART 또는 NCBI 데이터베이스에 기초하여 행해졌다. Fv3G 잔기 E509 및 D272는 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코데르마 리세이(수탁 번호 AAP57755), 트리코데르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Fv3G 폴리펩티드"는 일부 태양에서, 서열 번호 56의 잔기 20 내지 780 중에, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 또는 750개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Fv3G 폴리펩티드는 바람직하게는 잔기 E509 및 D272가 고유 Fv3G와 비교하여, 변경되지 않는다. Fv3G 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되

지 않는다. Fv3G 폴리펩티드는 적절하게는 도 30b에 나타난 고유 Fv3G의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fv3G 폴리펩티드는 도 30b에 나타난 성숙 Fv3G 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv3G 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Fv3G 폴리펩티드는 적절하게는 서열 번호 56의 아미노산 서열에 대하여 또는 서열 번호 56의 잔기 (i) 22 내지 292, (ii) 22 내지 629, (iii) 22 내지 780, (iv) 373 내지 629, 또는 (v) 373 내지 780에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Fv3G 폴리펩티드"는 또한 돌연변이체 Fv3G 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Fv3G 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Fv3G 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Fv3G의 능력을 향상시키는 아미노산 치환이 Fv3G 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Fv3G 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Fv3G 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fv3G 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fv3G 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Fv3G 폴리펩티드 아미노산 치환은 아미노산 E509 및/또는 D272에서 일어날 수 있다. 일부 태양에서, Fv3G 폴리펩티드 아미노산 치환은 아미노산 D101, R107, L150, R165, K198, H199, R209, M237, Y240, D272, W273, S455, 및/또는 E509 중 하나 이상에서 일어날 수 있다. 돌연변이체 Fv3G 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Fv3G 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Fv3G 서열(서열 번호 56)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하고, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 56의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 모티프를 포함한다.

특정 태양에서, 본 발명의 Fv3G 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Fv3G 서열(서열 번호 56)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 56의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기

로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Fv3G 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169 중 하나 이상 또는 모두를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함할 수 있다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Fv3G를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Fv3D

Fv3D의 아미노산 서열(서열 번호 58)은 도 31b 및 43에 나타나 있다. 서열 번호 58은 미성숙 Fv3D의 서열이다. Fv3D는 서열 번호 58의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 58의 위치 20 내지 811에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 31b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Fv3D 잔기 E534 및 D301은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Fv3D 폴리펩티드"는 일부 태양에서, 서열 번호 58의 잔기 20 내지 811 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 또는 750개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Fv3D 폴리펩티드는 바람직하게는 잔기 E534 및 D301이 고유 Fv3D와 비교하여, 변경되지 않는다. Fv3D 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리의 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv3D 폴리펩티드는 적절하게는 도 31b에 나타난 고유 Fv3D의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fv3D 폴리펩티드는 도 31b에 나타난 성숙 Fv3D 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv3D 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Fv3D 폴리펩티드는 적절하게는 서열 번호 58의 아미노산 서열에 대하여 또는 서열 번호 58의

잔기 (i) 20 내지 321, (ii) 20 내지 651, (iii) 20 내지 811, (iv) 423 내지 651, 또는 (v) 423 내지 811에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Fv3D 폴리펩티드"는 또한 돌연변이체 Fv3D 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Fv3D 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Fv3D 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Fv3D의 능력을 향상시키는 아미노산 치환이 Fv3D 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Fv3D 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Fv3D 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fv3G 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fv3D 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Fv3D 폴리펩티드 아미노산 치환은 아미노산 E534 및/또는 D301에서 일어날 수 있다. 일부 태양에서, Fv3D 폴리펩티드 아미노산 치환은 아미노산 D111, R117, L160, R175, K208, H209, R219, M266, Y269, D301, W302, S472, 및/또는 E534 중 하나 이상에서 일어날 수 있다. 돌연변이체 Fv3D 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Fv3D 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Fv3D 서열(서열 번호 58)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 58의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

특정 태양에서, 본 발명의 Fv3D 폴리펩티드는 2개의 β -글루코시다제 서열로 된 하이브리드/융합/키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Fv3D 서열(서열 번호 58)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 연속 아미노산 잔기의 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 58의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Fv3D 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는

바람직하게는 서열 번호 164 내지 169의 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Fv3D를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Tr3A

Tr3A의 아미노산 서열(서열 번호 62)은 도 33b 및 43에 나타나 있다. Tr3A는 트리코테르마 리세이 Bgl1로도 알려져 있다. 서열 번호 62는 미성숙 Tr3A의 서열이다. Tr3A는 서열 번호 62의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 62의 위치 20 내지 744에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 33b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Tr3A 잔기 E472 및 D267은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오피데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Tr3A 폴리펩티드"는 일부 태양에서, 서열 번호 62의 잔기 20 내지 744 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Tr3A 폴리펩티드는 바람직하게는 잔기 E472 및 D267이 고유 Tr3A와 비교하여, 변경되지 않는다. Tr3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Tr3A 폴리펩티드는 적절하게는 도 33b에 나타난 고유 Tr3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Tr3A 폴리펩티드는 도 33b에 나타난 성숙 Tr3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Tr3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Tr3A 폴리펩티드는 적절하게는 서열 번호 62의 아미노산 서열에 대하여 또는 서열 번호 62의 잔기 (i) 20 내지 287, (ii) 22 내지 611, (iii) 20 내지 744, (iv) 362 내지 611, 또는 (v) 362 내지 744에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Tr3A 폴리펩티드"는 또한 돌연변이체 Tr3A 폴리펩티드를 지칭할 수 있다. 아미노산

치환은 분자의 β -글루코시다제 활성을 향상시키도록 Tr3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Tr3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Tr3A의 능력을 향상시키는 아미노산 치환이 Tr3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Tr3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Tr3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tr3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tr3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Tr3A 폴리펩티드 아미노산 치환은 아미노산 E472 및/또는 D267에서 일어날 수 있다. 일부 태양에서, Tr3A 폴리펩티드 아미노산 치환은 아미노산 D92, R98, L141, R156, K189, H190, R200, M232, Y235, D267, W268, S415, 및/또는 E472 중 하나 이상에서 일어날 수 있다. 돌연변이체 Tr3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Tr3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이드리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Tr3A 서열(서열 번호 62)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 64, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 62의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Tr3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Tr3A 서열(서열 번호 62)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 62의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Tr3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 서열 모티프를 포함한다.

특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Tr3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실물 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함할 수 있다.

Tr3B

Tr3B의 아미노산 서열(서열 번호 64)은 도 34b 및 43에 나타나 있다. Tr3B는 "트리코테르마 리세이 Bgl3" 또는 "트리코테르마 리세이 Cel3B"로도 알려져 있다. 서열 번호 64는 미성숙 Tr3B의 서열이다. Tr3B는 서열 번호 64의 위치 1 내지 18에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 64의 위치 19 내지 874에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 34b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Tr3B 잔기 E516 및 D287은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Tr3B 폴리펩티드"는 일부 태양에서, 서열 번호 64의 잔기 19 내지 874 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 또는 850개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Tr3B 폴리펩티드는 바람직하게는 잔기 E516 및 D287이 고유 Tr3B와 비교하여, 변경되지 않는다. Tr3B 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Tr3B 폴리펩티드는 적절하게는 도 34b에 나타난 고유 Tr3B의 예측된 전체 보존 도메인을 포함한다. 예시적인 Tr3A 폴리펩티드는 도 34b에 나타난 성숙 Tr3B 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Tr3B 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Tr3B 폴리펩티드는 적절하게는 서열 번호 64의 아미노산 서열에 대하여 또는 서열 번호 64의 잔기 (i) 19 내지 307, (ii) 19 내지 640, (iii) 19 내지 874, (iv) 407 내지 640, 또는 (v) 407 내지 874에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Tr3B 폴리펩티드"는 또한 돌연변이체 Tr3B 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Tr3B 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Tr3B 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Tr3B의 능력을 향상시키는 아미노산 치환이 Tr3B 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Tr3B 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체

Tr3B 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tr3B 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tr3B 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Tr3B 폴리펩티드 아미노산 치환은 아미노산 E516 및/또는 D287에서 일어날 수 있다. 일부 태양에서, Tr3B 폴리펩티드 아미노산 치환은 아미노산 D99, R105, L148, R163, K196, H197, R207, M252, Y255, D287, W288, S457, 및/또는 E516 중 하나 이상에서 일어날 수 있다. 돌연변이체 Tr3B 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Tr3B 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이드브리드/융합을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Tr3B 서열(서열 번호 64)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 66, 68, 70, 72, 74, 76, 78 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 64의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Tr3B 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169 중 하나 이상의 폴리펩티드 서열 모티프를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Tr3B 서열(서열 번호 64)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 64의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Tr3B 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 서열 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이드브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Tr3B를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Te3A

Te3A의 아미노산 서열(서열 번호 66)은 도 35b 및 43에 나타나 있다. Te3A는 "Abg2"로도 알려져 있다. 서열 번호 66은 미성숙 Te3A의 서열이다. Te3A는 서열 번호 66의 위치 1 내지 19에 해당하는 예측된 신호 서열(밀줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 66의 위치 20 내지 857에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 35b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Te3A 잔기 E505 및 D277은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 탈리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니케르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Te3A 폴리펩티드"는 일부 태양에서, 서열 번호 66의 잔기 20 내지 857 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 또는 800개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Te3A 폴리펩티드는 바람직하게는 잔기 E505 및 D277이 고유 Te3A와 비교하여, 변경되지 않는다. Te3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Te3A 폴리펩티드는 적절하게는 도 35b에 나타난 고유 Te3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Te3A 폴리펩티드는 도 35b에 나타난 성숙 Te3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Te3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Te3A 폴리펩티드는 적절하게는 서열 번호 66의 아미노산 서열에 대하여 또는 서열 번호 66의 잔기 (i) 20 내지 297, (ii) 20 내지 629, (iii) 20 내지 857, (iv) 396 내지 629, 또는 (v) 396 내지 857에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Te3A 폴리펩티드"는 또한 돌연변이체 Te3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Te3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Te3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Te3A의 능력을 향상시키는 아미노산 치환이 Te3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Te3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Te3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Te3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Te3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Te3A 폴

리펩티드 아미노산 치환은 아미노산 E505 및/또는 D277에서 일어날 수 있다. 일부 태양에서, Te3A 폴리펩티드 아미노산 치환은 아미노산 D92, R98, L141, R156, K189, H190, R200, M242, Y245, D277, W278, S447, 및/또는 E505 중 하나 이상에서 일어날 수 있다. 돌연변이체 Te3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Te3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Te3A 서열(서열 번호 66)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 66의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Te3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이브리드/융합 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Te3A 서열(서열 번호 66)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 66의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Te3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이

유래된 Te3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

An3A

An3A의 아미노산 서열(서열 번호 68)은 도 36b 및 43에 나타나 있다. An3A는 "아스페르길루스 니게르 Bglu"로도 알려져 있다. 서열 번호 68은 미성숙 An3A의 서열이다. An3A는 서열 번호 68의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해, 서열 번호 68의 위치 20 내지 860에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 36b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터 베이스에 기초하여 행해졌다. An3A 잔기 E509 및 D277은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 헤마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오티데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "An3A 폴리펩티드"는 일부 태양에서, 서열 번호 68의 잔기 20 내지 860 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 또는 800개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. An3A 폴리펩티드는 바람직하게는 잔기 E509 및 D277이 고유 An3A와 비교하여, 변경되지 않는다. An3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀 리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. An3A 폴리펩티드는 적절하게는 도 36b에 나타난 고유 An3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 An3A 폴리펩티드는 도 36b에 나타난 성숙 An3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 An3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 An3A 폴리펩티드는 적절하게는 서열 번호 68의 아미노산 서열에 대하여 또는 서열 번호 68의 잔기 (i) 20 내지 300, (ii) 20 내지 634, (iii) 20 내지 860, (iv) 400 내지 634, 또는 (v) 400 내지 860에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "An3A 폴리펩티드"는 또한 돌연변이체 An3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 An3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 An3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 An3A의 능력을 향상시키는 아미노산 치환이 An3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 An3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 An3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 An3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 An3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, An3A 폴리펩티드 아미노산 치환은 아미노산 E509 및/또는 D277에서 일어날 수 있다. 일부 태양에서, An3A 폴리펩티드 아미노산 치환은 아미노산 D92, R98, L141, R156, K189, H190, R200, M245, Y248, D277, W278, S451, 및/또는

E509 중 하나 이상에서 일어날 수 있다. 돌연변이체 An3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, An3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이브리드/융합을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 An3A 서열(서열 번호 68)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 68의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 An3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 An3A 서열(서열 번호 68)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 68의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 An3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 An3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실을 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직

하계는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Fo3A

Fo3A의 아미노산 서열(서열 번호 70)은 도 37b 및 43에 나타나 있다. 서열 번호 70은 미성숙 Fo3A의 서열이다. Fo3A는 서열 번호 70의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 70의 위치 20 내지 899에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 37b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Fo3A 잔기 E536 및 D307은 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포름(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Fo3A 폴리펩티드"는 일부 태양에서, 서열 번호 70의 잔기 20 내지 899 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 또는 850개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fo3A 폴리펩티드는 바람직하게는 잔기 E536 및 D307이 고유 Fo3A와 비교하여, 변경되지 않는다. Fo3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fo3A 폴리펩티드는 적절하게는 도 37b에 나타난 고유 Fo3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fo3A 폴리펩티드는 도 37b에 나타난 성숙 Fo3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fo3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Fo3A 폴리펩티드는 적절하게는 서열 번호 70의 아미노산 서열에 대하여 또는 서열 번호 70의 잔기 (i) 20 내지 327, (ii) 20 내지 660, (iii) 20 내지 899, (iv) 428 내지 660, 또는 (v) 428 내지 899에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Fo3A 폴리펩티드"는 또한 돌연변이체 Fo3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Fo3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Fo3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Fo3A의 능력을 향상시키는 아미노산 치환이 Fo3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Fo3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Fo3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fo3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Fo3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Fo3A 폴리펩티드 아미노산 치환은 아미노산 E536 및/또는 D307에서 일어날 수 있다. 일부 태양에서, Fo3A 폴리펩티드 아미노산 치환은 아미노산 D119, R125, L168, R183, K216, H217, R227, M272, Y275, D307, W308, S477, 및/또는 E536 중 하나 이상에서 일어날 수 있다. 돌연변이체 Fo3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Fo3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이브리드/융합을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Fo3A 서

열(서열 번호 70)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 70의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Fo3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Fo3A 서열(서열 번호 70)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 72, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 70의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Fo3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Fo3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실을 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정

실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Gz3A

Gz3A의 아미노산 서열(서열 번호 72)은 도 38b 및 43에 나타나 있다. 서열 번호 72는 미성숙 Gz3A의 서열이다. Gz3A는 서열 번호 72의 위치 1 내지 18에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 72의 위치 19 내지 886에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 38b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Gz3A 잔기 E523 및 D294는 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니케르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Gz3A 폴리펩티드"는 일부 태양에서, 서열 번호 72의 잔기 19 내지 886 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 또는 850개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Gz3A 폴리펩티드는 바람직하게는 잔기 E536 및 D307이 고유 Gz3A와 비교하여, 변경되지 않는다. Gz3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Gz3A 폴리펩티드는 적절하게는 도 38b에 나타난 고유 Gz3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Gz3A 폴리펩티드는 도 38b에 나타난 성숙 Gz3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Gz3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Gz3A 폴리펩티드는 적절하게는 서열 번호 72의 아미노산 서열에 대하여 또는 서열 번호 72의 잔기 (i) 19 내지 314, (ii) 19 내지 647, (iii) 19 내지 886, (iv) 415 내지 647, 또는 (v) 415 내지 886에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Gz3A 폴리펩티드"는 또한 돌연변이체 Gz3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Gz3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Gz3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Gz3A의 능력을 향상시키는 아미노산 치환이 Gz3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Gz3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Gz3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Gz3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Gz3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Gz3A 폴리펩티드 아미노산 치환은 아미노산 E536 및/또는 D307에서 일어날 수 있다. 일부 태양에서, Gz3A 폴리펩티드 아미노산 치환은 아미노산 D106, R112, L155, R170, K203, H204, R214, M259, Y262, D294, W295, S464, 및/또는 E523 중 하나 이상에서 일어날 수 있다. 돌연변이체 Gz3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Gz3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Gz3A 서열(서열 번호 72)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 72의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며,

제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 74, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Gz3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Gz3A 서열(서열 번호 72)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 74, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 72의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Gz3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 바람직하게는 서열 번호 164 내지 169의 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 서열 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Gz3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Nh3A

Nh3A의 아미노산 서열(서열 번호 74)은 도 39b 및 43에 나타나 있다. 서열 번호 74는 미성숙 Nh3A의 서열이다. Nh3A는 서열 번호 74의 위치 1 내지 19에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 74의 위치 20 내지 880에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 39b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Nh3A 잔기 E523 및 D294는 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포름(수탁 번호 BGL FOXG_02349), 아스페르길루스 니케르(수탁 번호 CAK48740), 탈라로마이세스 에베르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Nh3A 폴리펩티드"는 일부 태양에서, 서열 번호 74의 잔기 20 내지 880 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800 또는 850개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Nh3A 폴리펩티드는 바람직하게는 잔기 E523 및 D294가 고유 Nh3A와 비교하여, 변경되지 않는다. Nh3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Nh3A 폴리펩티드는 적절하게는 도 39b에 나타난 고유 Nh3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Nh3A 폴리펩티드는 도 39b에 나타난 성숙 Nh3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Nh3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Nh3A 폴리펩티드는 적절하게는 서열 번호 74의 아미노산 서열에 대하여 또는 서열 번호 74의 잔기 (i) 20 내지 295, (ii) 20 내지 647, (iii) 20 내지 880, (iv) 414 내지 647, 또는 (v) 414 내지 880에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Nh3A 폴리펩티드"는 또한 돌연변이체 Nh3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Nh3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Nh3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Nh3A의 능력을 향상시키는 아미노산 치환이 Nh3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Nh3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Nh3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Nh3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Nh3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Nh3A 폴리펩티드 아미노산 치환은 아미노산 E523 및/또는 D294에서 일어날 수 있다. 일부 태양에서, Nh3A 폴리펩티드 아미노산 치환은 아미노산 D106, R112, L155, R170, K203, H204, R214, M259, Y262, D294, W295, S464, 및/또는 E523 중 하나 이상에서 일어날 수 있다. 돌연변이체 Nh3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Nh3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Nh3A 서열(서열 번호 74)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 74의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 76, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Nh3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을

포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Nh3A 서열(서열 번호 74)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 76, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 74의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Nh3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 서열 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 서열 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Nh3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실 정도 또는 그 손실률 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Vd3A

Vd3A의 아미노산 서열(서열 번호 76)은 도 40b 및 43에 나타나 있다. 서열 번호 76은 미성숙 Vd3A의 서열이다. Vd3A는 서열 번호 76의 위치 1 내지 18에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 76의 위치 19 내지 890에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다.

신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 40b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Vd3A는 예를 들어, cNPG 및 셀로비 오스를 사용하는 효소적 검정법에서, 그리고 기질로서 회색 암모니아로 전처리된 옥수수 속대의 가수분해에서 β -글루코시다제 활성을 갖는 것으로 나타났다. Vd3A 잔기 E524 및 D295는 각각, 예를 들어 포도스포라 안세리 나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소나이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타 나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Vd3A 폴리펩티드"는 일부 태양에서, 서열 번호 76의 잔기 19 내지 890 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 또는 850개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Vd3A 폴리펩티드는 바람직하게는 잔기 E524 및 D295가 고유 Vd3A와 비교하여, 변경되지 않는다. Vd3A 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Vd3A 폴리펩티드는 적절하게는 도 40b에 나타난 고유 Vd3A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Nh3A 폴리펩티드는 도 40b에 나타난 성숙 Vd3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Vd3A 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Vd3A 폴리펩티드는 적절하게는 서열 번호 76의 아미노산 서열에 대하여 또는 서열 번호 76의 잔기 (i) 19 내지 296, (ii) 19 내지 649, (iii) 19 내지 890, (iv) 415 내지 649, 또는 (v) 415 내지 890에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Vd3A 폴리펩티드"는 또한 돌연변이체 Vd3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Vd3A 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Vd3A 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Vd3A의 능력을 향상시키는 아미노산 치환이 Vd3A 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Vd3A 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Vd3A 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Vd3A 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Vd3A 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Vd3A 폴리펩티드 아미노산 치환은 아미노산 E524 및/또는 D295에서 일어날 수 있다. 일부 태양에서, Vd3A 폴리펩티드 아미노산 치환은 아미노산 D107, R113, L156, R171, K204, H205, R215, M260, Y263, D295, W296, S465, 및/또는 E524 중 하나 이상에서 일어날 수 있다. 돌연변이체 Vd3A 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Vd3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/하이드리드/융합을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Vd3A 서열(서열 번호 76)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 76의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 78, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Vd3A 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약

60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Vd3A 서열(서열 번호 76)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 78, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 76의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Vd3A 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 서열 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Vd3A를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 동안 관련된 효소 활성 손실을 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Pa3G

Pa3G의 아미노산 서열(서열 번호 78)은 도 41b 및 43에 나타나 있다. 서열 번호 78은 미성숙 Pa3G의 서열이다. Pa3G는 서열 번호 78의 위치 1 내지 19에 해당하는 예측된 신호 서열(밀줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 78의 위치 20 내지 805에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 신호 서열 예측은 SignalP-NN 알고리즘으로 행해졌다. 예측된 보존 도메인은 도 41b에서 볼드체로 되어 있다. 도메인 예측은 Pfam, SMART, 또는 NCBI 데이터베이스에 기초하여 행해졌다. Pa3G 잔기 E517 및 D289는 각각,

예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 넥트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Pa3G 폴리펩티드"는 일부 태양에서, 서열 번호 78의 잔기 20 내지 805 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 또는 750개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Pa3G 폴리펩티드는 바람직하게는 잔기 E517 및 D289가 고유 Pa3G와 비교하여, 변경되지 않는다. Pa3G 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Pa3G 폴리펩티드는 적절하게는 도 41b에 나타난 고유 Pa3G의 예측된 전체 보존 도메인을 포함한다. 예시적인 Pa3G 폴리펩티드는 도 41b에 나타난 성숙 Pa3G 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Pa3G 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Pa3G 폴리펩티드는 적절하게는 서열 번호 78의 아미노산 서열에 대하여 또는 서열 번호 78의 잔기 (i) 20 내지 354, (ii) 20 내지 660, (iii) 20 내지 805, (iv) 449 내지 660, 또는 (v) 449 내지 805에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Pa3G 폴리펩티드"는 또한 돌연변이체 Vd3A 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Pa3G 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Pa3G 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 그의 능력을 향상시키는 아미노산 치환이 Pa3G 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Pa3G 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Pa3G 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Pa3G 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Pa3G 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Pa3G 폴리펩티드 아미노산 치환은 아미노산 E517 및/또는 D289에서 일어날 수 있다. 일부 태양에서, Pa3G 폴리펩티드 아미노산 치환은 아미노산 D101, R107, L150, R165, K199, H209, R215, M254, Y257, D289, W290, S458, 및/또는 E517 중 하나 이상에서 일어날 수 있다. 돌연변이체 Pa3G 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Pa3G 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Pa3G 서열(서열 번호 78)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 78의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 79 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Pa3G 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Pa3G 서열(서열 번호 78)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58,

60, 62, 64, 66, 68, 70, 72, 74, 76, 및 79 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 78의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Pa3G 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Pa3G를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

Tn3B

Tn3B의 아미노산 서열(서열 번호 79)은 도 42 및 43에 나타나 있다. 서열 번호 79는 미성숙 Tn3B의 서열이다. SignalP-NN 알고리즘(<http://www.cbs.dtu.dk>)에 의해, 예측된 신호 서열이 제공되지 않았다. Tn3B 잔기 E458 및 D242는 각각, 예를 들어 포도스포라 안세리나(수탁 번호 XP_001912683), 버티실리움 달리아에, 벡트리아 해마토코카(수탁 번호 XP_003045443), 지베렐라 제아에(수탁 번호 XP_386781), 푸사리움 옥시스포룸(수탁 번호 BGL FOXG_02349), 아스페르길루스 니게르(수탁 번호 CAK48740), 탈라로마이세스 에메르소니이(수탁 번호 AAL69548), 트리코테르마 리세이(수탁 번호 AAP57755), 트리코테르마 리세이(수탁 번호 AAA18473), 푸사리움 베르티실리오이데스, 및 씨모토가 네아폴리타나(수탁 번호 Q0GC07) 등으로부터의 상기 언급된 GH3 글루코시다제의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다(도 43 참조). 본 명세서에 사용되는 "Tn3B 폴리펩티드"는 일부 태양에서, 서열 번호 79의 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300,

350, 400, 450, 500, 550, 600, 650, 700, 또는 750개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Tn3B 폴리펩티드는 바람직하게는 잔기 E458 및 D242가 고유 Tn3B와 비교하여, 변경되지 않는다. Tn3B 폴리펩티드는 바람직하게는 도 43의 정렬에 나타난 바와 같이, 본 명세서에 기재된 GH3 패밀리 β -글루코시다제 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Tn3B 폴리펩티드는 적절하게는 도 43에 나타난 고유 Tn3B의 예측된 전체 보존 도메인을 포함한다. 예시적인 Tn3B 폴리펩티드는 도 42에 나타난 성숙 Tn3B 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Tn3B 폴리펩티드는 바람직하게는 β -글루코시다제 활성을 갖는다.

따라서, 본 발명의 Tn3B 폴리펩티드는 적절하게는 서열 번호 79의 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, 본 발명의 "Tn3B 폴리펩티드"는 또한 돌연변이체 Tn3B 폴리펩티드를 지칭할 수 있다. 아미노산 치환은 분자의 β -글루코시다제 활성을 향상시키도록 Tn3B 폴리펩티드에 도입될 수 있다. 예를 들어, 그 기질에 대한 Tn3B 폴리펩티드의 결합 친화성을 증가시키거나, β -D-글루코시드에서의 말단 비환원성 잔기의 가수분해를 촉매하는 Tn3B의 능력을 향상시키는 아미노산 치환이 Tn3B 폴리펩티드에 도입될 수 있다. 일부 태양에서, 돌연변이체 Tn3B 폴리펩티드는 하나 이상의 보존적 아미노산 치환을 포함한다. 일부 태양에서, 돌연변이체 Tn3B 폴리펩티드는 하나 이상의 비보존적 아미노산 치환을 포함한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tn3B 폴리펩티드 CD에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 Tn3B 폴리펩티드 CBM에 존재한다. 일부 태양에서, 하나 이상의 아미노산 치환은 CD 및 CBM 둘 다에 존재한다. 일부 태양에서, Tn3B 폴리펩티드 아미노산 치환은 아미노산 E458 및/또는 D242에서 일어날 수 있다. 일부 태양에서, Tn3B 폴리펩티드 아미노산 치환은 아미노산 D58, R64, L116, R130, K163, H164, R174, M207, Y210, D242, W243, S370, 및/또는 E458 중 하나 이상에서 일어날 수 있다. 돌연변이체 Tn3B 폴리펩티드(들)는 적절하게는 β -글루코시다제 활성을 갖는다.

일부 태양에서, Tn3B 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라/융합/하이브리드를 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Tn3B 서열(서열 번호 79)에 대하여 약 60%, 65%, 70%, 75%, 또는 80% 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 78 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 79의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하며, 제2 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 78 중 어느 하나의 적어도 약 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다.

특정 태양에서, 본 발명의 Tn3B 폴리펩티드는 2개의 β -글루코시다제 서열로 된 키메라 또는 키메라 구축물을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 78 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 Tn3B 서열(서열 번호 79)에 대하여 약 60%, 65%, 70%, 75%, 80% 또는 그 이상의 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 및 78 중 어느 하나의 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 서열 번호 79의 적어도 50개의 연속 아미노산 잔기로 된 C-말단 서열을 포함한다.

일부 태양에서, 제1 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 N-말단에 위치하는 한편, 제2 β -글루코시다제 서열은 키메라 β -글루코시다제 폴리펩티드의 C-말단에 위치한다. 특정 실시형태에서, 제1, 제2, 또는 둘 모두의 β -글루코시다제 서열은 하나 이상의 글리코실화 부위를 추가로 포함한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 서로 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일

부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 영역, 또는 루프-유사 구조를 나타내는 서열을 포함한다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않는다. 일부 실시형태에서, 링커 도메인은 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기를 포함하는 루프 영역을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 연결하는 링커 도메인은 중앙에 위치한다(즉, 키메라 폴리펩티드의 N- 또는 C-말단에 위치하지 않음). 일부 태양에서, 키메라 β -글루코시다제의 N-말단 서열은 Tn3B 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 200, 250, 300, 350, 400, 450, 500, 550, 또는 600개의 잔기로 된 서열을 포함한다. 일부 태양에서, N-말단 서열은 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 164 내지 169의 모티프를 포함한다. 일부 태양에서, C-말단 서열은 β -글루코시다제 폴리펩티드 또는 그의 변이체로부터 유래되는 길이가 적어도 50, 75, 100, 125, 150, 175, 또는 200개의 아미노산 잔기로 된 서열을 포함한다. 일부 태양에서, C-말단 서열은 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두, 또는 바람직하게는 서열 번호 170의 모티프를 포함한다. 특정 실시형태에서, β -글루코시다제 폴리펩티드, 그의 변이체, 또는 그의 하이브리드 또는 키메라는 하나 이상의 글리코실화 부위를 추가로 포함한다. 하나 이상의 글리코실화 부위는 C-말단 서열 내 또는 N-말단 서열 내, 또는 두 서열 내에 위치될 수 있다.

일부 태양에서, 본 발명의 비천연 셀룰라제 또는 헤미셀룰라제 조성물은 하나 이상의 천연 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 키메라 β -글루코시다제의 C-말단 또는 N-말단 서열이 유래된 Tn3B를 비롯한 고유 효소에 비해, 향상된 안정성을 갖는다. 일부 태양에서, 향상된 안정성은 저장, 발현 또는 생성 공정 동안 단백질 가수분해 안정성 향상을 포함한다. 일부 태양에서, 향상된 안정성은 저장 또는 생성 조건 동안 관련된 효소 활성 손실률 또는 그 손실 정도 감소를 포함하며, 여기서 효소 활성 손실은 바람직하게는 약 50% 미만, 약 40% 미만, 약 20% 미만, 더욱 바람직하게는 약 15% 미만, 또는 더욱더 바람직하게는 약 10% 미만이다. 일부 태양에서, N-말단 서열 또는 C-말단 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되는 루프 서열을 포함할 수 있다. N-말단 및 C-말단 서열은 서로 바로 인접해 있거나 서로 직접 연결될 수 있다. 다른 태양에서, N-말단 서열 및 C-말단 서열은 링커 도메인을 통하여 연결될 수 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 β -글루코시다제 활성을 포함한다. 일부 태양에서, 비천연 셀룰라제 조성물은 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성 중 하나 이상을 추가로 포함한다.

핵산

예시적인 β -글루코시다제 핵산에는 β -글루코시다제 폴리펩티드의 적어도 하나의 활성을 갖는 폴리펩티드, 폴리펩티드 단편, 펩티드, 또는 융합 폴리펩티드를 암호화하는 핵산이 포함된다. 예시적인 β -글루코시다제 폴리펩티드 및 핵산에는 본 명세서에 기재된 임의의 공급원 유기체로부터의 천연 폴리펩티드 및 핵산, 및 본 명세서에 기재된 임의의 공급원 유기체로부터의 돌연변이체 폴리펩티드 및 핵산이 포함된다. 예시적인 β -글루코시다제 핵산에는 예를 들어, 제한 없이 하나 이상의 하기 유기체로부터 단리된 β -글루코시다제가 포함된다: 크리니펠리스 스카펠라, 마크로포미나 파세올리나, 마이셀리오프토라 씨모필라, 소르다리아 피미콜라, 볼루텔라 콜레토티코이데스, 티엘라비아 테레스트리스, 아크레모니움 *sp.*, 엑시디아 글란돌로사, 포메스 포멘타리우스, 스폰기펠리스 *sp.*, 리조플릭티스 로세아, 리조무코르 푸실루스, 파이코마이세스 니테우스, 카에토스틸룸 프레세니이, 디플로디아 고사이피나, 울로스포라 빌그라미이, 사코볼루스 딜루텔루스, 페니실리움 베루콜로숨, 페니실리움 크라이소게눔, 씨모마이세스 베루코수스, 디아포르테 사인게네시아, 콜레토티리움 라게나리움, 니그로스포라 *sp.*, 자일라리아 하이폭실론, 넥트리아 피네아, 소르다리아 마크로스포라, 티엘라비아 씨모필라, 카에토미움 모로룸, 카에토미움 비르센스, 카에토미움 브라실리엔시스, 카에토미움 쿠니콜로룸, 사이스파스토스포라 보니넨시스, 클라도리눔 포에쿤디시움, 사이탈리디움 씨모필라, 글리오콜라디움 카테놀라툼, 푸사리움 옥시스포룸 *ssp.* 라이코페르시시, 푸사리움 옥시스포룸 *ssp.* 파시플로라, 푸사리움 솔라니, 푸사리움 안구이오이데스, 푸사리움 포아에, 후미콜라 니그레센스, 후미콜라 그리세아, 파나에올루스 레티루기스, 트라메테스 상귀네아, 스킴조필룸 코뮤네, 트리코테움 로세움, 마이크로스페롭시스 *sp.*, 악소볼루스 스틱토이데우스 *spej.*, 포로니아 푼크타타, 노둘리스포룸 *sp.*, 트리코테르마 *sp.*(예를 들어, 트리코테르마 리세이) 및 실린드로카르폰 *sp.*

본 개시내용은 적어도 약 10개, 예를 들어 적어도 약 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200,

250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 또는 2000개의 뉴클레오타이드의 영역에 걸쳐, 서열 번호 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 46, 47, 48, 49, 50, 51, 53, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 또는 77의 핵산에 대하여 서열에 대하여 적어도 약 70%, 예를 들어 적어도 약 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%, 또는 완전한(100%) 서열 동일성을 갖는 핵산 서열을 포함하는 단리된, 합성, 또는 재조합 핵산을 제공한다. 본 개시내용은 또한 헤미셀룰로스 분해 활성(예를 들어, 자일라나제, β -자일로시다제, 및/또는 L- α -아라비노푸라노시다제 활성)을 갖는 적어도 하나의 폴리펩티드를 암호화하는 핵산을 제공한다. 게다가, 본 개시내용은 셀룰로스 분해 활성(예를 들어, β -글루코시다제 활성, 또는 엔도글루카나제 활성)을 갖는 폴리펩티드를 암호화하는 핵산을 제공한다.

본 개시내용의 핵산은 또한 서열 번호 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79의 서열을 포함하는 효소 또는 효소의 성숙 부위, 또는 폴리펩티드 서열 모티프를 포함하는 GH61 엔도글루카나제 효소 또는 그러한 효소의 성숙 부위를 암호화하는 단리된, 합성, 또는 재조합 핵산을 포함한다: (1) 서열 번호 84 및 88; (2) 서열 번호 85 및 88; (3) 서열 번호 86; (4) 서열 번호 87; (5) 서열 번호 84, 88 및 89; (6) 서열 번호 85, 88 및 89; (7) 서열 번호 84, 88 및 90; (8) 서열 번호 85, 88 및 90; (9) 서열 번호 84, 88 및 91; (10) 서열 번호 85, 88 및 91; (11) 서열 번호 84, 88, 89 및 91; (12) 서열 번호 84, 88, 90 및 91; (13) 서열 번호 85, 88, 89 및 91; 및 (14) 서열 번호 85, 88, 90 및 91, 및 그들의 하위서열(예를 들어, 보존 도메인 또는 탄수화물 결합 도메인 ("CBM"), 및 그들의 변이체.

본 개시내용은 구체적으로 Fv3A, Pf43A, Fv43E, Fv39A, Fv43A, Fv43B, Pa51A, Gz43A, Fo43A, Af43A, Pf51A, AfuXyn2, AfuXyn5, Fv43D, Pf43B, Fv43B, Fv51A, 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Xyn2, 트리코테르마 리세이 Bx11, 트리코테르마 리세이 Bgl1(Tr3A), 트리코테르마 리세이 Eg4, 트리코테르마 리세이 Bgl3(Tr3B), Pa3D, Fv3G, Fv3D, Fv3C, Te3A, An3A, Fo3A, Gz3A, Nh3A, Vd3A, Pa3G 또는 Tn3B 폴리펩티드, 그의 변이체, 돌연변이체, 또는 하이브리드 또는 키메라 폴리펩티드를 암호화하는 핵산을 제공한다. 일부 태양에서, 본 개시내용은 예를 들어, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열을 포함하는 키메라 또는 융합 효소를 암호화하는 핵산을 제공하며, 여기서 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 상이한 유기체로부터 유래된다. 특정 태양에서, 제1 β -글루코시다제 서열은 N-말단에 존재하고, 제2 β -글루코시다제는 하이브리드 또는 키메라 β -글루코시다제 폴리펩티드의 C-말단에 존재한다. 특정 태양에서, 제1 β -글루코시다제 서열, 또는 더욱 구체적으로는 제1 β -글루코시다제 서열의 C-말단은 제2 β -글루코시다제 서열, 또는 더욱 구체적으로는 제2 β -글루코시다제 서열의 N-말단에 직접 인접하거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제는 직접 인접하거나 연결되어 있지 않지만, 제1 β -글루코시다제 서열은 링커 서열 또는 도메인을 통하여 제2 β -글루코시다제 서열에 작동가능하게 결합되거나 연결되어 있다. 일부 예에서, 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 136 내지 148로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 149 내지 156으로 나타내는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함한다. 특히, 2개 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하며, 2개 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 서로 직접 연결되거나 바로 인접해 있다. 일부 태양에서, 제1 β -글루코시다제 서열은 제2 β -글루코시다제 서열에 직접 연결되거나 바로 인접해 있지 않지만, 제1 및 제2 β -글루코시다제는 링커 서열을 통하여 연결되어 있다. 특정 실시형태에서, 링커 서열은 중앙에 위치한다. 특정한 구체예에서, 제1 β -글루코시다제 서열은 Fv3C 폴리펩티드의 서열, 예를 들어 길이가 적어도 200개의 아미노산 잔기로 된 N-말단 서열을 포함한다. 일부 실시형태에서, 제2 β -글루코시다제 서열은 트리코테르마 리세이 Bgl3 폴리펩티드의 서열, 예를 들어 길이가 적어도 50개의 아미노산 잔기로 된 C-말단 서열을 포함한다. 특정예에서, β -글루코시다제 폴리펩티드는 하이브리드 또는 키메라 Fv3C 폴리펩티드, 또는 트리코테르마 리세이 Bgl3(Tr3B) 폴리펩티드이며, 서열 번호 159의 아미노산 서열을 포함한다. 또 하나의 예에서, β -글루코시다제 폴리펩티드는 제3 β -글루코시다제 폴리펩티드 서열로부터 유래되는 링커 서열을 임의로 포함하는 하이브리드 또는 키메라 Fv3C 폴리펩티드, 또는 트리코테르마 리세이 Bgl3 폴리펩티드이며, 여기서

β -글루코시다제 폴리펩티드는 서열 번호 135의 아미노산 서열을 포함한다. 키메라 또는 융합 효소는 일부 태양에서, 적절하게는 링커 서열을 포함하며, 따라서, 본 개시내용은 키메라 효소를 암호화하는 핵산을 제공하고, 그의 N-말단 서열, C-말단 서열, 또는 하위서열 중 어느 하나가 유래되는 β -글루코시다제 폴리펩티드인 것으로 간주될 수 있다. 예를 들어, 하이브리드 Fv3C/Bg13 폴리펩티드는 Fv3C 폴리펩티드, 그의 변이체, 트리코테르마 리세이 Bg13 폴리펩티드, 그의 변이체, 또는 키메라 Fv3C/Bg13 폴리펩티드 또는 그의 변이체인 것으로 간주될 수 있다. 또 하나의 예에서, 하이브리드 Fv3C/Te3A/Bg13 폴리펩티드는 Fv3C 폴리펩티드 또는 그의 변이체, 트리코테르마 리세이 Bg13 폴리펩티드 또는 그의 변이체, Te3A 폴리펩티드 또는 그의 변이체, 또는 키메라 Fv3C/Te3A/Bg13/ 폴리펩티드 또는 그의 변이체인 것으로 간주될 수 있다.

폴리뉴클레오티드 서열과 관련하여 사용되는 경우, 용어 "변이체"는 유전자의 서열 또는 그의 암호화 서열과 관련된 폴리뉴클레오티드 서열을 포함할 수 있다. 또한, 이러한 정의는 예를 들어, "대립 형질", "스플라이스", "중" 또는 "다형성" 변이체를 포함할 수 있다. 스플라이스 변이체는 참조 폴리뉴클레오티드에 대하여 상당한 동일성을 가질 수 있으나, mRNA 가공 동안의 엑손의 선택적 스플라이싱 때문에, 일반적으로 더 많거나 더 적은 잔기를 가질 것이다. 해당하는 폴리펩티드는 추가의 기능성 도메인을 가지거나 도메인이 존재하지 않을 수 있다. 중 변이체는 종마다 다른 폴리뉴클레오티드 서열이다. 얻어진 폴리펩티드는 더욱 상술한 바와 같이, 일반적으로 서로에 대하여 상당한 아미노산 동일성을 가질 것이다. 다형성 변이체는 주어진 종의 개체 간의 특정 유전자의 폴리뉴클레오티드 서열의 변이이다.

예를 들어, 본 개시내용은 단리된 핵산 분자를 제공하며, 여기서, 핵산 분자는 하기의 것을 암호화한다:

- (1) 서열 번호 54의 아미노산 서열에 대하여 또는 서열 번호 54의 잔기 (i) 18 내지 282, (ii) 18 내지 601, (iii) 18 내지 733, (iv) 356 내지 601 또는 (v) 356 내지 733에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (2) 서열 번호 56의 아미노산 서열에 대하여 또는 서열 번호 56의 잔기 (i) 22 내지 292, (ii) 22 내지 629, (iii) 22 내지 780, (iv) 373 내지 629 또는 (v) 373 내지 780에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (3) 서열 번호 58의 아미노산 서열에 대하여 또는 서열 번호 58의 잔기 (i) 20 내지 321, (ii) 20 내지 651, (iii) 20 내지 811, (iv) 423 내지 651 또는 (v) 423 내지 811에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (4) 서열 번호 60의 아미노산 서열에 대하여 또는 서열 번호 60의 잔기 (i) 20 내지 327, (ii) 22 내지 600, (iii) 20 내지 899, (iv) 428 내지 899 또는 (v) 428 내지 660에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (5) 서열 번호 62의 아미노산 서열에 대하여 또는 서열 번호 62의 잔기 (i) 20 내지 287, (ii) 22 내지 611, (iii) 20 내지 744, (iv) 362 내지 611 또는 (v) 362 내지 744에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (6) 서열 번호 64의 아미노산 서열에 대하여 또는 서열 번호 64의 잔기 (i) 19 내지 307, (ii) 19 내지 640, (iii) 19 내지 874, (iv) 407 내지 640 또는 (v) 407 내지 874에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (7) 서열 번호 66의 아미노산 서열에 대하여 또는 서열 번호 66의 잔기 (i) 20 내지 297, (ii) 20 내지 629, (iii) 20 내지 857, (iv) 396 내지 629 또는 (v) 396 내지 857에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (8) 서열 번호 68의 아미노산 서열에 대하여 또는 서열 번호 68의 잔기 (i) 20 내지 300, (ii) 20 내지 634, (iii) 20 내지 860, (iv) 400 내지 634 또는 (v) 400 내지 860에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (9) 서열 번호 70의 아미노산 서열에 대하여 또는 서열 번호 70의 잔기 (i) 20 내지 327, (ii) 20 내지 660, (iii) 20 내지 899, (iv) 428 내지 660 또는 (v) 428 내지 899에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;
- (10) 서열 번호 72의 아미노산 서열에 대하여 또는 서열 번호 72의 잔기 (i) 19 내지 314, (ii) 19 내지 647, (iii) 19 내지 886, (iv) 415 내지 647 또는 (v) 415 내지 886에 대하여 적어도 80%, 85%, 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;

(11) 서열 번호 74의 아미노산 서열에 대하여 또는 서열 번호 74의 잔기 (i) 20 내지 295, (ii) 20 내지 647, (iii) 20 내지 880, (iv) 414 내지 647 또는 (v) 414 내지 880에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;

(12) 서열 번호 76의 아미노산 서열에 대하여 또는 서열 번호 76의 잔기 (i) 19 내지 296, (ii) 19 내지 649, (iii) 19 내지 890, (iv) 415 내지 649 또는 (v) 415 내지 890에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드;

(13) 서열 번호 78의 아미노산 서열에 대하여 또는 서열 번호 78의 잔기 (i) 20 내지 354, (ii) 20 내지 660, (iii) 20 내지 805, (iv) 449 내지 660 또는 (v) 449 내지 805에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드; 또는

(14) 서열 번호 79의 아미노산 서열에 대하여 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드.

또한, 본 개시내용은 하기의 것을 제공한다:

(1) 서열 번호 53에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 53의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(2) 서열 번호 55에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 55의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(3) 서열 번호 57에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 57의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(4) 서열 번호 59에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 59의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(5) 서열 번호 61에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 61의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(6) 서열 번호 63에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 63의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(7) 서열 번호 65에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 65의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(8) 서열 번호 67에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 67의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(9) 서열 번호 69에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 69의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(10) 서열 번호 71에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 71의 상보체 또는 그의 단편과 혼성화할 수 있는 핵산;

(11) 서열 번호 73에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 73의 상보체 또는 그 단편과 혼성화할 수 있는 핵산;

(12) 서열 번호 75에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 75의 상보체 또는 그 단편과 혼성화할 수 있는 핵산; 또는

(13) 서열 번호 77에 대하여 적어도 90%(예를 들어, 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상)의 서열 동일성을 갖는 핵산, 또는 높은 엄격성 조건 하에서 서열 번호 77의 상보체 또는 그 단편과 혼성화할 수 있는 핵산.

본 명세서에 사용되는 용어 "낮은 엄격성, 중간 엄격성, 높은 엄격성 또는 매우 높은 엄격성 조건 하에서 혼성화한다"는 혼성화 및 세척을 위한 조건을 기술한다. 혼성화 반응을 수행하기 위한 지침은 문헌[Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1 - 6.3.6]에서 찾을 수 있다. 수성 및 비수성 방법은 상기 문헌에 기재되어 있으며, 어느 하나의 방법이 사용될 수 있다. 본 명세서에서 언급되는 특정 혼성화 조건은 하기와 같다: 1) 약 45℃에서 6X 염화나트륨/시트르산나트륨(SSC)에 이어서 50℃ 이상에서 0.2X SSC, 0.1% SDS에서의 2회 세척의 낮은 엄격성 혼성화 조건(세척 온도는 낮은 엄격성 조건의 경우에 55℃로 증가될 수 있다); 2) 약 45℃에서 6X SSC에 이어서 60℃에서 0.2X SSC, 0.1% SDS에서의 1회 이상의 세척의 중간 엄격성 혼성화 조건; 3) 약 45℃에서 6X SSC에 이어서 65℃에서 0.2X SSC, 0.1% SDS에서 1회 이상의 세척의 높은 엄격성 혼성화 조건; 그리고 바람직하게는 4) 65℃에서 0.5M 인산나트륨, 7% SDS에 이어서 65℃에서 0.2X SSC, 1% SDS에서 1회 이상의 세척의 매우 높은 엄격성 혼성화 조건. 매우 높은 엄격성 조건(4)는 달리 규정되지 않는 한 바람직한 조건이다.

핵산을 분리하기 위한 방법의 예

본 개시내용의 β-글루코시다제 및 기타 핵산은 표준 방법을 이용하여 분리될 수 있다. 대상으로 하는 공급원 유기체(예컨대, 박테리아 계통)으로부터 원하는 핵산을 얻는 방법은 분자 생물학 분야에서 통상적이며 공지되어 있다. 공지된 서열의 PCR 증폭, 핵산 합성, 계통 라이브러리 스크리닝, 코스미드 라이브러리 스크리닝을 비롯하여, 핵산을 분리하는 표준 방법은 국제 특허 공개 제WO 2009/076676 A2호 및 미국 특허 출원 제12/335,071호에 기술되어 있다.

숙주 세포의 예

본 개시내용은 본 개시내용의 하나 이상의 효소를 발현하도록 조작된 숙주 세포를 제공한다. 적절한 숙주 세포는 임의의 미생물의 세포(예를 들어, 박테리아, 원생생물, 조류, 진균(예를 들어, 효모 또는 사상 진균), 또는 기타 미생물의 세포)를 포함하며, 바람직하게는 박테리아, 효모, 또는 사상 진균의 세포이다.

박테리아 속의 적절한 숙주 세포에는 에스케리키아(*Escherichia*), 바실루스, 락토바실루스(*Lactobacillus*), 슈도모나스(*Pseudomonas*) 및 스트렙토마이세스의 세포가 포함되나 이들에 한정되지 않는다. 박테리아 종의 적절한 세포에는 에스케리키아 콜라이(*Escherichia coli*), 바실루스 서브틸리스(*Bacillus subtilis*), 바실루스 리케니포르미스(*Bacillus licheniformis*), 락토바실루스 브레비스(*Lactobacillus brevis*), 슈도모나스 아에루지노사(*Pseudomonas aeruginosa*), 및 스트렙토마이세스 리비단스(*Streptomyces lividans*)의 세포가 포함되나 이들에 한정되지 않는다.

효모 속의 적절한 숙주 세포에는 사카로마이세스, 스킴조사카로마이세스(*Schizosaccharomyces*), 칸디다(*Candida*), 한센울라(*Hansenula*), 피치아(*Pichia*), 클루이베로마이세스(*Kluyveromyces*) 및 파피아(*Phaffia*)의 세포가 포함되나 이들에 한정되지 않는다. 효모 종의 적절한 세포에는 사카로마이세스 세레비지애, 스킴조사카로마이세스 폼베(*Schizosaccharomyces pombe*), 칸디다 알비칸스(*Candida albicans*), 한센울라 폴리모르파(*Hansenula polymorpha*), 피치아 파스토리스(*Pichia pastoris*), 피치아 카나덴시스(*P. canadensis*), 클루이베로마이세스 마르시아누스(*Kluyveromyces marxianus*) 및 파피아 로도지마(*Phaffia rhodozyma*)의 세포가 포함되나 이들에 한정되지 않는다.

사상 진균의 적절한 숙주 세포에는 아문 진균류(*Eumycotina*)의 모든 사상형(filamentous form)이 포함된다. 사상 진균 속의 적절한 세포에는 예를 들어, 아크레모니움(*Acremonium*), 아스페르길루스, 아우레오바시디움(*Aureobasidium*), 비어칸테라(*Bjerkandera*), 세리포리옵시스(*Ceriporiopsis*), 크리소스포리움(*Chrysosporium*),

코프리누스(*Coprinus*), 코리올루스(*Coriolus*), 코리나스쿠스(*Corynascus*), 카에토미움(*Chaetomium*), 크립토코커스(*Cryptococcus*), 필로바시디움(*Filobasidium*), 푸사리움, 지베렐라(*Gibberella*), 후미콜라, 마그나포르테(*Magnaporthe*), 무코르, 마이셀리오프토라(*Myceliophthora*), 무코르, 네오칼리마스틱스(*Neocallimastix*), 뉴로스포라, 파에실로마이세스(*Paecilomyces*), 페니실리움, 파네로차에테(*Phanerochaete*), 플레비아(*Phlebia*), 피로마이세스(*Piromyces*), 플레우로투스(*Pleurotus*), 사이탈리디움(*Scytalidium*), 스킴조필룸(*Schizophyllum*), 스포로트리쿰(*Sporotrichum*), 탈라로마이세스(*Talaromyces*), 써모아스쿠스(*Thermoascus*), 티엘라비아(*Thielavia*), 톨리포클라디움(*Tolypocladium*), 트라메테스(*Trametes*) 및 트리코데르마의 세포가 포함되나, 이들에 한정되지 않는다.

사상 진균 종의 적절한 세포에는 아스페르길루스 아와모리(*Aspergillus awamori*), 아스페르길루스 푸미가투스(*Aspergillus fumigatus*), 아스페르길루스 포에티더스(*Aspergillus foetidus*), 아스페르길루스 야포니쿠스(*Aspergillus japonicus*), 아스페르길루스 니둘란스, 아스페르길루스 니게르, 아스페르길루스 오리자에, 크리스로스포리움 록크노웬스(*Chrysosporium lucknowense*), 푸사리움 박트리디오이데스(*Fusarium bactridioides*), 푸사리움 세레알리스(*Fusarium cerealis*), 푸사리움 크루크웰렌스(*Fusarium crookwellense*), 푸사리움 쿨모룸(*Fusarium culmorum*), 푸사리움 그라미네아룸(*Fusarium graminearum*), 푸사리움 그라미눔(*Fusarium graminum*), 푸사리움 헤테로스포룸(*Fusarium heterosporum*), 푸사리움 네군디(*Fusarium negundi*), 푸사리움 옥시스포룸, 푸사리움 레티쿨라툼(*Fusarium reticulatum*), 푸사리움 로세움(*Fusarium roseum*), 푸사리움 삼부시눔(*Fusarium sambucinum*), 푸사리움 사코크로움(*Fusarium sarcochroum*), 푸사리움 스포로트리키오이데스(*Fusarium sporotrichioides*), 푸사리움 숄푸레움(*Fusarium sulphureum*), 푸사리움 토룰로숨(*Fusarium torulosum*), 푸사리움 트리코테시오이데스(*Fusarium trichothecioides*), 푸사리움 베네나툼(*Fusarium venenatum*), 비어칸데라 아두스타(*Bjerkandera adusta*), 세리포리옵시스 아네이리나(*Ceriporiopsis aneirina*), 세리포리옵시스 아네이리나, 세리포리옵시스 카레지예아(*Ceriporiopsis caregiea*), 세리포리옵시스 질베슨스(*Ceriporiopsis gilvescens*), 세리포리옵시스 파노신타(*Ceriporiopsis pannocinta*), 세리포리옵시스 리블로사(*Ceriporiopsis rivulosa*), 세리포리옵시스 수브루파(*Ceriporiopsis subrufa*), 세리포리옵시스 수브베르미스포라(*Ceriporiopsis subvermisporea*), 코프리누스 시네레우스(*Coprinus cinereus*), 코리올루스 히르수투스(*Coriolus hirsutus*), 후미콜라 인솔렌스(*Humicola insolens*), 후미콜라 라누지노사(*Humicola lanuginosa*), 무코르 미에헤이(*Mucor miehei*), 마이셀리오프토라 써모필라(*Myceliophthora thermophila*), 뉴로스포라 크라사(*Neurospora crassa*), 뉴로스포라 인터메디아(*Neurospora intermedia*), 페니실리움 푸르푸로제눔(*Penicillium purpurogenum*), 페니실리움 카네스스(*Penicillium canescens*), 페니실리움 솔리툼(*Penicillium solitum*), 페니실리움 푸니쿨로숨, 파네로차에테 크리스로스포리움(*Phanerochaete chrysosporium*), 플레비아 라디아테(*Phlebia radiata*), 플레우로투스 에린지이(*Pleurotus eryngii*), 탈라로마이세스 플라부스(*Talaromyces flavus*), 티엘라비아 테레스트리스(*Thielavia terrestris*), 트라메테스 빌로사(*Trametes villosa*), 트라메테스 베르시콜로르(*Trametes versicolor*), 트리코데르마 하지아눔(*Trichoderma harzianum*), 트리코데르마 코닌지이(*Trichoderma koningii*), 트리코데르마 롱지브라키아툼(*Trichoderma longibrachiatum*), 트리코데르마 리세이 또는 트리코데르마 비리데(*Trichoderma viride*)의 세포가 포함되나 이들에 한정되지 않는다.

본 개시내용은 추가로, Fv3A, Pf43A, Fv43E, Fv39A, Fv43A, Fv43B, Pa51A, Gz43A, Fo43A, Af43A, Pf51A, AfuXyn2, AfuXyn5, Fv43D, Pf43B, Fv43B, Fv51A, 트리코데르마 리세이 Xyn3, 트리코데르마 리세이 Xyn2, 트리코데르마 리세이 Bx11, 트리코데르마 리세이 Bg11(Tr3A), GH61 엔도글루카나제, 트리코데르마 리세이 Eg4, Pa3D, Fv3G, Fv3D, Fv3C, Tr3B, Te3A, An3A, Fo3A, Gz3A, Nh3A, Vd3A, Pa3G 또는 Tn3B 폴리펩티드, 또는 그의 변이체 중 1개 이상, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상을 발현하도록 조작된 재조합 숙주 세포를 제공한다.

특정 실시형태에서, 2개 이상의 셀룰라제 서열 및/또는 헤미셀룰라제 서열로부터 유래되는 하이브리드 또는 키메라 효소를 발현하는 재조합 숙주 세포가 고려된다. 일부 태양에서, 하이브리드 또는 키메라 효소는 2개 이상의 β -글루코시다제 서열을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 136 내지 148의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 149 내지 156으로부터 선택되는 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열은 N-말단에 존재하고, 제2 β -글루코시다제 서열은 하이

브리드 또는 키메라 폴리펩티드의 C-말단에 존재한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접하거나 직접 연결되어 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다. 특정 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRRSPG(서열 번호 171), 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함하며, 이의 변형은 비변형된 대응 폴리펩티드, 또는 하이브리드 또는 키메라 폴리펩티드의 키메라 부분이 유래되는 폴리펩티드에 비해, 하이브리드 또는 키메라 폴리펩티드의 안정성을 향상시킨다. 특정 실시형태에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않지만, 링커 도메인은 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열의 변형, 예를 들어, 서열의 단축, 연장, 결실, 대체, 치환, 아니면 변경에 의해, 루프 서열의 잔기의 절단을 감소시킨다. 다른 실시형태에서, 루프 서열의 변형은 루프 서열 외측 부위에서의 잔기 절단을 감소시킨다.

특정 실시형태에서, 2개 이상의 셀룰라제 서열 및/또는 헤미셀룰라제 서열로부터 유래되는 하이브리드 또는 키메라 효소를 발현하는 재조합 숙주 세포가 고려된다. 일부 태양에서, 하이브리드 또는 키메라 효소는 2개 이상의 β -글루코시다제 서열을 포함한다. 일부 실시형태에서, 길이가 적어도 약 200개의 연속 아미노산 잔기로 되어 있으며, 서열 번호 60의 동일한 길이의 서열에 대하여 적어도 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 갖는 제1 서열; 및 길이가 적어도 약 50개의 연속 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 갖는 제2 서열을 포함하는 하이브리드 또는 키메라 효소를 발현하는 재조합 숙주 세포가 고려된다. 대안적인 실시 형태에서, 길이가 적어도 약 200개의 연속 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 갖는 제1 서열; 및 길이가 적어도 약 50개의 연속 아미노산 잔기로 되어 있으며, 서열 번호 60의 서열에 대하여 적어도 약 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 갖는 제2 서열을 포함하는 하이브리드 또는 키메라 효소를 발현하는 재조합 숙주 세포가 고려된다. 특정 실시형태에서, 제1 β -글루코시다제 서열은 N-말단에 존재하고, 제2 β -글루코시다제 서열은 하이브리드 또는 키메라 폴리펩티드의 C-말단에 존재한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접하거나 직접 연결되어 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다. 특정 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRRSPG(서열 번호 171), 또는 FD(R/K)YNIT(서열 번호 172)의 서열을 포함하는 루프 서열을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함하며, 이의 변형은 비변형된 대응 폴리펩티드, 또는 하이브리드 또는 키메라 폴리펩티드의 키메라 부분이 유래되는 폴리펩티드에 비해, 하이브리드 또는 키메라 폴리펩티드의 안정성을 향상시킨다. 특정 실시형태에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 서열을 포함하지 않지만, 링커 도메인은 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열의 변형, 예를 들어, 서열의 단축, 연장, 결실, 대체, 치환, 아니면 변경에 의해, 루프 서열의 잔기의 절단을 감소시킨다. 다른 실시형태에서, 루프 서열의 변형은 루프 서열 외측 부위에서의 잔기 절단을 감소시킨다.

일부 태양에서, 재조합 숙주 세포는 하나 이상의 키메라 효소, 예를 들어, Fv3C 융합 효소, 트리코테르마 리세이 Bgl3 융합 효소, Fv3C/Bgl3 융합 효소, Te3A 융합 효소, 또는 Fv3C/Te3A/Bgl3 융합 효소를 발현한다. 본 명세서의 개시내용에서, 용어 "XX 융합 효소", "XX 키메라 효소" 및 "XX 하이브리드 효소"는 XX 효소로부터 유래되는 적어도 하나의 키메라 부분을 갖는 효소를 지칭하도록 교호적으로 사용된다. 예를 들어, Fv3C 융합 또는 키메라 효소는 Fv3C/Bgl3 하이브리드 효소(또한 Bgl3 키메라 효소임), 또는 Fv3C/Te3A/Bgl3 하이브리드 효소(또한 Te3A 또는 Bgl3 키메라 효소임)를 지칭할 수 있다.

재조합 숙주 세포는 예를 들어, 재조합 트리코테르마 리세이 숙주 세포이다. 특정예에서, 본 개시내용은 재조합 진균, 예컨대 재조합 트리코테르마 리세이를 제공하며, 이는 Fv3A, Pf43A, Fv43E, Fv39A, Fv43A, Fv43B, Pa51A, Gz43A, Fo43A, Af43A, Pf51A, AfuXyn2, AfuXyn5, Fv43D, Pf43B, Fv43B, Fv51A, 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Xyn2, 트리코테르마 리세이 Bx11, 트리코테르마 리세이 Bgl1(Tr3A), 트리코테르마 리세이 Bgl3(Tr3B), GH61 엔도글루카나제, 트리코테르마 리세이 Eg4, Pa3D, Fv3G, Fv3D, Fv3C, Fv3C 융합/키메라 효소, Fv3C/Bgl3, Fv3C/Te3A/Bgl3 융합/키메라 효소, Te3A, An3A, Fo3A, Gz3A, Nh3A, Vd3A, Pa3G 또는 Tn3B 폴리펩티드, 또는 그의 변이체 또는 돌연변이체(예를 들어, 그의 하이브리드 또는 키메라 폴리펩티드를 포함)

중 1개 이상, 2개 이상, 3개 이상, 4개 이상, 또는 5개 이상을 발현하도록 조작된다.

본 개시내용은 적어도 하나의 자일라나제, 적어도 하나의 β -자일로시다제, 및 하나의 L- α -아라비노푸라노시다제를 재조합에 의해 발현하도록 조작된 숙주 세포, 예를 들어 재조합 진균 숙주 세포 또는 재조합 사상 진균을 제공한다. 본 개시내용은 또한 재조합 숙주 세포, 예를 들어, 재조합 진균 숙주 세포 또는 재조합 사상 진균, 예컨대 재조합 트리코테르마 리세이를 제공하며, 이는 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Xyn2, 트리코테르마 리세이 Bx11, 트리코테르마 리세이 Bg11, GH61 엔도글루카나제, 트리코테르마 리세이 Eg4, 또는 그의 변이체 중 하나 이상 이외에도, Fv3A, Pf43A, Fv43E, Fv39A, Fv43A, Fv43B, Pa51A, Gz43A, Fo43A, Af43A, Pf51A, AfuXyn2, AfuXyn5, Fv43D, Pf43B, Fv43B, Fv51A, Pa3D, Fv3G, Fv3D, Fv3C, Fv3C 융합 효소, 트리코테르마 리세이 Bg13(Tr3B), 트리코테르마 리세이 Bg13 융합 효소, Fv3C/Bg13 융합 효소, Tr3A, Te3A, Te3A 융합 효소, Fv3C/Te3A/Bg13 융합 효소, An3A, Fo3A, Gz3A, Nh3A, Vd3A, Pa3G 또는 Tn3B 폴리펩티드 중 1, 2, 3, 4, 5개, 또는 그 이상을 발현하도록 조작된다. 재조합 숙주 세포는 예를 들어, 트리코테르마 리세이 숙주 세포이다.

본 개시내용은 또한 재조합 숙주 세포, 예를 들어, 재조합 진균 숙주 세포 또는 재조합 유기체, 예를 들어, 사상 진균, 예컨대 재조합 트리코테르마 리세이를 제공하며, 이는 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Bg11, 트리코테르마 리세이 Bg13(Tr3B), 트리코테르마 리세이 Bg13 융합 효소, Fv3A, Fv43D, 및 Fv51A 폴리펩티드를 재조합에 의해 발현하도록 조작된다. 예를 들어, 재조합 숙주 세포는 적절하게는 트리코테르마 리세이 숙주 세포이다. 재조합 진균은 적절하게는 재조합 트리코테르마 리세이이다. 본 개시내용은 예를 들어, 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Bg11, 트리코테르마 리세이 Bg13 융합 효소, Fv3A, Fv43D, 및 Fv51A 폴리펩티드를 재조합에 의해 발현하도록 조작된 트리코테르마 리세이 숙주 세포를 제공한다.

프로모터 및 벡터의 예

또한, 본 개시내용은 상술한 핵산을 포함하는 발현 카세트 및/또는 벡터를 제공한다. 적절하게는, 본 개시내용의 효소를 암호화하는 핵산은 프로모터에 작동가능하게 연결된다. 프로모터는 당업계에 공지되어 있다. 숙주 세포에서 기능하는 임의의 프로모터는 본 개시내용의 β -글루코시다제 및/또는 임의의 기타 핵산의 발현을 위해 사용될 수 있다. 다양한 숙주 세포에서 본 개시내용의 β -글루코시다제 핵산 및/또는 임의의 기타 핵산의 발현을 구동시키는데 유용한 개시 조절 영역 또는 프로모터는 다수 존재하며, 당업자에게 잘 알려져 있다(예를 들어, 국제 특허 공개 제WO 2004/033646호 및 상기 공보에 인용된 참고문헌 참조). 사실상, 이들 핵산을 구동할 수 있는 임의의 프로모터가 사용될 수 있다.

구체적으로, 사상 진균 숙주에서의 재조합 발현이 요구되는 경우, 프로모터는 사상 진균 프로모터일 수 있다. 핵산은 예를 들어, 이중 프로모터의 제어 하에 존재할 수 있다. 또한, 핵산은 구성적 프로모터 또는 유도성 프로모터의 제어 하에서 발현될 수 있다. 사용될 수 있는 프로모터의 예에는 셀룰라제 프로모터, 자일라나제 프로모터, 1818 프로모터(이전에, 트리코테르마의 EST 맵핑(mapping)에 의해 고도로 발현되는 단백질로 동정된)가 포함되나, 이에 한정되지 않는다. 예를 들어, 프로모터는 적절하게는 셀로비오하이드롤라제, 엔도글루카나제 또는 β -글루코시다제 프로모터일 수 있다. 특히 적절한 프로모터는 예를 들어, 트리코테르마 리세이 셀로비오하이드롤라제, 엔도글루카나제, 또는 β -글루코시다제 프로모터일 수 있다. 예를 들어, 프로모터는 셀로비오하이드롤라제 I(*cbh1*) 프로모터이다. 프로모터의 비제한적인 예에는 *cbh1*, *cbh2*, *egl1*, *egl2*, *egl3*, *egl4*, *egl5*, *pki1*, *gpd1*, *xyn1*, 또는 *xyn2* 프로모터가 포함된다. 프로모터의 추가의 비제한적인 예에는 트리코테르마 리세이 *cbh1*, *cbh2*, *egl1*, *egl2*, *egl3*, *egl4*, *egl5*, *pki1*, *gpd1*, *xyn1*, 또는 *xyn2* 프로모터가 포함된다.

본 명세서에 사용되는 용어 "작동가능하게 연결된"은 선택된 뉴클레오티드 서열(예를 들어, 본 명세서에 기재된 폴리펩티드를 암호화하는)이 프로모터와 인접하게 존재하여 프로모터가 선택된 DNA의 발현을 조절할 수 있게 하는 것을 의미한다. 또한, 프로모터는 전사 및 번역의 방향의 면에서 선택된 뉴클레오티드 서열의 업스트림에 위치한다. "작동가능하게 연결된"은 뉴클레오티드 서열 및 조절 서열(들)이 적절한 분자(예를 들어, 전사 활성화제 단백질)가 조절 서열(들)에 결합되는 경우 유전자 발현을 가능하게 하는 방식으로 연결되는 것을 의미한다.

본 명세서에 기재된 임의의 β -글루코시다제 및/또는 기타 핵산은 하나 이상의 벡터에 포함될 수 있다. 따라서, 본 개시내용의 임의의 β -글루코시다제를 암호화하는 하나 이상의 핵산 및/또는 기타 핵산을 갖는 벡터도 본 명세서에 기재된다. 일부 태양에서, 벡터는 발현 제어 서열의 제어 하에 핵산을 포함한다. 일부 태양에서, 발현 제어 서열은 고유 발현 제어 서열이다. 일부 태양에서, 발현 제어 서열은 비고유 발현 제어 서열이다. 일부 태양에서, 벡터는 선택 마커 또는 선택가능한 마커를 포함한다. 일부 태양에서, 하나 이상의

β -글루코시다제(들)는 선택가능한 마커 없이 세포의 염색체로 통합된다.

적절한 벡터는 사용된 숙주 세포에 적합한 것들이다. 적절한 벡터는 예를 들어, 박테리아, 바이러스(예를 들어, 박테리오파지 T7 또는 M-13 유래의 파지), 코스미드, 효모 또는 식물로부터 유래될 수 있다. 적절한 벡터는 숙주 세포에 저, 중, 또는 고 카피수로 유지될 수 있다. 이러한 벡터를 얻어 사용하기 위한 프로토콜은 당업자에게 공지되어 있다(예를 들어, 문헌[Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor, 1989] 참조).

일부 태양에서, 발현 벡터는 또한 종결 서열을 포함한다. 종결 제어 영역도 또한 숙주 세포 고유의 다양한 유전자로부터 유래될 수 있다. 일부 태양에서, 종결 서열 및 프로모터 서열은 동일한 공급원으로부터 유래된다.

β -글루코시다제 핵산은 표준 기술을 이용하여, 벡터, 예컨대 발현 벡터로 혼입될 수 있다(문헌[Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1982]).

일부 태양에서, 본 개시내용에 기재된 하나 이상의 β -글루코시다제(들) 및/또는 하나 이상의 임의의 기타 핵산을 천연 세포에서 현재 발견된 것보다 훨씬 높은 수준으로 과발현시키는 것이 바람직할 수 있다. 일부 실시형태에서, 본 개시내용에 기재된 β -글루코시다제(들) 및/또는 하나 이상의 임의의 기타 핵산을 천연 세포에서 현재 발견된 것보다 훨씬 낮은 수준으로 저발현(예를 들어, 돌연변이, 불활성화, 또는 결실)시키는 것이 바람직할 수 있다.

형질전환법의 예

β -글루코시다제 핵산 또는 이들을 포함하는 벡터는 숙주 세포로 DNA 구축물 또는 벡터를 도입하기 위한 표준 기술, 예컨대 형질전환, 전기천공법, 핵 미세주입(nuclear microinjection), 형질도입, 트랜스펙션(예를 들어, 리포펙션 매개된 또는 DEAE-덱스트린 매개된 트랜스펙션, 또는 재조합 파지 바이러스를 이용한 트랜스펙션), 인산칼슘 DNA 침전물을 사용한 인큐베이션, DNA 코팅된 미세분사물에 의한 고속 충격(high velocity bombardment with DNA-coated microprojectile), 및 원형질체 융합을 이용하여, 숙주 세포(예를 들어, 본 명세서에 기재된 식물 세포, 진균 세포, 효모 세포, 또는 박테리아 세포)로 삽입될 수 있다. 통상적인 형질전환 기술은 당업계에 공지되어 있다(예를 들어, 문헌[Current Protocols in Molecular Biology (F. M. Ausubel *et al.* (eds) Chapter 9, 1987]; 문헌[Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor, 1989]; 및 문헌[Campbell *et al.*, *Curr. Genet.* 16:53-56, 1989] 참조). 도입된 핵산은 염색체 DNA에 통합되거나 염색체외 복제 서열로서 유지될 수 있다. 형질전환체는 당업계에 공지된 임의의 방법에 의해 선택될 수 있다.

세포 배양 배지의 예

일반적으로, 미생물은 본 명세서에 기재된 폴리펩티드의 생성에 적합한 세포 배양 배지에서 배양된다. 당업계에 공지된 절차 및 다양성을 이용하여, 탄소원 및 질소원 및 무기 염류를 포함하는 적절한 영양 배지에서 배양이 일어난다. 성장 및 셀룰라제 생성을 위한 적절한 배양 배지, 온도 범위 및 다른 조건이 당업계에 공지되어 있다. 비제한적인 예로서, 트리코테르마 리세이에 의한 셀룰라제 생성을 위한 전형적인 온도 범위는 24°C 내지 28°C이다.

세포 배양 조건의 예

박테리아 배양물의 유지 및 성장을 위해 적합한 재료 및 방법은 당업계에 주지되어 있다. 예시적인 기술은 문헌[*Manual of Methods for General Bacteriology* Gerhardt *et al.*, eds), American Society for Microbiology, Washington, D.C. (1994) or Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA]에서 찾을 수 있다. 일부 태양에서, 세포는 숙주 세포에 삽입된 핵산에 의해 암호화되는 하나 이상의 β -글루코시다제 폴리펩티드의 발현을 가능하게 하는 조건 하에서 배양 배지에서 배양된다. 표준 세포 배양 조건은 세포를 배양하는데 사용될 수 있다. 일부 태양에서, 세포는 적절한 온도, 가스 혼합물, 및 pH에서 성장되어 유지된다. 일부 태양에서, 세포는 적절한 세포 배지에서 성장된다.

본 발명의 조성물

또한, 본 개시내용은 하나 이상의 상술한 폴리펩티드가 풍부한 조작된 효소 조성물(예를 들어, 셀룰라제 조성물) 또는 발효 브로쓰를 제공한다. 일부 태양에서, 상기 조성물은 셀룰라제 조성물이다. 셀룰라제 조성물

은 예를 들어, 사상 진균 셀룰라제 조성물, 예를 들어 트리코테르마 셀룰라제 조성물일 수 있다. 일부 태양에서, 상기 조성물은 하나 이상의 셀룰라제 폴리펩티드를 암호화하는 하나 이상의 핵산을 포함하는 세포이다. 일부 태양에서, 상기 조성물은 셀룰라제 활성을 포함하는 발효 브로쓰이며, 여기서 브로쓰는 바이오매스 시료에 존재하는 셀룰로스를 약 50 wt. %를 초과하여 당으로 전환시킬 수 있다. 본 명세서에 사용되는 용어 "발효 브로쓰"는 발효에 의해 생성되며, 발효 후에 회수 및/또는 정제되지 않고/않거나 최소로 회수 및/또는 정제되는 효소 제제를 말한다. 발효 브로쓰는 사상 진균의 발효 브로쓰, 예를 들어, 트리코테르마, 후미콜라, 푸사리움, 아스페르길루스, 뉴로스포라, 페니실리움, 세팔로스פור리움, 아클리아, 포도스포라, 엔도티아, 무코르, 코클리옴볼루스, 피리쿨라리아 또는 크리소스포리움 발효 브로쓰일 수 있다. 특히, 발효 브로쓰는 예를 들어, 트리코테르마 *spp.*, 예를 들어, 트리코테르마 리세이 또는 페니실리움 *spp.*, 예를 들어, 페니실리움 푸니쿨로숨 중 하나일 수 있다. 발효 브로쓰는 또한 적절하게는 무세포 발효 브로쓰일 수 있다. 일 태양에서, 본 발명의 임의의 셀룰라제, 세포 또는 발효 브로쓰 조성물은 하나 이상의 헤미셀룰라제를 추가로 포함할 수 있다. 일 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 특정 실시형태에서, 발효 브로쓰는 제한된 생성-후 처리, 예를 들어, 정제, 한외여과, 여과, 또는 세포 사멸 단계와 함께 사용될 수 있으며, 그와 같이, 발효 브로쓰는 전체 브로쓰 제제에 사용된다고 한다. 일부 태양에서, 전체 셀룰라제 조성물은 트리코테르마 리세이에서 발현된다. 일부 태양에서, 전체 셀룰라제 조성물은 트리코테르마 리세이 통합 균주 H3A에서 발현된다. 일부 태양에서, 전체 셀룰라제 조성물은 트리코테르마 리세이 통합 균주 H3A에서 발현되며, 여기서 트리코테르마 리세이 통합 균주 H3A에서 발현되는 폴리펩티드의 하나 이상의 성분이 결실되어 있다. 일부 태양에서, 전체 셀룰라제 조성물은 아스페르길루스 니게르 또는 그의 조작된 균주에서 발현된다. 일부 태양에서, 셀룰라제 조성물은 칼코폴루오르 검정법에 의해 측정된 것으로서, 적어도 0.1 내지 0.4 분율의 생성물을 달성할 수 있다. 일부 태양에서, 셀룰라제 조성물은 조성물의 전체 효소 중량의 0.1 내지 25 wt. %를 포함한다. 일부 태양에서, 셀룰라제 조성물은 하나 이상의 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 셀룰라제 조성물은 바이오매스에 존재하는 셀룰로스의 중량을 약 70%, 75%, 80%, 85%, 90%를 초과하여 당으로 전환시킬 수 있다. 일부 태양에서, 셀룰라제 조성물은 폴리펩티드를 포함하며, 여기서 당으로 전환되는 바이오매스 시료 중의 셀룰로스의 wt. %는 폴리펩티드를 포함하지 않는 셀룰라제 조성물에 비해 증가된다.

일부 태양에서, 상기 조성물은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 어느 하나에 대하여 적어도 약 60%, 예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물이다. 일부 태양에서, 셀룰라제 조성물은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 어느 하나에 대하여 적어도 약 60%, 예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 폴리펩티드를 포함하며, 여기서 셀룰라제 조성물은 바이오매스 기질에 존재하는 셀룰로스를 약 30 wt.%를 초과하여, 예를 들어, 약 40 wt.%, 45 wt.%, 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 또는 80 wt.%를 초과하여 당으로 전환시킬 수 있다. 특정 실시형태에서, 바이오매스 기질은 전형적으로는 바이오매스 기질에 대하여 본 명세서에 기재된 것과 같은 특정 적절한 전처리 공정이 행해진 결과, 고체, 겔, 반액체, 또는 액체 형태 중의 혼합물이다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99%) 서열 동일성을 갖는 폴리펩티드를 포함하며, 바이오매스 시료에 존재하는 셀룰로스를 약 30 wt.%를 초과하여(예를 들어, 약 40 wt.%, 45 wt.%, 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 또는 80 wt.%를 초과하여) 당으로 전환시킬 수 있는 셀룰라제 조성물은 전체 세포 조성물이다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99%) 서열 동일성을 갖는 폴리펩티드를 포함하며, 바이오매스 시료에 존재하는 셀룰로스를 약 30 wt.%를 초과하여, 예를 들어, 약 40 wt.%, 45 wt.%, 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 또는 80 wt.%를 초과하여 당으로 전환시킬 수 있는 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 트리코데르마 리세이에서 발현된다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%,

98%, 또는 99%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 트리코테르마 리세이 통합 균주 H3A에서 발견된다. 일부 태양에서, 트리코테르마 리세이 통합 균주 H3A에서 발견되는 폴리펩티드의 하나 이상의 성분이 결실되어 있다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 적어도 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 아스페르길루스 니게르 또는 그의 조작된 균주에서 발견된다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 칼코플루오르 검정법에 의해 측정된 것으로서, 적어도 0.1 내지 0.4 분율의 생성물을 달성할 수 있다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 적어도 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 조성물의 단백질의 총 중량의 0.1 내지 25 wt.%(예를 들어, 0.5 내지 22 wt.%, 1 내지 20 wt.%, 5 내지 19 wt.%, 7 내지 18 wt.%, 9 내지 17 wt.%, 10 내지 15 wt.%)로 포함된다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 적어도 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 하나 이상의 헤미셀룰라제를 추가로 포함한다. 일부 태양에서, 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 적어도 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하는 셀룰라제 조성물은 바이오매스에 존재하는 셀룰로스의 중량을 약 50%를 초과하여(예를 들어, 약 55%, 60%, 65%, 70%, 75%, 80%, 85%, 또는 90%를 초과하여) 당으로 전환시킬 수 있다. 일부 태양에서, 셀룰라제 조성물은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79의 아미노산 서열 중 적어도 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 또는 90%) 서열 동일성을 갖는 폴리펩티드를 포함하며, 당으로 전환되는 바이오매스 시료 중의 셀룰로스의 wt.%는 폴리펩티드를 포함하지 않는 셀룰라제 조성물에 비해 증가된다.

일부 태양에서, 셀룰라제 조성물은 비천연 셀룰라제 조성물이며, 2개 이상의 β -글루코시다제 서열의 키메라/하이브리드/융합을 포함하고, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 동일한 길이의 (제1 β -글루코시다제 서열에 대하여) Fv3C 연속 서열(서열 번호 60)에 대하여 약 60%(예를 들어, 약 65%, 70%, 75%, 80%) 이상의 서열 동일성을 포함하고, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 (제2 β -글루코시다제 서열에 대하여) 연속 서열에 대하여 적어도 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%) 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 키메라 폴리펩티드의 N-말단에 존재하는 한편, 제2 β -글루코시다제 서열은 키메라 폴리펩티드의 C-말단에 존재한다. 일부 태양에서, 셀룰라제 조성물은 전체 세포 조성물이다. 일부 태양에서, 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다.

일부 태양에서, 셀룰라제 조성물은 비천연 셀룰라제 조성물이며, 2개 이상의 β -글루코시다제 서열의 키메라 또는 하이브리드를 포함하고, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 (제1 β -글루코시다제 서열에 대하여) 연속 서열에 대하여 약 60%(예를 들어, 약 65%, 70%, 75%, 80%) 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하고, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 동일한 길이의 (제2 β -글루코시다제 서열에 대하여) Fv3C 연속 서열(서열 번호 60)에 대하여 적어도 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%) 서열 동일성을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 키메라 폴리펩티드의 N-말단에 존재하는 한편, 제2 β -글루코시다제 서열은 키메라 폴리펩티드의 C-말단에 존재한다. 일부 태양에서, 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다.

특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 하이브리드 또는 키메라 β -글루코시다제 폴리펩티드의 중앙에 위치된다(즉, N-말단 또는 C-말단이 아님). 특정 실시형태에서, 제1 β -글루코시다

제 서열 또는 제2 β -글루코시다제 서열, 또는 이들 서열 둘 다는 하나 이상의 글리코실화 부위를 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 예를 들어, FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열은 제1 및 제2 β -글루코시다제 서열을 결합하는 링커 서열을 제공한다. 일부 태양에서, 셀룰라제 조성물은 전체 세포 조성물이다. 일부 태양에서, 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다.

일부 태양에서, 셀룰라제 조성물은 비천연 셀룰라제 조성물이며, 2개 이상의 β -글루코시다제 서열의 키메라 또는 하이브리드를 포함하고, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 동일한 길이의 (제1 β -글루코시다제 서열에 대하여) Fv3C 연속 서열(서열 번호 60)에 대하여 약 60%(예를 들어, 약 65%, 70%, 75%, 80%) 이상의 서열 동일성을 포함하고, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있으며, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나의 동일한 길이의 (제2 β -글루코시다제 서열에 대하여) 연속 서열에 대하여 적어도 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%) 서열 동일성을 포함하거나, 서열 번호 170의 폴리펩티드 서열 모티프를 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 키메라 폴리펩티드의 N-말단에 존재하는 한편, 제2 β -글루코시다제 서열은 키메라 폴리펩티드의 C-말단에 존재한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 하이브리드 또는 키메라 β -글루코시다제 폴리펩티드의 중앙에 위치된다(즉, N-말단 또는 C-말단이 아님). 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열, 또는 이들 서열 둘 다는 하나 이상의 글리코실화 부위를 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 예를 들어, FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열은 제1 및 제2 β -글루코시다제 서열을 결합하는 링커 서열을 제공한다. 일부 태양에서, 셀룰라제 조성물은 전체 세포 조성물이다. 일부 태양에서, 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다.

일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다. 일부 태양에서, 셀룰라제 조성물은 비천연 셀룰라제 조성물이며, 2개 이상의 β -글루코시다제 서열의 키메라 또는 하이브리드를 포함하고, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개(예를 들어, 적어도 약 250, 300, 350, 400, 또는 450개) 연속 아미노산 잔기로 되어 있으며, 서열 번호 136 내지 148의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하는 것인 한편; 제2 β -글루코시다제 서열은 길이가 적어도 약 50개(예를 들어, 적어도 약 50, 75, 100, 120, 150, 180, 200, 220, 또는 250개)의 연속 아미노산 잔기로 되어 있으며, 서열 번호 149 내지 156의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하는 것이다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 일부 태양에서, 제1 β -글루코시다제 서열은 키메라 폴리펩티드의 N-말단에 존재하는 한편, 제2 β -글루코시다제 서열은 키메라 폴리펩티드의 C-말단에 존재한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 하이브리드 또는 키메라 β -글루코시다제 폴리펩티드의 중앙에 위치된다(즉, N-말단 또는 C-말단이 아님). 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열, 또는 이들 서열 둘 다는 하나 이상의 글리코실화 부위를 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 예를 들어, FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열은 제1 및 제2 β -글루코시다제 서열을 결합하는 링커 서열을 제공한다. 일부 태양에서, 셀룰라제 조성물은 전체 세포 조성물이다. 일부 태양에서, 셀룰라제 조성물은 발효 브로쓰이다. 일부 태양에서, 발효 브로쓰는 전체 셀룰라제를 포함한다. 일부 태양에서, 발효 브로쓰는 무세포 발효 브로쓰이다.

헤미셀룰라제 조성물

일부 태양에서, 본 발명의 임의의 셀룰라제 조성물은 하나 이상의 헤미셀룰라제를 추가로 포함한다. 그런 경우에는, 그러면, 셀룰라제 조성물은 또한 헤미셀룰라제 조성물이다. 일부 태양에서, 본 발명의 헤미셀룰라제 조성물은 자일라나제, β -자일로시다제, L- α -아라비노푸라노시다제, 및 그 조합으로부터 선택되는 헤미셀룰라제를 포함한다. 일부 태양에서, 본 발명의 헤미셀룰라제 조성물은 적어도 하나의 자일라나제를 포함한다. 일부 태양에서, 적어도 하나의 자일라나제는 트리코테르마 리세이 Xyn2, 트리코테르마 리세이 Xyn3, AfuXyn2, 및 AfuXyn5로 이루어지는 그룹으로부터 선택된다. 일부 태양에서, 본 발명의 헤미셀룰라제 조성물은 적어도 하나의 β -자일로시다제를 포함한다. 일부 태양에서, β -자일로시다제는 β -자일로시다제로부터 선택되는 그룹 1 β -자일로시다제, 예를 들어, Fv3A 및 Fv43A를 포함한다. 일부 태양에서, β -자일로시다제는 β -자일로시다제로부터 선택되는 그룹 2 β -자일로시다제, 예를 들어, Pf43A, Fv43D, Fv39A, Fv43E, Fo43E, Fv43B, Pa51A, Gz43A, 및 트리코테르마 리세이 Bx11을 포함한다. 일부 태양에서, 본 발명의 셀룰라제 조성물은 그룹 1 또는 그룹 2의 β -자일로시다제로부터 선택되는 단일 β -자일로시다제를 포함한다. 일부 태양에서, 본 발명의 셀룰라제 조성물은 2개의 β -자일로시다제를 포함하며, 여기서 하나의 β -자일로시다제는 그룹 1로부터 선택되고, 다른 하나는 그룹 2로부터 선택된다. 일부 태양에서, 본 발명의 헤미셀룰라제 조성물은 적어도 하나의 L- α -아라비노푸라노시다제를 포함한다. 일부 태양에서, 적어도 하나의 L- α -아라비노푸라노시다제는 Af43A, Fv43B, Pf51A, Pa51A, 및 Fv51A로 이루어지는 그룹으로부터 선택된다.

자일라나제

일부 태양에서, 셀룰라제 조성물은 적어도 하나의 적절한 자일라나제를 포함하는, 헤미셀룰라제 조성물이다. 일부 태양에서, 적어도 하나의 자일라나제는 트리코테르마 리세이 Xyn2, 트리코테르마 리세이 Xyn3, AfuXyn2, 및 AfuXyn5로 이루어지는 그룹으로부터 선택된다.

임의의 자일라나제(EC 3.2.1.8)는 하나 이상의 자일라나제로서 사용될 수 있다. 적절한 자일라나제에는 예를 들어, 칼도셀룸 사카롤리티쿰(*Caldocellum saccharolyticum*) 자일라나제(문헌[Luthi et al. 1990, Appl. Environ. Microbiol. 56(9):2677-2683]), 씨모토가 마리티마(*Thermotoga maritima*) 자일라나제(문헌[Winterhalter & Liebel, 1995, Appl. Environ. Microbiol. 61(5):1810-1815]), 씨모토가 Sp. 군주 FJSS-B.1 자일라나제(문헌[Simpson et al. 1991, Biochem. J. 277, 413-417]), 바실루스 서큘란스(*Bacillus circulans*) 자일라나제(BcX) (미국 특허 제5,405,769), 아스페르길루스 니게르 자일라나제(문헌[Kinoshita et al. 1995, Journal of Fermentation and Bioengineering 79(5):422-428]), 스트렙토마이세스 리비단스(*Streptomyces lividans*) 자일라나제 (문헌[Shareck et al. 1991, Gene 107:75-82; Morosoli et al. 1986 Biochem. J. 239:587-592; Kluepfel et al. 1990, Biochem. J. 287:45-50]), 바실루스 서브틸리스 자일라나제(문헌[Bernier et al. 1983, Gene 26(1):59-65]), 셀룰로모나스 피미(*Cellulomonas fimi*) 자일라나제(문헌[Clarke et al., 1996, FEMS Microbiology Letters 139:27-35]), 슈도모나스 플루오레스센스(*Pseudomonas fluorescens*) 자일라나제(문헌[Gilbert et al. 1988, Journal of General Microbiology 134:3239-3247]), 클로스트리디움 씨모셀룸(*Clostridium thermocellum*) 자일라나제(문헌[Dominguez et al., 1995, Nature Structural Biology 2:569-576]), 바실루스 푸밀루스(*Bacillus pumilus*) 자일라나제(문헌[Nuyens et al. Applied Microbiology and Biotechnology 2001, 56:431-434; Yang et al. 1998, Nucleic Acids Res. 16(14B):7187]), 클로스트리디움 아세토부틸리쿰(*Clostridium acetobutylicum*) P262 자일라나제(문헌[Zappe et al. 1990, Nucleic Acids Res. 18(8):2179]), 또는 트리코테르마 하지아눔 자일라나제(문헌[Rose et al. 1987, J. Mol. Biol.194(4):755-756])가 포함된다.

Xyn2

일부 태양에서, 본 발명의 셀룰라제 조성물은 Xyn2를 추가로 포함한다. 트리코테르마 리세이 Xyn2의 아미노산 서열(서열 번호 43)은 도 25 및 59b에 나타나 있다. 서열 번호 43은 미성숙 트리코테르마 리세이 Xyn2의 서열이다. 트리코테르마 리세이 Xyn2는 서열 번호 43의 잔기 1 내지 33에 해당하는 예측된 프리프로펩티드 서열(도 25에서 밑줄 그어짐)을 가지며; 위치 16과 17 사이의 예측된 신호 서열의 절단에 의해, 위치 32와 33 사이에서 켁신(kexin)-유사 프로테아제에 의해 처리되는 프로펩티드가 제공되어, 서열 번호 43의 잔기 33 내지 222에 해당하는 서열을 갖는 성숙 단백질이 생성되는 것으로 예측된다. 예측된 보존 도메인은 도 25에서 볼드체로 되어 있다. 트리코테르마 리세이 Xyn2는 간접적으로 효소가 전처리된 바이오매스에서 또는 분리된 헤미셀룰로스에서 작용하는 경우 자일로비오시다제의 존재 하에서 증가된 자일로로스 단량체 생성을 촉매하는 그의 능력의 관찰에 의해 엔도자일라나제 활성을 갖는 것으로 나타났다. 보존된 산성 잔기는 E118, E123 및 E209를 포함한다. 본 명세서에 사용되는 "트리코테르마 리세이 Xyn2 폴리펩티드"는 서열 번호 43의 잔기 33 내지 222 중에서, 적어도 50, 75, 100, 125, 150 또는 175개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%,

91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. 트리코테르마 리세이 Xyn2 폴리펩티드는 바람직하게는 잔기 E118, E123 및 E209가 고유 트리코테르마 리세이 Xyn2와 비교하여, 변경되지 않는다. 트리코테르마 리세이 Xyn2 폴리펩티드는 바람직하게는 도 59b의 정렬에 나타난 바와 같이, 트리코테르마 리세이 Xyn2, AfuXyn2 및 AfuXyn5 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. 트리코테르마 리세이 Xyn2 폴리펩티드는 적절하게는 도 25에 나타난 고유 트리코테르마 리세이 Xyn2의 예측된 전체 보존 도메인을 포함한다. 예시적인 트리코테르마 리세이 Xyn2 폴리펩티드는 도 25에 나타난 성숙 트리코테르마 리세이 Xyn2 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 트리코테르마 리세이 Xyn2 폴리펩티드는 바람직하게는 자일라나제 활성을 갖는다.

Xyn3

일부 태양에서, 본 발명의 셀룰라제 조성물은 Xyn3를 추가로 포함한다. 트리코테르마 리세이 Xyn3의 아미노산 서열(서열 번호 42)은 도 24b에 나타나 있다. 서열 번호 42는 미성숙 트리코테르마 리세이 Xyn3의 서열이다. 트리코테르마 리세이 Xyn3는 서열 번호 42의 잔기 1 내지 16에 해당하는 예측된 신호 서열(도 24b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 42의 잔기 17 내지 347에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 24b에서 볼드체로 되어 있다. 트리코테르마 리세이 Xyn3는 간접적으로 효소가 전처리된 바이오매스에서 또는 단리된 헤미셀룰로스에서 작용하는 경우 자일로비오시다제의 존재 하에서 증가된 자일로스 단량체 생성을 촉매하는 그의 능력의 관찰에 의해 엔도자일라나제 활성을 갖는 것으로 나타났다. 보존된 촉매 잔기는 트리코테르마 리세이 Xyn3에 대하여 33% 서열 동일성을 갖는 스트렙토마이세스 할스테디(*Streptomyces halstedii*)로부터의 다른 GH10 패밀리 효소, Xys1 델타와의 정렬에 의해 결정시 E91, E176, E180, E195 및 E282를 포함한다(문헌[Canals *et al.*, 2003, Act Crystalogr. D Biol. 59:1447-53]). 본 명세서에 사용되는 "트리코테르마 리세이 Xyn3 폴리펩티드"는 서열 번호 42의 잔기 17 내지 347 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250 또는 300개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. 트리코테르마 리세이 Xyn3 폴리펩티드는 바람직하게는 잔기 E91, E176, E180, E195 및 E282가 고유 트리코테르마 리세이 Xyn3와 비교하여, 변경되지 않는다. 트리코테르마 리세이 Xyn3 폴리펩티드는 바람직하게는 트리코테르마 리세이 Xyn3와 Xys1 델타 사이에 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. 트리코테르마 리세이 Xyn3 폴리펩티드는 적절하게는 도 24b에 나타난 고유 트리코테르마 리세이 Xyn3의 예측된 전체 보존 도메인을 포함한다. 예시적인 트리코테르마 리세이 Xyn3 폴리펩티드는 도 24b에 나타난 성숙 트리코테르마 리세이 Xyn3 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 트리코테르마 리세이 Xyn3 폴리펩티드는 바람직하게는 자일라나제 활성을 갖는다.

AfuXyn2

일부 태양에서, 본 발명의 셀룰라제 조성물은 AfuXyn2를 추가로 포함한다. AfuXyn2의 아미노산 서열(서열 번호 24)은 도 19b 및 59b에 나타나 있다. 서열 번호 24는 미성숙 AfuXyn2의 서열이다. AfuXyn2는 서열 번호 24의 잔기 1 내지 18에 해당하는 예측된 신호 서열(도 19b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 24의 잔기 19 내지 228에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 GH11 보존 도메인은 도 19b에서 볼드체로 되어 있다. AfuXyn2는 간접적으로 효소가 전처리된 바이오매스에서 또는 단리된 헤미셀룰로스에서 작용하는 경우 자일로비오시다제의 존재 하에서 증가된 자일로스 단량체 생성을 촉매하는 그의 능력의 관찰에 의해 엔도자일라나제 활성을 갖는 것으로 나타났다. 보존된 촉매 잔기는 E124, E129 및 E215를 포함한다. 본 명세서에 사용되는 "AfuXyn2 폴리펩티드"는 서열 번호 24의 잔기 19 내지 228 중에서, 적어도 50, 75, 100, 125, 150, 175 또는 200개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. AfuXyn2 폴리펩티드는 바람직하게는 잔기 E124, E129 및 E215가 고유 AfuXyn2와 비교하여, 변경되지 않는다. AfuXyn2 폴리펩티드는 바람직하게는 도 59b의 정렬에 나타난 바와 같은 AfuXyn2, AfuXyn5 및 트리코테르마 리세이 Xyn2 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. AfuXyn2 폴리펩티드는 적절하게는 도 19b에 나타난 고유 AfuXyn2의 예측된 전체 보존 도메인을 포함한다. 예시적인 AfuXyn2 폴리펩티드는 도 19b에 나타난 성숙 AfuXyn2 서열에 대

하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 AfuXyn2 폴리펩티드는 바람직하게는 자일라나제 활성을 갖는다.

AfuXyn5

일부 태양에서, 본 발명의 셀룰라제 조성물은 AfuXyn5를 추가로 포함한다. AfuXyn5의 아미노산 서열(서열 번호 26)은 도 20b 및 59b에 나타나 있다. 서열 번호 26은 미성숙 AfuXyn5의 서열이다. AfuXyn5는 서열 번호 26의 잔기 1 내지 19에 해당하는 예측된 신호 서열(도 20b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 26의 잔기 20 내지 313에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 GH11 보존 도메인은 도 20b에서 볼드체로 되어 있다. AfuXyn5는 간접적으로 효소가 전처리된 바이오매스에서 또는 단리된 헤미셀룰로스에서 작용하는 경우 자일로비오시다제의 존재 하에서 증가된 자일로스 단량체 생성을 촉매하는 그의 능력의 관찰에 의해 엔도자일라나제 활성을 갖는 것으로 나타났다. 보존된 촉매 잔기는 E119, E124 및 E210을 포함한다. 예측된 CBM은 수많은 소수성 잔기를 특징으로 하는 C-말단 근처에 존재하며, 긴 세린-, 트레오닌-풍부 시리즈의 아미노산이 뒤따른다. 영역은 도 59b에 밑줄 그어져 나타나 있다. 본 명세서에 사용되는 "AfuXyn5 폴리펩티드"는 서열 번호 26의 잔기 20 내지 313 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250 또는 275개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. AfuXyn5 폴리펩티드는 바람직하게는 잔기 E119, E120 및 E210이 고유 AfuXyn5와 비교하여, 변경되지 않는다. AfuXyn5 폴리펩티드는 바람직하게는 도 59b의 정렬에 나타난 바와 같은 AfuXyn5, AfuXyn2, 및 트리코테르마 리세이 Xyn2 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. AfuXyn5 폴리펩티드는 적절하게는 도 20b에 나타난 고유 AfuXyn5의 전체 예측된 CBM 및/또는 고유 AfuXyn5의 예측된 전체 보존 도메인(밑줄)을 포함한다. 예시적인 AfuXyn5 폴리펩티드는 도 20b에 나타난 성숙 AfuXyn5 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 AfuXyn5 폴리펩티드는 바람직하게는 자일라나제 활성을 갖는다.

자일라나제(들)는 본 개시내용의 셀룰라제 조성물의 약 0.05 wt.% 내지 약 50 wt.%를 구성하며, 여기서 wt.%는 주어진 조성물 중의 모든 효소의 합한 중량에 대한 자일라나제(들)의 합한 중량을 나타낸다. 자일라나제(들)는 하한치가 0.05 wt.%, 1 wt.%, 1.5 wt.%, 2 wt.%, 3 wt.%, 4 wt.%, 5 wt.%, 6 wt.%, 7 wt.%, 8 wt.%, 9 wt.%, 10 wt.%, 12 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 30 wt.%, 40 wt.%, 또는 45 wt.%이고, 상한치가 5 wt.%, 10 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 30 wt.%, 35 wt.%, 40 wt.%, 또는 50 wt.%인 범위로 존재할 수 있다. 적절하게는, 본 발명의 효소 조성물 중의 하나 이상의 자일라나제의 합한 중량은 예를 들어, 효소 조성물 중의 모든 효소의 총 중량의 약 0.05 wt.% 내지 약 50 wt.%(예를 들어, 0.05 wt.%, 1 wt.%, 2 wt.%, 3 wt.% 내지 50 wt.%, 3 wt.% 내지 40 wt.%, 3 wt.% 내지 30 wt.%, 3 wt.% 내지 20 wt.%, 5 wt.% 내지 20 wt.%, 10 wt.% 내지 30 wt.%, 15 wt.% 내지 35 wt.%, 20 wt.% 내지 40 wt.%, 20 wt.% 내지 50 wt.% 등)를 구성할 수 있다.

자일라나제는 자일라나제를 암호화하는 내인성 또는 외인성 유전자를 발현함으로써 생성될 수 있다. 자일라나제는 경우에 따라, 과발현되거나 저발현될 수 있다.

β -자일로시다제

일부 태양에서, 본 발명의 셀룰라제 조성물은 적어도 하나의 β -자일로시다제를 포함한다. 일부 태양에서, 셀룰라제 조성물은 예를 들어, Fv3A 및 Fv43A로 이루어진 그룹으로부터 선택되는 적어도 하나의 그룹 1 β -자일로시다제를 포함한다. 일부 태양에서, 셀룰라제 조성물은 예를 들어, Pf43A, Fv43D, Fv39A, Fv43E, Fo43E, Fv43B, Pa51A, Gz43A, 및 트리코테르마 리세이 Bx11로 이루어지는 그룹으로부터 선택되는 적어도 하나의 그룹 2 β -자일로시다제를 포함한다. 일부 태양에서, 셀룰라제 조성물은 단일 β -자일로시다제를 포함하며, β -자일로시다제는 그룹 1 또는 그룹 2 중 하나로부터 선택된다. 일부 태양에서, 셀룰라제 조성물은 2개의 β -자일로시다제를 포함하며, 여기서 하나의 β -자일로시다제는 그룹 1로부터 선택되고, 다른 하나는 그룹 2로부터 선택된다.

임의의 β -자일로시다제(EC 3.2.1.37)는 적절한 β -자일로시다제로서 사용될 수 있다. 적절한 β -자일로시다제에는 예를 들어, 탈라로마이세스 에메르소니이 Bx11(문헌[Reen *et al.* 2003, Biochem Biophys Res Commun. 305(3):579-85]), 지오바실루스 스테아로써모필러스(*G. stearothermophilus*) β -자일로시다제(문헌[Shallom *et al.* 2005, Biochemistry 44:387-397]), 스테이네르네마 썬모필룸(*S. thermophilum*) β -자일로시다제(문헌[Zanoelo *et al.* 2004, J. Ind. Microbiol. Biotechnol. 31:170-176]), 트리코테르마 리그노룸(*T. lignorum*)

β -자일로시다제(문헌[Schmidt, 1998, Methods Enzymol. 160:662-671]), 아스페르길루스 아와모리 β -자일로시다제(문헌[Kurakake *et al.* 2005, Biochim. Biophys. Acta 1726:272-279]), 아비쿨라리아 베르시콜로르(*A. versicolor*) β -자일로시다제(문헌[Andrade *et al.* 2004, Process Biochem. 39:1931-1938]), 스트랩토마이세스 *sp.* β -자일로시다제(문헌[Pinphanichakarn *et al.* 2004, World J. Microbiol. Biotechnol. 20:727-733]), 썬모토가 마리티마 β -자일로시다제(문헌[Xue and Shao, 2004, Biotechnol. Lett. 26:1511-1515]), 트리코테르마 *sp.* SY β -자일로시다제(문헌[Kim *et al.* 2004, J. Microbiol. Biotechnol. 14:643-645]), 아스페르길루스 니게르 β -자일로시다제(문헌[Oguntimein and Reilly, 1980, Biotechnol. Bioeng. 22:1143-1154]), 또는 페니실리움 워트만니(*P. wortmanni*) β -자일로시다제(문헌[Matsuo *et al.* 1987, Agric. Biol. Chem. 51:2367-2379])가 포함된다. 적절한 β -자일로시다제는 숙주 유기체에 의해 내인적으로 생성될 수 있거나, 숙주 유기체에 의해 재조합에 의해 클로닝되고/되거나 발현될 수 있다. 게다가, 적절한 β -자일로시다제는 정제되거나 단리된 형태로 셀룰라제 조성물에 첨가될 수 있다.

Fv3A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv3A 폴리펩티드를 포함한다. Fv3A의 아미노산 서열(서열 번호 2)은 도 8b 및 56에 나타나 있다. 서열 번호 2는 미성숙 Fv3A의 서열이다. Fv3A는 서열 번호 2의 잔기 1 내지 23에 해당하는 예측된 신호 서열(밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 2의 잔기 24 내지 766에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 8b에서 볼드체로 되어 있다. Fv3A는 예를 들어, 기질로서 *p*-니트로페닐- β -자일로피라노시드, 자일로비오스, 혼합된 선행 자일로-올리고머, 헤미셀룰로스로부터의 분지형 아라비녹실란 올리고머 또는 희석 암모니아로 전처리된 옥수수 속대를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기는 D291 인 반면에, 인접 잔기, S290 및 C292가 기질 결합에 관여하는 것으로 예상된다. E175 및 E213은 다른 GH3 및 GH39 효소에 걸쳐 보존되며, 촉매 기능을 갖는 것으로 예상된다. 본 명세서에 사용되는 "Fv3A 폴리펩티드"는 서열 번호 2의 잔기 24 내지 766 중에서, 적어도 50개, 예를 들어, 적어도 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700개의 연속 아미노산 잔기에 대하여 적어도 85%, 예를 들어, 적어도 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv3A 폴리펩티드는 바람직하게는 잔기 D291, S290, C292, E175 및 E213이 공유 Fv3A와 비교하여, 변경되지 않는다. Fv3A 폴리펩티드는 바람직하게는 도 56의 정렬에 나타난 바와 같이, Fv3A와 트리코테르마 리세이 Bx11 사이에 보존되는 아미노산 잔기의 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99%가 변경되지 않는다. Fv3A 폴리펩티드는 적절하게는 도 8b에 나타난 바와 같은 공유 Fv3A의 예측된 전체 보존 도메인을 포함한다. 본 발명의 예시적인 Fv3A 폴리펩티드는 도 8b에 나타난 성숙 Fv3A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv3A 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Fv3A 폴리펩티드는 적절하게는 서열 번호 2의 아미노산 서열에 대하여 또는 서열 번호 2의 잔기 (i) 24 내지 766, (ii) 73 내지 321, (iii) 73 내지 394, (iv) 395 내지 622, (v) 24 내지 622, 또는 (vi) 73 내지 622에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Fv43A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv43A 폴리펩티드를 포함한다. Fv43A의 아미노산 서열(서열 번호 10)은 도 12b 및 도 57에 제공된다. 서열 번호 10은 미성숙 Fv43A의 서열이다. Fv43A는 서열 번호 10의 잔기 1 내지 22에 해당하는 예측된 신호 서열(도 12b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 10의 잔기 23 내지 449에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 도 12b에서, 예측된 보존 도메인은 볼드체로 되어 있으며, 예측된 CBM은 대문자로 되어 있으며, CD와 CBM을 분리하는 예측된 링커는 이탤릭체로 되어 있다. Fv43A는 예를 들어, 기질로서 4-니트로페닐- β -D-자일로피라노시드, 자일로비오스, 혼합된 선행 자일로-올리고머, 헤미셀룰로스로부터의 분지형 아라비녹실란 올리고머 및/또는 선행 자일로-올리고머를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기에는 D34 또는 D62 중 어느 하나, D148 및 E209가 포함된다. 본 명세서에 사용되는 "Fv43A 폴리펩티드"는 서열 번호 10의 잔기 23 내지 449 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 또는 400개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv43A 폴리펩티드

는 바람직하게는 잔기 D34 또는 D62, D148 및 E209가 고유 Fv43A와 비교하여, 변경되지 않는다. Fv43A 폴리펩티드는 바람직하게는 도 57의 정렬에서 Fv43A 및 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 모두의 다른 아미노산 서열을 포함하는 효소의 패밀리 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv43A 폴리펩티드는 적절하게는 도 12b에 나타난 바와 같은 고유 Fv43A의 예측된 전체 CBM 및/또는 고유 Fv43A의 예측된 전체 보존 도메인, 및/또는 Fv43A의 링커를 포함한다. 예시적인 Fv43A 폴리펩티드는 도 12b에 나타난 성숙 Fv43A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv43A 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Fv43A 폴리펩티드는 적절하게는 서열 번호 10의 아미노산 서열에 대하여 또는 서열 번호 10의 잔기 (i) 23 내지 449, (ii) 23 내지 302, (iii) 23 내지 320, (iv) 23 내지 448, (v) 303 내지 448, (vi) 303 내지 449, (vii) 321 내지 448, 또는 (viii) 321 내지 449에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Pf43A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Pf43A 폴리펩티드를 포함한다. Pf43A의 아미노산 서열(서열 번호 4)은 도 9b 및 57에 나타나 있다. 서열 번호 4는 미성숙 Pf43A의 서열이다. Pf43A는 서열 번호 4의 잔기 1 내지 20에 해당하는 예측된 신호 서열(도 9b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 4의 잔기 21 내지 445에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 도 9b에서, 예측된 보존 도메인은 볼드체로 되어 있으며, 예측된 CBM은 대문자로 되어 있으며, CD와 CBM을 분리하는 예측된 링커는 이탤릭체로 되어 있다. Pf43A는 예를 들어, 기질로서 *p*-니트로페닐- β -자일로피라노시드, 자일로비오스, 혼합된 선형 자일로-올리고머 또는 희석 암모니아로 전처리된 옥수수 속대를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기에는 D32 또는 D60 중 어느 하나, D145 및 E206이 포함된다. 도 57에 밑줄 그어져 있는 C-말단 영역은 예측된 CBM이다. 본 명세서에 사용되는 "Pf43A 폴리펩티드"는 서열 번호 4의 잔기 21 내지 445 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 또는 400개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Pf43A 폴리펩티드는 바람직하게는 잔기 D32 또는 D60, D145 및 E206이 고유 Pf43A와 비교하여, 변경되지 않는다. Pf43A는 바람직하게는 도 57의 정렬에서 Pf43A 및 1, 2, 3, 4, 5, 6, 7 또는 8개 모두의 다른 아미노산 서열을 포함하는 단백질의 패밀리에 걸쳐 보존되는 것으로 관찰되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. 본 발명의 Pf43A 폴리펩티드는 적절하게는 하기 도메인 중 2개 이상 또는 모두를 포함한다: 도 9b에 나타난 바와 같은 Pf43A의 (1) 예측된 CBM, (2) 예측된 보존 도메인 및 (3) 링커. 본 발명의 예시적인 Pf43A 폴리펩티드는 도 9b에 나타난 성숙 Pf43A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Pf43A 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 포함한다.

따라서, 본 발명의 Pf43A 폴리펩티드는 서열 번호 4의 아미노산 서열에 대하여 또는 서열 번호 4의 잔기 (i) 21 내지 445, (ii) 21 내지 301, (iii) 21 내지 323, (iv) 21 내지 444, (v) 302 내지 444, (vi) 302 내지 445, (vii) 324 내지 444, 또는 (viii) 324 내지 445에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Fv43D

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv43D 폴리펩티드를 추가로 포함한다. Fv43D의 아미노산 서열(서열 번호 28)은 도 21b 및 57에 나타나 있다. 서열 번호 28은 미성숙 Fv43D의 서열이다. Fv43D는 서열 번호 28의 잔기 1 내지 20에 해당하는 예측된 신호 서열(도 21b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 28의 잔기 21 내지 350에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 21b에서 볼드체로 되어 있다. Fv43D는 예를 들어, 기질로서 *p*-니트로페닐- β -자일로피라노시드, 자일로비오스 및/또는 혼합된 선형 자일로-올리고머를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기에는 D37 또는 D72 중 어느 하나, D159 및 E251이 포함된다. 본 명세서에 사용되는 "Fv43D 폴리펩티드"는 서열 번호 28의 잔기 21 내지 350 중에서, 적어도 50, 75, 100, 125,

150, 175, 200, 250, 300 또는 320개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv43D 폴리펩티드는 바람직하게는 잔기 D37 또는 D72, D159 및 E251이 고유 Fv43D와 비교하여, 변경되지 않는다. Fv43D 폴리펩티드는 바람직하게는 도 57의 정렬에서 Fv43D 및 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 모두의 다른 아미노산 서열을 포함하는 효소의 그룹 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv43D 폴리펩티드는 적절하게는 도 21b에 나타난 고유 Fv43D의 예측된 전체 CD를 포함한다. 예시적인 Fv43D 폴리펩티드는 도 21b에 나타난 성숙 Fv43D 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv43D 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Fv43D 폴리펩티드는 서열 번호 28의 아미노산 서열에 대하여 또는 서열 번호 28의 잔기 (i) 20 내지 341, (ii) 21 내지 350, (iii) 107 내지 341, 또는 (iv) 107 내지 350에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Fv39A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv39A 폴리펩티드를 포함한다. Fv39A의 아미노산 서열(서열 번호 8)은 도 11b에 나타나 있다. 서열 번호 8은 미성숙 Fv39A의 서열이다. Fv39A2는 서열 번호 8의 잔기 1 내지 19에 해당하는 예측된 신호 서열(도 11b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 8의 잔기 20 내지 439에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 11b에서 볼드체로 나타나 있다. Fv39A는 예를 들어, 기질로서 *p*-니트로페닐- β -자일로피라노시드, 자일로비오스, 또는 혼합된 선형 자일로-올리고머를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. Fv39A 잔기 E168 및 E272는 각각, 씨모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*) (Uniprot 수탁 번호 P36906) 및 지오바실루스 스테아로씨모필러스(Uniprot 수탁 번호 Q9ZFM2)로부터의 상술한 GH39 자일로시다제와 Fv39A의 서열 정렬에 기초하여, 촉매 산-염기 및 친핵체로 기능하는 것으로 예측된다. 본 명세서에 사용되는 "Fv39A 폴리펩티드"는 서열 번호 8의 잔기 20 내지 439 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 또는 400개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv39A 폴리펩티드는 바람직하게는 잔기 E168 및 E272가 고유 Fv39A와 비교하여, 변경되지 않는다. Fv39A 폴리펩티드는 바람직하게는 Fv39A, 및 씨모안에어로박테리움 사카롤리티쿰 및 지오바실루스 스테아로씨모필러스로부터의 자일로시다제를 포함하는 패밀리 또는 효소 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98%, 또는 99%가 변경되지 않는다(상기 참조). Fv39A 폴리펩티드는 적절하게는 도 11b에 나타난 바와 같은 고유 Fv39A의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fv39A 폴리펩티드는 도 11b에 나타난 성숙 Fv39A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv39A 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Fv39A 폴리펩티드는 적절하게는 서열 번호 8의 아미노산 서열에 대하여 또는 서열 번호 8의 잔기 (i) 20 내지 439, (ii) 20 내지 291, (iii) 145 내지 291, 또는 (iv) 145 내지 439에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Fv43E

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv43E 폴리펩티드를 포함한다. Fv43E의 아미노산 서열(서열 번호 6)은 도 10b 및 57에 나타나 있다. 서열 번호 6은 미성숙 Fv43E의 서열이다. Fv43E는 서열 번호 6의 잔기 1 내지 18에 해당하는 예측된 신호 서열(도 10b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 6의 잔기 19 내지 530에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 10b에서 볼드체로 표시되어 있다. Fv43E는 예를 들어, 기질로서 4-니트로페닐- β -D-자일로피라노시드, 자일로비오스 및 혼합된 선형 자일로-올리고머 또는 희석 암모니아로 전처리된 옥수수 속대를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기에는 D40 또는 D71 중 어느 하나, D155 및 E241이 포함된다. 본 명세서에 사용되는 "Fv43E 폴리펩티드"는 서열 번호 6의 잔기 19 내지 530 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450 또는 500개의 연속 아미노산 잔기에

대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv43E 폴리펩티드는 바람직하게는 잔기 D40 또는 D71, D155 및 E241이 고유 Fv43E와 비교하여, 변경되지 않는다. Fv43E 폴리펩티드는 바람직하게는 도 57의 정렬에서 Fv43E 및 1, 2, 3, 4, 5, 6, 7 또는 8개 모두의 다른 아미노산 서열을 포함하는 효소의 패밀리 중에서 보존되는 것으로 관찰되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv43E 폴리펩티드는 적절하게는 도 10b에 나타난 바와 같은 고유 Fv43E의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fv43E 폴리펩티드는 도 10b에 나타난 성숙 Fv43E 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv43E 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Fv43E 폴리펩티드는 적절하게는 서열 번호 6의 아미노산 서열에 대하여 또는 서열 번호 6의 잔기 (i) 19 내지 530, (ii) 29 내지 530, (iii) 19 내지 300, 또는 (iv) 29 내지 300에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

Fv43B

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv43B 폴리펩티드를 포함한다. Fv43B의 아미노산 서열(서열 번호 12)은 도 13b 및 57에 나타나 있다. 서열 번호 12는 미성숙 Fv43B의 서열이다. Fv43B는 서열 번호 12의 잔기 1 내지 16에 해당하는 예측된 신호 서열(도 13b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 12의 잔기 17 내지 574에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 13b에서 볼드체로 되어 있다. Fv43B는 예를 들어, 기질로서 4-니트로페닐- β -D-자일로피라노시드 및 *p*-니트로페닐- α -L-아라비노푸라노시드를 사용하는 제1 효소적 검정법에서 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 모두를 갖는 것으로 나타났다. 제2 효소적 검정법에서, 다른 자일로시다제 효소의 존재 하에서 분지형 아라비노-자일로올리고머로부터 아라비노스의 방출이 촉매되고, 올리고머 혼합물로부터 증가된 자일로스 방출이 촉매되는 것으로 나타났다. 예측된 촉매 잔기에는 D38 또는 D68 중 어느 하나, D151 및 E236이 포함된다. 본 명세서에 사용되는 "Fv43B 폴리펩티드"는 서열 번호 12의 잔기 17 내지 574 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500 또는 550개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv43B 폴리펩티드는 바람직하게는 잔기 D38 또는 D68, D151 및 E236이 고유 Fv43B와 비교하여, 변경되지 않는다. Fv43B 폴리펩티드는 바람직하게는 도 57의 정렬에서 Fv43B 및 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 모두의 다른 아미노산 서열을 포함하는 효소의 패밀리 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv43B 폴리펩티드는 적절하게는 도 13b 및 도 57에 나타난 바와 같은 고유 Fv43B의 예측된 전체 보존 도메인을 포함한다. 예시적인 Fv43B 폴리펩티드는 도 13b에 나타난 성숙 Fv43B 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv43B 폴리펩티드는 바람직하게는 β -자일로시다제 활성, L- α -아라비노푸라노시다제 활성, 또는 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 다를 갖는다.

따라서, 본 발명의 Fv43B 폴리펩티드는 서열 번호 12의 아미노산 서열에 대하여 또는 서열 번호 12의 잔기 (i) 17 내지 574, (ii) 27 내지 574, (iii) 17 내지 303, 또는 (iv) 27 내지 303에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성, L- α -아라비노푸라노시다제 활성, 또는 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 다를 갖는다.

Pa51A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Pa51A 폴리펩티드를 포함한다. Pa51A의 아미노산 서열(서열 번호 14)은 도 14b 및 58에 나타나 있다. 서열 번호 14는 미성숙 Pa51A의 서열이다. Pa51A는 서열 번호 14의 잔기 1 내지 20에 해당하는 예측된 신호 서열(도 14b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 14의 잔기 21 내지 676에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 도 14b에서 볼드체로 되어 있다. Pa51A는 예를 들어, 인공 기질 *p*-니트로페닐- β -자일로피라노시드 및 *p*-니트로페닐- α -L-아라비노푸라노시드를 사용하는 효소적 검정법에서 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 모두를 갖는 것으로 나타났다. 다른 자일로시다제 효소의 존

재 하에서 분지형 아라비노-자일로 올리고머로부터 아라비노스의 방출이 촉매되고, 올리고머 혼합물로부터 증가된 자일로스 방출이 촉매되는 것으로 나타났다. 보존된 산성 잔기에는 E43, D50, E257, E296, E340, E370, E485 및 E493이 포함된다. 본 명세서에 사용되는 "Pa51A 폴리펩티드"는 서열 번호 14의 잔기 21 내지 676 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600 또는 650개의 연속 아미노산 잔기에 대하여 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Pa51A 폴리펩티드는 바람직하게는 잔기 E43, D50, E257, E296, E340, E370, E485 및 E493이 고유 Pa51A와 비교하여, 변경되지 않는다. Pa51A 폴리펩티드는 바람직하게는 도 58의 정렬에 나타난 바와 같이 Pa51A, Fv51A 및 Pf51A를 포함하는 효소의 그룹 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Pa51A 폴리펩티드는 적절하게는 도 14b에 나타난 바와 같은 고유 Pa51A의 예측된 보존 도메인을 포함한다. 예시적인 Pa51A 폴리펩티드는 도 14b에 나타난 성숙 Pa51A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Pa51A 폴리펩티드는 바람직하게는 β -자일로시다제 활성, L- α -아라비노푸라노시다제 활성, 또는 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 다를 갖는다.

따라서, 본 발명의 Pa51A 폴리펩티드는 적절하게는 서열 번호 14의 아미노산 서열에 대하여 또는 서열 번호 14의 잔기 (i) 21 내지 676, (ii) 21 내지 652, (iii) 469 내지 652, 또는 (iv) 469 내지 676에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성, L- α -아라비노푸라노시다제 활성, 또는 β -자일로시다제 활성 및 L- α -아라비노푸라노시다제 활성 둘 다를 갖는다.

Gz43A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Gz43A 폴리펩티드를 포함한다. Gz43A의 아미노산 서열(서열 번호 16)은 도 15b 및 57에 나타나 있다. 서열 번호 16은 미성숙 Gz43A의 서열이다. Gz43A는 서열 번호 16의 잔기 1 내지 18에 해당하는 예측된 신호 서열(도 15b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 16의 잔기 19 내지 340에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 보존 도메인은 도 15b에서 볼드체로 되어 있다. Gz43A는 예를 들어, 기질로서 *p*-니트로페닐- β -자일로피라노시드, 자일로비오스, 또는 혼합된 및/또는 선형 자일로-올리고머를 사용하는 효소적 검정법에서 β -자일로시다제 활성을 갖는 것으로 나타났다. 예측된 촉매 잔기에는 D33 또는 D68 중 어느 하나, D154 및 E243이 포함된다. 본 명세서에 사용되는 "Gz43A 폴리펩티드"는 서열 번호 16의 잔기 19 내지 340 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 또는 300개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Gz43A 폴리펩티드는 바람직하게는 잔기 D33 또는 D68, D154 및 E243이 고유 Gz43A와 비교하여, 변경되지 않는다. Gz43A 폴리펩티드는 바람직하게는 도 57의 정렬에서 Gz43A 및 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 모두의 다른 아미노산 서열을 포함하는 효소의 그룹 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Gz43A 폴리펩티드는 적절하게는 도 15b에 나타난 바와 같은 고유 Gz43A의 예측된 보존 도메인을 포함한다. 예시적인 Gz43A 폴리펩티드는 도 15b에 나타난 성숙 Gz43A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Gz43A 폴리펩티드는 바람직하게는 β -자일로시다제 활성을 갖는다.

따라서, 본 발명의 Gz43A 폴리펩티드는 적절하게는 서열 번호 16의 아미노산 서열에 대하여 또는 서열 번호 16의 잔기 (i) 19 내지 340, (ii) 53 내지 340, (iii) 19 내지 383, 또는 (iv) 53 내지 383에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 β -자일로시다제 활성을 갖는다.

β -자일로시다제(들)는 적절하게는 본 발명의 셀룰라제 또는 헤미셀룰라제 조성물 중의 효소의 총 중량의 약 0 wt.% 내지 약 75 wt.%(예를 들어, 약 0.1 wt.% 내지 약 50 wt.%, 약 1 wt.% 내지 약 40 wt.%, 약 2 wt.% 내지 약 35 wt.%, 약 5 wt.% 내지 약 30 wt.%, 약 10 wt.% 내지 약 25 wt.%)를 구성한다. 서로에 대한 임의의 단백질 쌍의 비는 본 명세서의 개시내용에 기초하여 용이하게 계산될 수 있다. 본 명세서에 개시된 중량 백분율로부터 유도가능한 임의의 중량비의 효소를 포함하는 조성물이 고려된다. β -자일로시다제 함량은 하한치가 블렌드/조성물 중의 효소의 총 중량의 약 0 wt.%, 0.05 wt.%, 0.5 wt.%, 1 wt.%, 2 wt.%, 3 wt.%, 4 wt.%, 5 wt.%, 6 wt.%, 7 wt.%, 8 wt.%, 9 wt.%, 10 wt.%, 12 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 30 wt.%, 40 wt.%, 45 wt.%,

또는 50 wt.%이고, 상한치가 상기 조성물 중의 효소의 총 중량의 약 10 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 30 wt.%, 35 wt.%, 40 wt.%, 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.% 또는 70 wt.%인 범위로 될 수 있다. 예를 들어, β -자일로시다제(들)는 적절하게는 상기 조성물 중의 효소의 총 중량의 약 2 wt.% 내지 약 30 wt.%; 약 10 wt.% 내지 약 20 wt.%; 약 3 wt.% 내지 약 10 wt.%, 또는 약 5 wt.% 내지 약 9 wt.%를 나타낸다.

β -자일로시다제는 β -자일로시다제를 암호화하는 내인성 또는 외인성 유전자를 발현시킴으로써 생성될 수 있다. β -자일로시다제는 경우에 따라, 과발현되거나 저발현될 수 있다. 대안적으로, β -자일로시다제는 숙주 유기체에 대하여 이중일 수 있으며, 숙주 유기체로 재조합에 의해 발현된다. 게다가, β -자일로시다제는 정제되거나 단리된 형태로 본 발명의 셀룰라제 또는 헤미셀룰라제 조성물에 첨가될 수 있다.

L- α -아라비노푸라노시다제

일부 태양에서, 본 발명의 셀룰라제 조성물은 적어도 하나의 L- α -아라비노푸라노시다제를 포함한다. 일부 태양에서, 적어도 하나의 L- α -아라비노푸라노시다제는 Af43A, Fv43B, Pf51A, Pa51A, 및 Fv51A로 이루어지는 그룹으로부터 선택된다. 일부 태양에서, Pa51A, Fv43A는 L- α -아라비노푸라노시다제 활성 및 β -자일로시다제 활성 둘 다를 갖는다.

임의의 적절한 유기체로부터의 L- α -아라비노푸라노시다제(EC 3.2.1.55)는 하나 이상의 L- α -아라비노푸라노시다제로서 사용될 수 있다. 적절한 L- α -아라비노푸라노시다제에는 예를 들어, 아스페르길루스 오리자에(문헌[Numan & Bhosle, J. Ind. Microbiol. Biotechnol. 2006, 33:247-260]), 아스페르길루스 소자에(*A. sojae*)(문헌[Oshima *et al.* J. Appl. Glycosci. 2005, 52:261-265]), 바실루스 브레비스(*B. brevis*)(문헌[Numan & Bhosle, J. Ind. Microbiol. Biotechnol. 2006, 33:247-260]), 바실루스 스테아로써모필루스(*B. stearothermophilus*)(문헌[Kim *et al.*, J. Microbiol. Biotechnol. 2004, 14:474-482]), 비피도박테리움 브레베(*B. breve*)(문헌[Shin *et al.*, Appl. Environ. Microbiol. 2003, 69:7116-7123]), 비피도박테리움 롱굼(*B. longum*)(문헌[Margolles *et al.*, Appl. Environ. Microbiol. 2003, 69:5096-5103]), 클로스트리디움 썬모셀룸(문헌[Taylor *et al.*, Biochem. J. 2006, 395:31-37]), 푸사리움 옥시스포룸(문헌[Panagiotou *et al.*, Can. J. Microbiol. 2003, 49:639-644]), 푸사리움 옥시스포룸 품종 디안티(*F. oxysporum f. sp. dianthi*)(문헌[Numan & Bhosle, J. Ind. Microbiol. Biotechnol. 2006, 33:247-260]), 지오바실루스 스테아로써모필루스 T-6(문헌[Shallom *et al.*, J. Biol. Chem. 2002, 277:43667-43673]), 호데움 불가레(*H. vulgare*)(문헌[Lee *et al.*, J. Biol. Chem. 2003, 278:5377-5387]), 페니실리움 크라이소게눔(문헌[Sakamoto *et al.*, Biophys. Acta 2003, 1621:204-210]), 페니실리움 *sp.*(문헌[Rahman *et al.*, Can. J. Microbiol. 2003, 49:58-64]), 슈도모나스 셀룰로사(*P. cellulosa*)(문헌[Numan & Bhosle, J. Ind. Microbiol. Biotechnol. 2006, 33:247-260]), 리조무코르푸실루스(문헌[Rahman *et al.*, Carbohydr. Res. 2003, 338:1469-1476]), 스트렙토마이세스 카트레우시스(*S. chartreusis*), 스트렙토마이세스 썬모비올라쿠스(*S. thermoviolacus*), 썬모안에어로박터 에탄올리쿠스(*T. ethanolicus*), 썬모바실루스 자일라닐리티쿠스(*T. xylanilyticus*)(문헌[Numan & Bhosle, J. Ind. Microbiol. Biotechnol. 2006, 33:247-260]), 썬모비피다 푸스카(*T. fusca*)(문헌[Tuncer and Ball, Folia Microbiol. 2003, (Praha) 48:168-172]), 썬모토가 마리티마(문헌[Miyazaki, Extremophiles 2005, 9:399-406]), 트리코테르마 *sp.* SY(문헌[Jung *et al.* Agric. Chem. Biotechnol. 2005, 48:7-10]), 아스페르길루스 카와치(*A. kawachii*)(문헌[Koseki *et al.*, Biochim. Biophys. Acta 2006, 1760:1458-1464]), 푸사리움 옥시스포룸 품종 디안티(문헌[Chacon-Martinez *et al.*, Physiol. Mol. Plant Pathol. 2004, 64:201-208]), 썬모바실루스 자일라닐리티쿠스(문헌[Debeche *et al.*, Protein Eng. 2002, 15:21-28]), 후미콜라 인솔렌스, 메리필루스 지잔테우스(*M. giganteus*)(문헌[Sorensen *et al.*, Biotechnol. Prog. 2007, 23:100-107]) 또는 라파누스 사티부스(*R. sativus*)(문헌[Kotake *et al.* J. Exp. Bot. 2006, 57:2353-2362])의 L- α -아라비노푸라노시다제가 포함된다. 적절한 L- α -아라비노푸라노시다제는 숙주 유기체에 의해 내인적으로 생성될 수 있거나, 숙주 유기체에 의해 재조합에 의해 클로닝되고/되거나 발현될 수 있다. 게다가, 적절한 L- α -아라비노푸라노시다제는 정제되거나 단리된 형태로 셀룰라제 조성물에 첨가될 수 있다.

Af43A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Af43A 폴리펩티드를 포함한다. Af43A의 아미노산 서열(서열 번호 20)은 도 17b 및 57에 나타나 있다. 서열 번호 20은 미성숙 Af43A의 서열이다. 예측된 보존 도메인은 도 17b에서 볼드체로 되어 있다. Af43A는 예를 들어, 기질로서 *p*-니트로페닐- α -L-아라비노푸라노시드를 사용하는 효소적 검정법에서 L- α -아라비노푸라노시다제 활성을 갖는 것으로 나타났다. Af43A는 엔도자일라나제의 작용을 통해 헤미셀룰로스로부터 방출되는 올리고머의 세트로부터의 아라비노스의 방출을 촉매하는 것으로 나타났다.

예측된 측매 잔기에는 D26 또는 D58 중 어느 하나, D139 및 E227이 포함된다. 본 명세서에 사용되는 "Af43A 폴리펩티드"는 서열 번호 20의 적어도 50, 75, 100, 125, 150, 175, 200, 250 또는 300개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Af43A 폴리펩티드는 바람직하게는 잔기 D26 또는 D58, D139 및 E227이 고유 Af43A와 비교하여, 변경되지 않는다. Af43A 폴리펩티드는 바람직하게는 도 57의 정렬에서 Af43A 및 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 모두의 다른 아미노산 서열을 포함하는 효소의 그룹 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Af43A 폴리펩티드는 적절하게는 도 17b에 나타난 바와 같은 고유 Af43A의 예측된 보존 도메인을 포함한다. 예시적인 Af43A 폴리펩티드는 서열 번호 20에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다. 본 발명의 Af43A 폴리펩티드는 바람직하게는 L- α -아라비노푸라노시다제 활성을 갖는다.

따라서, 본 발명의 Af43A 폴리펩티드는 적절하게는 서열 번호 20의 아미노산 서열에 대하여 또는 서열 번호 20의 잔기 (i) 15 내지 558, 또는 (ii) 15 내지 295에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 L- α -아라비노푸라노시다제 활성을 갖는다.

Pf51A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Pf51A 폴리펩티드를 포함한다. Pf51A의 아미노산 서열(서열 번호 22)은 도 18b 및 58에 나타나 있다. 서열 번호 22는 미성숙 Pf51A의 서열이다. Pf51A는 서열 번호 22의 잔기 1 내지 22에 해당하는 예측된 신호 서열(도 18b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 22의 잔기 21 내지 642에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 도 18b에서 볼드체로 되어 있다. Pf51A는 예를 들어, 기질로서 4-니트로페닐- α -L-아라비노푸라노시드를 사용하는 효소적 검정법에서 L- α -아라비노푸라노시다제 활성을 갖는 것으로 나타났다. Pf51A는 엔도자일라나제의 작용을 통해 헤미셀룰로스로부터 방출되는 올리고머의 세트로부터의 아라비노스의 방출을 촉매하는 것으로 나타났다. 예측된 보존 산성 잔기에는 E43, D50, E248, E287, E331, E360, E472 및 E480이 포함된다. 본 명세서에 사용되는 "Pf51A 폴리펩티드"는 서열 번호 22의 잔기 21 내지 642 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550 또는 600개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 지칭한다. Pf51A 폴리펩티드는 바람직하게는 잔기 E43, D50, E248, E287, E331, E360, E472 및 E480이 고유 Pf51A와 비교하여, 변경되지 않는다. Pf51A 폴리펩티드는 바람직하게는 도 58의 정렬에 나타난 바와 같이 Pf51A, Pa51A 및 Fv51A 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Pf51A 폴리펩티드는 적절하게는 도 18b에 고유 Pf51A의 예측된 보존 도메인을 포함한다. 예시적인 Pf51A 폴리펩티드는 도 18b에 나타난 성숙 Pf51A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Pf51A 폴리펩티드는 바람직하게는 L- α -아라비노푸라노시다제 활성을 갖는다.

따라서, 본 발명의 Pf51A 폴리펩티드는 적절하게는 서열 번호 22의 아미노산 서열에 대하여 또는 서열 번호 22의 잔기 (i) 21 내지 632, (ii) 461 내지 632, (iii) 21 내지 642, 또는 (iv) 461 내지 642에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 L- α -아라비노푸라노시다제 활성을 갖는다.

Fv51A

일부 태양에서, 본 발명의 셀룰라제 조성물은 Fv51A 폴리펩티드를 포함한다. Fv51A의 아미노산 서열(서열 번호 32)은 도 23b 및 58에 나타나 있다. 서열 번호 32는 미성숙 Fv51A의 서열이다. Fv51A는 서열 번호 32의 잔기 1 내지 19에 해당하는 예측된 신호 서열(도 23b에서 밑줄 그어짐)을 가지며; 신호 서열의 절단에 의해 서열 번호 32의 잔기 20 내지 660에 해당하는 서열을 갖는 성숙 단백질이 제공되는 것으로 예측된다. 예측된 L- α -아라비노푸라노시다제 보존 도메인은 도 23b에서 볼드체로 되어 있다. Fv51A는 예를 들어, 기질로서 4-니트로페닐- α -L-아라비노푸라노시드를 사용하는 효소적 검정법에서 L- α -아라비노푸라노시다제 활성을 갖는 것으로 나타났다. Fv51A는 엔도자일라나제의 작용을 통해 헤미셀룰로스로부터 방출되는 올리고머의 세트로부터의 아라비노스의 방출을 촉매하는 것으로 나타났다. 보존 잔기는 E42, D49, E247, E286, E330, E359, E479 및 E487를

포함한다. 본 명세서에 사용되는 "Fv51A 폴리펩티드"는 서열 번호 32의 잔기 20 내지 660 중에서, 적어도 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 또는 625개의 연속 아미노산 잔기에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 갖는 서열을 포함하는 폴리펩티드 및/또는 그의 변이체를 말한다. Fv51A 폴리펩티드는 바람직하게는 잔기 E42, D49, E247, E286, E330, E359, E479 및 E487이 고유 Fv51A와 비교하여, 변경되지 않는다. Fv51A 폴리펩티드는 바람직하게는 도 58의 정렬에 나타난 바와 같이 Fv51A, Pa51A 및 Pf51A 중에서 보존되는 아미노산 잔기의 적어도 70%, 80%, 90%, 95%, 98% 또는 99%가 변경되지 않는다. Fv51A 폴리펩티드는 적절하게는 도 23b에 나타난 고유 Fv51A의 예측된 보존 도메인을 포함한다. 예시적인 Fv51A 폴리펩티드는 도 23b에 나타난 성숙 Fv51A 서열에 대하여 적어도 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일성을 갖는 서열을 포함한다. 본 발명의 Fv51A 폴리펩티드는 바람직하게는 L- α -아라비노푸라노시다제 활성을 갖는다.

따라서, 본 발명의 Fv51A 폴리펩티드는 적절하게는 서열 번호 32의 아미노산 서열에 대하여 또는 서열 번호 32의 잔기 (i) 21 내지 660, (ii) 21 내지 645, (iii) 450 내지 645, 또는 (iv) 450 내지 660에 대하여 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 아미노산 서열을 포함한다. 폴리펩티드는 적절하게는 L- α -아라비노푸라노시다제 활성을 갖는다.

L- α -아라비노푸라노시다제(들)는 적절하게는 본 개시내용의 셀룰라제 또는 헤미셀룰라제 조성물 중의 효소의 총량의 약 0.05 wt.% 내지 약 30 wt.%(예를 들어, 약 0.1 wt.% 내지 약 25 wt.%, 약 0.5 wt.% 내지 약 20 wt.%, 약 1 wt.% 내지 약 10 wt.%)를 구성하며, 여기서 wt.%는 소정 조성물 중의 모든 효소의 합한 중량에 대하여 L- α -아라비노푸라노시다제(들)의 합한 중량을 나타낸다. L- α -아라비노푸라노시다제(들)는 하한치가 0.05 wt.%, 0.5 wt.%, 1 wt.%, 2 wt.%, 3 wt.%, 4 wt.%, 5 wt.%, 6 wt.%, 7 wt.%, 8 wt.%, 9 wt.%, 10 wt.%, 12 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 또는 28 wt.%이고, 상한치가 5 wt.%, 10 wt.%, 15 wt.%, 20 wt.%, 25 wt.%, 또는 30 wt.%인 범위로 존재할 수 있다. 예를 들어, 하나 이상의 L- α -아라비노푸라노시다제(들)는 적절하게는 본 발명의 셀룰라제 또는 헤미셀룰라제 조성물 중의 효소의 총 중량의 약 2 wt.% 내지 약 30 wt.%(예를 들어, 약 2 wt.% 내지 약 30 wt.%, 약 5 wt.% 내지 약 30 wt.%, 약 5 wt.% 내지 약 10 wt.%, 약 10 wt.% 내지 약 30 wt.%, 약 20 wt.% 내지 약 30 wt.%, 약 25 wt.% 내지 약 30 wt.%, 약 2 wt.% 내지 약 10 wt.%, 약 5 wt.% 내지 약 15 wt.%, 약 10 wt.% 내지 약 25 wt.%, 약 20 wt.% 내지 약 30 wt.% 등)를 구성할 수 있다.

L- α -아라비노푸라노시다제는 L- α -아라비노푸라노시다제를 암호화하는 내인성 또는 외인성 유전자를 발현시킴으로써 생성될 수 있다. L- α -아라비노푸라노시다제는 경우에 따라, 과발현되거나 저발현될 수 있다. 대안적으로, L- α -아라비노푸라노시다제는 숙주 유기체에 대하여 이중될 수 있으며, 숙주 유기체로 재조합에 의해 발현된다. 게다가, L- α -아라비노푸라노시다제는 정제되거나 단리된 형태로 본 발명의 셀룰라제 또는 헤미셀룰라제 조성물에 첨가될 수 있다.

세포 조성물

일부 태양에서, 본 발명은 셀룰라제 활성을 갖는 폴리펩티드를 암호화하는 핵산을 포함하는 세포를 고려한다. 일부 태양에서, 세포는 트리코데르마 리세이 세포이다. 일부 태양에서, 세포는 아스페르길루스 니게르 세포이다. 일부 태양에서, 세포는 임의의 미생물의 세포(예를 들어, 박테리아, 원생생물, 조류, 진균(예를 들어, 효모 또는 사상 진균), 또는 기타 미생물의 세포)를 포함하며, 바람직하게는 박테리아, 효모, 또는 사상 진균의 세포이다. 박테리아 속의 적절한 숙주 세포에는 에스케리키아, 바실루스, 락토바실루스, 슈도모나스 및 스트렙토마이세스의 세포가 포함되나 이들에 한정되지 않는다. 박테리아 종의 적절한 세포에는 에스케리키아 콜라이, 바실루스 서브틸리스, 바실루스 리케니포르미스, 락토바실루스 브레비스, 슈도모나스 아에루지노사, 및 스트렙토마이세스 리비단스의 세포가 포함되나 이들에 한정되지 않는다. 효모 속의 적절한 숙주 세포에는 사카로마이세스, 스킨조사카로마이세스, 칸디다, 한세울라, 피치아, 클루이베로마이세스 및 파피아의 세포가 포함되나 이들에 한정되지 않는다. 효모 종의 적절한 세포에는 사카로마이세스 세레비지에, 스킨조사카로마이세스 폼베, 칸디다 알비칸스, 한세울라 폴리모르파, 피치아 파스토리스, 피치아 카나덴시스, 클루이베로마이세스 마르시아누스 및 파피아 로도지마의 세포가 포함되나 이들에 한정되지 않는다. 사상 진균의 적절한 숙주 세포에는 아문진균류(*Eumycotina*)의 모든 사상형(filamentous form)이 포함된다. 사상 진균 속의 적절한 세포에는 아크레모니움, 아스페르길루스, 아우레오바시디움, 비어칸테라, 세리포리움시스, 크리소스포리움, 코프리누스, 코리올루스, 코리나스쿠스, 카에토미움, 크립토크커스, 필로바시디움, 푸사리움, 지베렐라, 후미콜라, 마그나포르테, 무코르, 마이셀리오프토라, 무코르, 네오칼리마스틱스, 뉴로스포라, 파에실로마이세스, 페니실리움, 파네로차에테, 플레비아, 피로마이세스, 플레우로투스, 사이탈리디움, 스킨조필름, 스포르트릭움, 탈라로마이세

스, 씨모아스쿠스, 티엘라비아, 툴리포클라디움, 트라메테스, 및 트리코테르마의 세포가 포함되나 이들에 한정되지 않는다. 사상 진균 종의 적절한 세포에는 아스페르길루스 아와모리, 아스페르길루스 푸마타투스, 아스페르길루스 포에티더스, 아스페르길루스 야포니쿠스, 아스페르길루스 니들란스, 아스페르길루스 니게르, 아스페르길루스 오리자에, 크리스소스포리움 록크노웬스, 푸사리움 박트리디오이데스, 푸사리움 세레알리스, 푸사리움 크루크웰렌스, 푸사리움 쿨모룸, 푸사리움 그라미네아툼, 푸사리움 그라미눔, 푸사리움 헤테로스포룸, 푸사리움 네균디, 푸사리움 옥시스포룸, 푸사리움 레티쿨라툼, 푸사리움 로세움, 푸사리움 삼부시눔, 푸사리움 사코크로움, 푸사리움 스포로트리키오이데스, 푸사리움 술푸레움, 푸사리움 톨로로숨, 푸사리움 트리코테시오이데스, 푸사리움 베네타툼, 비어칸데라 아두스타, 세리포리움시스 아네이리나, 세리포리움시스 아네이리나, 세리포리움시스 카레지에아, 세리포리움시스 질베슨스, 세리포리움시스 파노신타, 세리포리움시스 리볼로사, 세리포리움시스 수브루파, 세리포리움시스 수브베르미스포라, 코프리누스 시네레우스, 코리올루스 히르수투스, 후미콜라 인솔렌스, 후미콜라 라누지노사, 무코르 미에헤이, 마이셀리오프토라 씨모필라, 뉴로스포라 크라사, 뉴로스포라 인터메디아, 페니실리움 푸르푸로제눔, 페니실리움 카네슨스, 페니실리움 솔리툼, 페니실리움 푸니쿨로숨, 파네로차에테 크리스소스포리움, 플레비아 라디아테, 플레우로투스 에린지이, 탈라로마이세스 플라부스, 티엘라비아 테레스트리스, 트라메테스 빌로사, 트라메테스 베르시콜로르, 트리코테르마하지아눔, 트리코테르마 코닌지이, 트리코테르마 론지브라키아툼, 트리코테르마 리세이, 및 트리코테르마 비리테의 세포가 포함되나 이들에 한정되지 않는다. 일부 태양에서, 세포는 트리코테르마 리세이 세포이다. 일부 태양에서, 세포는 아스페르길루스 니게르 세포이다. 일부 태양에서, 세포는 하나 이상의 헤미셀룰라제를 암호화하는 하나 이상의 핵산을 추가로 포함한다. 일부 태양에서, 세포는 적어도 2개의 베타-글루코시다제의 키메라인 베타-글루코시다제 효소를 포함하는 비천연 셀룰라제 조성물을 포함한다.

일부 태양에서, 본 발명은 서열 번호 60, 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70 wt.%, 75%, 80 wt.%, 85%, 90%, 91 wt.%, 92 wt.%, 93 wt.%, 94 wt.%, 95 wt.%, 96 wt.%, 97 wt.%, 98 wt.%, 99 wt.%) 서열 동일성을 갖는 폴리펩티드를 암호화하는 핵산을 포함하는 세포를 고려한다. 일부 태양에서, 세포는 적어도 하나의 헤미셀룰라제 활성, 예를 들어 β -자일로시다제, L- α -아라비노푸라노시다제, 또는 자일라나제 활성을 갖는 폴리펩티드를 암호화하는 핵산을 추가로 포함한다. 일부 태양에서, 본 발명은 또한 2개 이상의 β -글루코시다제 서열의 키메라를 포함하는 세포를 고려하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 연속 스트레치(stretch)에 대하여 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79로부터 선택되는 하나의 아미노산 서열의 동일한 길이의 연속 스트레치에 대하여 약 60%(예를 들어, 약 65%, 약 65%, 약 70%, 약 75%, 약 80%) 이상의 서열 동일성을 포함한다. 특정 태양에서, 본 발명은 2개 이상의 β -글루코시다제 서열의 키메라 또는 하이브리드를 포함하는 세포를 고려하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79로부터 선택되는 하나의 아미노산 서열의 동일한 길이의 연속 스트레치에 대하여 약 60%(예를 들어, 약 65%, 약 65%, 약 70%, 약 75%, 약 80%) 이상의 서열 동일성을 포함하거나, 서열 번호 164 내지 169의 폴리펩티드 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 연속 스트레치에 대하여 약 60%(예를 들어, 약 65%, 약 65%, 약 70%, 약 75%, 약 80%) 이상의 서열 동일성을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열, 제2 β -글루코시다제 서열, 또는 제1 및 제2 β -글루코시다제 서열 둘 다는 하나 이상의 글리코실화 부위를 포함한다. 특정 실시형태에서, β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역, 또는 루프-유사 구조를 암호화하는 서열을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역을 포함할 수 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다(즉, 키메라 분자의 N-말단에 위치하거나 그 근처에 위치하지 않음, 또는 C-말단에 위치하거나 그 근처에 위치하지 않음).

특정 태양에서, 본 발명은 2개 이상의 β -글루코시다제 서열의 키메라 또는 하이브리드를 포함하는 세포를 고려하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기(예를 들어, 길이가 약 250,

300, 350 또는 400개의 아미노산 잔기)로 되어 있고, 서열 번호 136 내지 148의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하는 한편, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기(예를 들어, 길이가 약 120, 150, 170, 200, 또는 220개의 아미노산 잔기)로 되어 있고, 서열 번호 149 내지 156의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열, 제2 β -글루코시다제 서열, 또는 제1 및 제2 β -글루코시다제 서열 둘 다는 하나 이상의 글리코실화 부위를 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역, 또는 루프-유사 구조를 암호화하는 서열을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역을 포함할 수 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다(즉, 키메라 분자의 N-말단에 위치하거나 그 근처에 위치하지 않음, 또는 C-말단에 위치하거나 그 근처에 위치하지 않음).

발효 브로쓰 조성물

일부 태양에서, 본 발명은 하나 이상의 셀룰라제 활성을 포함하는 발효 브로쓰를 고려하며, 여기서 브로쓰는 바이오매스 시료에 존재하는 셀룰로스를 약 50 wt.%를 초과하여 발효성 당으로 전환시킬 수 있다. 일부 태양에서, 발효 브로쓰는 바이오매스 시료에 존재하는 셀룰로스를 약 55 wt.%를 초과하여(예를 들어, 약 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%를 초과하여) 발효성 당으로 전환시킬 수 있다. 일부 태양에서, 발효 브로쓰는 하나 이상의 헤미셀룰라제 활성을 추가로 포함할 수 있다. 특정 태양에서, 본 발명은 서열 번호 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 83%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 갖는 적어도 하나의 β -글루코시다제 폴리펩티드를 포함하는 발효 브로쓰를 고려한다. 특정 태양에서, 본 발명은 적어도 2개의 β -글루코시다제 서열로 된 키메라인 하이브리드 또는 키메라 β -글루코시다제를 포함하는 발효 브로쓰를 고려한다.

일부 태양에서, 본 발명은 적어도 하나의 β -글루코시다제 활성을 포함하는 발효 브로쓰를 고려하며, 여기서 발효 브로쓰는 바이오매스 시료에 존재하는 셀룰로스를 약 50 wt.%를 초과하여(예를 들어, 약 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.% 또는 80 wt.%) 발효성 당으로 전환시킬 수 있다. 특정 실시형태에서, 발효 브로쓰는 Fv3C 셀룰라제 활성, Pa3D 셀룰라제 활성, Fv3G 활성, Fv3D 활성, Tr3A 활성, Tr3B 활성, Te3A 활성, An3A 활성, Fo3A 활성, Gz3A 활성, Nh3A 활성, Vd3A 활성, Pa3G 활성, 및/또는 Tn3B 활성을 포함하며, 여기서 브로쓰는 바이오매스 시료에 존재하는 셀룰로스를 약 50 wt.%를 초과하여(예를 들어, 약 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 또는 심지어는 80 wt.%를 초과하여) 당으로 전환시킬 수 있다.

일부 태양에서, 본 발명은 2개의 β -글루코시다제 서열로 된 키메라 또는 하이브리드를 포함하는 발효 브로쓰를 고려하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 서열에 대하여 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 하나의 동일한 길이의 서열에 대하여 적어도 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함한다. 일부 태양에서, 본 발명은 2개의 β -글루코시다제 서열로 된 키메라 또는 하이브리드를 포함하는 발효 브로쓰를 고려하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 하나의 동일한 길이의 서열에 대하여 적어도 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 서열에 대하여 적어도 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열, 제2 β -글루코시다제 서열, 또는 제1 및 제2 β -글루코시다제 서열 둘 다는 하나 이상의 글리코실화

부위를 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역, 또는 루프-유사 구조를 암호화하는 서열을 포함한다. 특정 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있거나 연결되어 있다. 일부 실시형태에서, 제1 β -글루코시다제 서열 및 제2 β -글루코시다제 서열은 직접 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 영역을 포함할 수 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다(즉, 키메라 분자의 N-말단 또는 C-말단에 위치하거나 그 근처에 위치하지 않음).

본 발명의 방법

일부 태양에서, 안정성을 향상시키도록 키메라 효소 골격(예를 들어, 엔도글루카나제, 셀로비오하이드롤라제, 및 β -글루코시다제와 같은 셀룰라제, 및 자일라나제, α -아라비노푸라노시다제, β -자일로시다제와 같은 헤미셀룰라제)을 형성하는 방법이 본 명세서에 제공된다. 일부 태양에서, 향상된 안정성은 효소가 적절하게 또는 전형적으로 사용되는 특정한 표준 조건 하에서의 단백질 가수분해에 의한 절단에 대하여 생성된 효소가 덜 민감하다는 점에서 향상된 단백질 가수분해 안정성이다. 일부 태양에서, 단백질 가수분해 안정성은 저장 동안의 안정성에 관한 것인데 반해, 다른 태양에서, 단백질 가수분해 안정성은 발현 및 생성 동안의 안정성에 관한 것으로, 더욱 효과적으로 효소를 생성할 수 있게 한다. 그와 같이, 향상된 안정성은 키메라 효소(즉, 그의 서열 또는 변이체 서열이 키메라 효소의 부분을 구성하는 효소)에 대한 공급원 효소인 비변형된 효소에 비해, 표준 저장 조건 하에서 또는 표준 발현 또는 생성 조건 하에서의 단백질 가수분해에 의한 절단 수준의 감소이다. 일부 태양에서, 향상된 안정성은 발현 및 생성 동안 향상된 저장 안정성 및 향상된 단백질 가수분해 안정성 둘 다에 반영된다. 그와 같이, 향상된 안정성은 저장, 및 발현 및 생성에 관한 표준 조건 하에서의 단백질 가수분해에 의한 절단 수준의 감소이다.

일부 태양에서, 바이오매스를 발효성 당으로 전환시키는데 효과적인 본 명세서에 개시된 임의의 조성물의 양과 바이오매스를 접촉시키는 것을 포함하는, 바이오매스를 당으로 전환시키는 방법이 본 명세서에 제공된다. 일부 태양에서, 바이오매스를 폴리펩티드로 처리하는 것을 포함하는 당화 공정이 본 명세서에 제공되며, 여기서 폴리펩티드는 셀룰라제 활성을 갖고, 상기 공정은 바이오매스를 적어도 약 50 wt.%(예를 들어, 적어도 약 55 wt.%, 적어도 약 60 wt.%, 적어도 약 65 wt.%, 적어도 약 70 wt.%, 적어도 약 75 wt.%, 또는 적어도 약 80 wt.%)로 발효성 당으로 전환시킨다. 일부 태양에서, 본 명세서에 개시된 임의의 조성물을 시판하는 방법이 본 명세서에 제공되며, 여기서 조성물은 에탄올 정제소 또는 기타 생화학물질 또는 바이오물질의 제조처에 공급되거나 판매되며, 임의로 조성물은 상기 에탄올 정제소 또는 기타 생화학물질 또는 바이오물질의 제조처 또는 그 부근에 위치하는 제조 시설에서 제조된다.

키메라 골격의 형성 방법

일부 태양에서, 본 발명은 특정한 β -글루코시다제 폴리펩티드의 향상된 안정성을 제공한다. 특정 태양에서, 향상된 안정성은 예를 들어, 표준 조건 하에서의 β -글루코시다제 폴리펩티드의 보다 낮은 단백질 가수분해 정도 또는 단백질 가수분해에 의한 절단 정도에 반영된 향상된 단백질 가수분해 안정성이며, 여기서 β -글루코시다제 폴리펩티드가 전형적으로 사용된다. 일부 태양에서, 향상된 단백질 가수분해 안정성은 저장, 발현 및/또는 생성 동안의 향상된 안정성이다. 그리하여, 향상된 단백질 가수분해 안정성은 보다 낮은 표준 저장, 발현 및/또는 생성 조건 하에서의 단백질 가수분해에 의한 절단 수준(예를 들어, 활성 손실 정도 또는 수준 감소에 반영된 것)에 반영되며, 여기서 β -글루코시다제 폴리펩티드가 전형적으로 사용되거나 적용된다.

다른 이중 발현된 단백질과 마찬가지로, 특정한 β -글루코시다제는 엑소게나제 프로테아제(exogenase protease)에 의해, 박테리아 또는 진균 숙주 세포에 의해 발현되는 프로테아제, 또는 생성 및 저장 공정 동안의 기타 외부력에 의해 생성 및 저장 동안 단백질 가수분해에 의한 절단이 일어나기 쉽다. 통상적으로, 그러한 단백질 가수분해는 공지된 단백질 가수분해 컨센서스 서열 또는 단백질의 일차 아미노산 서열의 절단 부위를 동정하고 프로테아제가 그 부위에서 단백질을 더 이상 절단할 수 없도록 아미노산을 돌연변이시켜 감소될 수 있다. 이러한 접근법은 폴리펩티드가 하나 초과와 프로테아제에 의해 단백질 가수분해에 의해 절단될 수도 있거나 절단이 효소에 의한 단백질 가수분해의 결과가 아닐 수도 있기 때문에 불리하다. 이러한 접근법은 또한 단백질 가수분해에 의한 절단이 단계적 선호도를 갖는 다수의 부위에서 일어나는 상황에 대응하기에 불충분하다. 예를 들어, 초기 단백질, 예를 들어, 대상으로 하는 β -글루코시다제 폴리펩티드가 단백질 가수분해에 의한 절단 메커니즘

을 통하여 특정 부위에서 초기에 절단될 수 있다. 그러나 일단 초기 절단 부위가 동정, 변형 또는 돌연변이되어, 더 이상 동일한 단백질 가수분해에 의한 절단 메커니즘의 영향을 받기 쉽지 않다면, 동일한 효소는 초기 절단 부위와는 상이한 부위에서 동일하거나 약간 상이한 단백질 가수분해에 의한 절단 메커니즘을 통하여 절단되는 것으로 판명된다. 물론, 제2 부위는 또한 동정, 변형 또는 돌연변이되어 더 이상 단백질 가수분해에 의한 절단의 영향을 받기 쉽지 않지만, 효소는 여전히 또 다른 부위에서 상술한 바와 동일하거나 상이한 메커니즘에 의해 단백질 가수분해에 의한 절단이 일어날 수 있다.

본 발명자들은 이중 발현된 폴리펩티드 상의 절단 부위가 진화적으로 관련된 효소의 이차 구조 간의 비교에 기초하여 동정될 수 있음을 알아냈다. 이중 발현, 생성, 및/또는 저장 동안에 절단되지 않는 관련된 효소의 아미노산 서열 및 예측된 이차 구조의 비교에 의해, 단백질의 이차 구조에 존재하는 루프 서열의 동정을 행할 수 있다. 그러나, 루프 서열은 절단이 일어날 수도 있고 일어나지 않을 수도 있다. 일부 실시형태에서, 실제 단백질 가수분해에 의한 절단은 루프 서열의 다운스트림 또는 업스트림에서 일어날 수 있다. 통상적인 접근법과 마찬가지로, 절단 부위 부근에서 개별 아미노산을 돌연변이시키고/시키거나 개별 아미노산 잔기 또는 잔기를 돌연변이시키기 보다는 오히려, 본 발명은 발현, 생성, 및/또는 저장 동안에 우수한 안정성을 갖는 폴리펩티드를 달성하기 위해 루프 도메인을 변경하는데, 예를 들어, 그러한 루프 도메인을 치환하거나, 아니면 루프 도메인의 길이 및/또는 서열을 변경하는데 접근한다. 특정 실시형태에서, 변경은 예를 들어, 절단되지 않은 진화적으로 관련된 효소와 관련하여 동정된 루프를 제거, 연장, 단축, 또는 치환하는 것을 포함할 수 있다. 게다가, 다수의 이중 발현된 폴리펩티드는 절단되기 쉬운 이차 구조를 제거하기 위해, 이러한 방법이 행해진 다음에, 변경되지 않은 키메라 폴리펩티드에 비해 전반적인 우수한 단백질 가수분해 안정성을 지니는 단일 키메라 골격에 융합될 수 있다. 특정한 아미노산 서열 모티프, 예를 들어, 도 68a에 나타난 것들은 충분한 활성 및 고 성능을 지닌 β -글루코시다제 하이브리드/키메라/융합 분자를 구축하는데 중요할 수 있음을 알 수 있었다.

본 발명자들은 또한 예를 들어, 문헌[Acta Cryst. (2010) D66, 486-501]에 기술된 바와 같이, 통상적인 3차원 효소 구조 틀, 예컨대 "쿠트(Coot)"로 명명되는 모델링 기법을 이용하여, 클리핑에 약하거나 클리핑에 저항성을 나타내는 특정한 GH3 패밀리의 β -글루코시다제의 공지된 3차원 구조를 비교하였다. 예를 들어, Fv3C 및 Te3A 둘 다가 트리코테르마 리세이 Bg11보다, 다수의 셀룰로스 기질에 대한 β -글루코시다제 활성 및 성능이 우수하다는 것을 알아냈다. 또한 Fv3C에 대하여 표준 저장 또는 생성 조건 하에서의 단백질 가수분해에 의한 절단이 일어나면, 이것이 시판용 또는 산업용 효소 조성물의 성분으로서 포함되는 것이 덜 효과적이거나 덜 바람직하게 한다는 것을 알아냈다. 쿠트와 같은 모델링 기법을 이용하여, 트리코테르마 리세이 Bg11과 비교한 Te3A, Fv3C의 공통 특징이 조사되었고, 도 70e에 나타난 바와 같이 4개의 삽입이 관찰되었다. 이러한 삽입으로부터, 잔기 및 아미노산 서열 모티프는 추가로, 도 70f 내지 70j에 나타난 바와 같이, Fv3C 및 Te3A에 존재하나, 트리코테르마 리세이 Bg11에 존재하지 않는 보존된 상호작용(예를 들어, 수소 결합, 글리코실화 부위)을 나타내는 것으로 밝혀졌다. 따라서, 도 68b에 나타난 것을 비롯하여 특정한 아미노산 서열 모티프가, 소정의 천연 β -글루코시다제, 또는 그의 돌연변이체, 또는 그의 하이브리드/키메라/융합 분자가 향상된 성능/활성 및 안정성을 갖는지의 여부를 결정하는 키인 것을 알아냈다.

이론에 구속되는 것은 아니지만, 향상된 단백질 안정성은 효소 활성을 감소시킬 수 있다. 효소 활성 감소는 바람직하게는 20% 미만, 더욱 바람직하게는 15% 미만, 더욱더 바람직하게는 10% 미만이다. 따라서, 효소, 예를 들어, 셀룰라제 효소 또는 헤미셀룰라제 효소의 루프 서열을 변경시켜 단백질 안정성을 향상시키는 방법이 본 명세서에 제공된다. 특정 실시형태에서, 루프 서열 자체는 단백질 가수분해에 의한 절단에 영향을 받기 쉽다. 다른 실시형태에서, 루프 서열 자체는 단백질 가수분해에 의한 절단에 영향을 받기는 쉽지 않지만, 루프 서열의 변경은 효소의 루프 서열로부터의 업스트림 또는 다운스트림 부위에서의 절단에 영향을 미칠 수 있다.

특정 실시형태에서, 루프 서열은 하이브리드 또는 키메라 효소, 예를 들어, 하이브리드 또는 키메라 β -글루코시다제에 존재하며, 이는 각각 상이한 β -글루코시다제로부터 유래되는 2개 이상의 β -글루코시다제 서열을 포함한다. 예를 들어, 하이브리드 또는 키메라 β -글루코시다제는 2개의 β -글루코시다제 서열을 포함할 수 있으며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 가지며, 제2 β -글루코시다제는 길이가 적어도 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 갖는다. 다른 예에서, 하이브리드 또는 키메라 β -글루코시다제는 2개의 β -글루코시다제 서열을 포함할 수 있으며, 여기서 제1 β -글루코시다제 서열은 길이가

적어도 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60% (예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 가지며, 제2 β -글루코시다제는 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 서열에 대하여 적어도 약 60% (예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 갖는다. 일부 실시형태에서, 길이가 적어도 약 200개의 아미노산 잔기로 된 제1 β -글루코시다제 서열은 하이브리드 효소의 N-말단에 있는 한편, 길이가 적어도 약 50개의 아미노산 잔기로 된 제2 β -글루코시다제 서열은 하이브리드 효소의 C-말단에 있다. 특정 실시형태에서, N-말단 또는 C-말단 β -글루코시다제 서열은 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있다. 특정 실시형태에서, N-말단 및 C-말단 β -글루코시다제 서열은 서로 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, N-말단 및 C-말단 β -글루코시다제 서열은 서로 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다. 일부 실시형태에서, 링커 도메인은 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열의 변경, 예를 들어, 루프 서열의 연장, 단축, 돌연변이, 결실(전체적으로 또는 부분적으로) 또는 치환에 의해, 생성된 하이브리드 또는 키메라 효소가 단백질 가수분해에 의한 절단에 좌우되지 않게 된다. 그리하여, 생성된 폴리펩티드 또는 키메라 폴리펩티드는 바람직하게는 그의 고유 대응물에 비해 향상된 안정성을 달성한다(예를 들어, 키메라 폴리펩티드인 경우에, 고유 대응물은 각 키메라 부분이 유래되는 고유 효소를 말함). 향상된 안정성은 표준 저장, 발현, 생성, 또는 사용 조건 동안에 분해 산물의 수준이 감소되거나 더 적어짐으로써 반영될 수 있다.

이중 발현된 폴리펩티드 및 키메라 폴리펩티드의 향상된 안정성은 저장, 발현 또는 기타 생성 공정 동안의 단백질 가수분해 안정성의 향상, 및 그러한 폴리펩티드가 사용되는 공정의 개선을 테스트하여 결정될 수 있다.

특정 실시형태에서, 루프 서열은 하이브리드 또는 키메라 효소, 예를 들어, 하이브리드 또는 키메라 β -글루코시다제에 존재하며, 이는 각각 상이한 β -글루코시다제로부터 유래되는 2개 이상의 β -글루코시다제 서열을 포함한다. 예를 들어, 하이브리드 또는 키메라 β -글루코시다제는 2개의 β -글루코시다제 서열을 포함할 수 있으며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 200개의 아미노산 잔기로 되어 있고, 서열 번호 136 내지 148의 아미노산 서열 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제는 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 149 내지 156의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함한다. 특히, 2개 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개, 또는 모두)를 포함하며, 2개 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있고, 서열 번호 170의 서열을 포함한다. 일부 실시형태에서, 길이가 적어도 약 200개의 아미노산 잔기로 된 제1 β -글루코시다제 서열은 하이브리드 효소의 N-말단에 있는 한편, 길이가 적어도 약 50개의 아미노산 잔기로 된 제2 β -글루코시다제 서열은 하이브리드 효소의 C-말단에 있다. 특정 실시형태에서, N-말단 또는 C-말단 β -글루코시다제 서열은 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있다. 특정 실시형태에서, N-말단 및 C-말단 β -글루코시다제 서열은 서로 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, N-말단 및 C-말단 β -글루코시다제 서열은 서로 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다. 일부 실시형태에서, 링커 도메인은 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열의 변경, 예를 들어, 루프 서열의 연장, 단축, 돌연변이, 결실(전체적으로 또는 부분적으로) 또는 치환에 의해, 생성된 하이브리드 또는 키메라 효소가 단백질 가수분해에 의한 절단에 좌우되지 않게 된다. 그리하여, 생성된 폴리펩티드 또는 키메라 폴리펩티드는 바람직하게는 그의 고유 대응물에 비해 향상된 안정성을 달성한다(예를 들어, 키메라 폴리펩티드인 경우에, 고유 대응물은 각 키메라 부분이 유래되는 고유 효소를 말함). 향상된 안정성은 표준 저장, 발현, 생성, 또는 사용 조건 동안에 분해 산물의 수준이 감소되거나 더 적어짐으로써 반영될 수 있다.

일부 태양에서, 루프 서열은 하이브리드 또는 키메라 효소, 예를 들어, 하이브리드 또는 키메라 β -글루코시다제에 존재하며, 이는 2개 이상의 효소 서열을 포함하고, 여기서 적어도 하나는 β -글루코시다제 서열인 반면에, 다른 하나는 다른 효소의 서열도 아니고, 하나의 β -글루코시다제도 아니다. 예를 들어, 키메라 효소의 적어도 하나의 키메라 부분이 유래되는 비- β -글루코시다제 서열은 다른 헤미셀룰라제 또는 셀룰라제, 예를 들어, 자일라나제, 엔도글루카나제, 자일로시다제, 아라비노푸라노시다제 등으로부터 선택될 수 있다. 키메라 폴리펩티드의 N-말단 도메인 및 C-말단 도메인은 서로 직접 인접할 수 있다. 대안적으로, N-말단 도메인 및 C-말단 도메인

인은 직접 인접하거나 연결되어 있지 않지만, 링커 서열을 통하여 연결되어 있다. 특정 실시형태에서, N-말단 또는 C-말단 β -글루코시다제 서열은 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있다. 특정 실시형태에서, 링커 도메인은 중앙에 위치한다. 일부 실시형태에서, 링커 도메인은 루프 서열을 포함한다. 특정 실시형태에서, 루프 서열의 변경, 예를 들어, 루프 서열의 연장, 단축, 돌연변이, 결실(전체적으로 또는 부분적으로) 또는 치환에 의해, 생성된 하이브리드 또는 키메라 효소가 단백질 가수분해에 의한 절단에 좌우되지 않게 된다. 그리하여, 생성된 폴리펩티드 또는 키메라 폴리펩티드는 바람직하게는 그의 고유 대응물에 비해 향상된 안정성을 달성한다(예를 들어, 키메라 폴리펩티드인 경우에, 고유 대응물은 각 키메라 부분이 유래되는 고유 효소를 말함). 향상된 안정성은 표준 저장, 발현, 생성, 또는 사용 조건 동안에 분해 산물의 수준이 감소되거나 더 적어짐으로써 반영될 수 있다. 특정 실시형태에서, 키메라 또는 하이브리드 폴리펩티드는 이중 셀룰라제 및/또는 헤미셀룰라제 활성을 가질 수 있다. 예를 들어, 본 발명의 키메라 또는 하이브리드 폴리펩티드는 β -글루코시다제 활성 및 자일라나제 활성 둘 다를 가질 수 있다. 일부 실시형태에서, 키메라 또는 하이브리드 폴리펩티드는 그의 키메라 부분의 고유 대응물에 비해 향상된 안정성을 가질 수 있다. 예를 들어, 변경된 루프 서열을 포함하는 키메라 β -글루코시다제-자일라나제 폴리펩티드는 키메라 폴리펩티드가 그의 β -글루코시다제 서열 및 그의 자일라나제 서열로부터 유래되는 β -글루코시다제 및 자일라나제에 비해 표준 저장, 발현, 생성 또는 사용 조건 하에서 향상된 안정성, 예를 들어, 향상된 단백질 가수분해 안정성을 가질 수 있다.

일부 태양에서, 본 발명은 셀룰라제 또는 헤미셀룰라제 효소의 안정성을 향상시키는 방법에 관한 것이며, 여기서 안정성은 예를 들어, 표준 저장, 발현, 생성, 또는 사용 조건 하에서 5% 이상, 10% 이상, 15% 이상, 20% 이상, 25% 이상, 또는 심지어는 30% 이상 향상된다. 안정성 향상은 특정한 표준 저장, 발현, 생성 또는 사용 조건에서 일정 기간 후에 절단되는 이러한 효소의 양을 측정함으로써 측정될 수 있다. 예를 들어, 안정성 향상은 예를 들어, 약 1 (예를 들어, 약 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 20, 24) 시간 이상에서 표준 저장 조건 하에서, 예를 들어, 주위 온도에서 또는 약 40°C, 45°C, 50°C의 고온에서, 또는 더욱더 높은 온도에서 분해 산물의 양에 의해 측정될 수 있다. 특정 실시 형태에서, 안정성 향상은 예를 들어, 약 1(예를 들어, 약 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 20, 24) 시간 이상에서 표준 생성 조건 하에서, 예를 들어, 50°C 초과(예를 들어, 50°C 초과, 55°C 초과, 60°C 초과, 또는 심지어는 65°C 초과)의 온도에서 잔존하는 온전한 생성물의 양을 검출하여 측정함으로써 측정될 수 있다.

바이오매스를 당으로 전환시키는 방법

일부 태양에서, 바이오매스를 발효성 당으로 전환시키는데 효과적인 본 명세서에 개시된 임의의 조성물의 양과 바이오매스를 접촉시키는 것을 포함하는, 바이오매스를 당으로 전환시키는 방법이 본 명세서에 제공된다. 일부 태양에서, 상기 방법은 바이오매스를 산 및/또는 염기로 전처리하는 것을 추가로 포함한다. 일부 태양에서, 산은 인산을 포함한다. 일부 태양에서, 염기는 수산화나트륨 또는 암모니아를 포함한다.

바이오매스

본 개시내용은 본 개시내용의 셀룰라제 또는 비천연 헤미셀룰라제 조성물을 이용한 바이오매스 당화의 방법 및 공정을 제공한다. 본 명세서에 사용되는 용어 "바이오매스"는 셀룰로스 및/또는 헤미셀룰로스(또한, 임의로 리그노셀룰로스를 바이오매스 물질 중의 리그닌)를 포함하는 임의의 조성물을 말한다. 본 명세서에 사용되는 바이오매스는 제한 없이, 종자, 낱알, 덩이줄기, 식품 가공 또는 생산 가공의 식물 폐기물 또는 부산물(예를 들어, 줄기), 옥수수(예를 들어, 옥수수 속대, 옥수수 대 등 포함), 목초(예를 들어, 인디안 그래스, 예컨대 소 르카스트럼 누탄스; 또는, 스위치그래스, 예를 들어, 파니쿰 종, 예컨대 파니쿰 비르가툼), 다년생 줄기(예를 들어, 물대), 목재(예를 들어, 목재 칩, 가공 폐기물 포함), 종이, 펄프 및 재생지(예를 들어, 신문지, 인쇄 용지 등 포함)를 포함한다. 다른 바이오매스 물질은 제한 없이, 감자, 대두(예를 들어, 평지씨), 보리, 호밀, 귀리, 밀, 비트(beet) 및 사탕수수 바가스(sugar cane bagasse)를 포함한다.

본 개시내용은 바이오매스 물질, 예를 들어, 자일란, 헤미셀룰로스, 셀룰로스 및/또는 발효성 당을 포함하는 물질을 포함하는 조성물을 본 개시내용의 폴리펩티드, 또는 본 개시내용의 핵산에 의해 암호화되는 폴리펩티드, 또는 본 개시내용의 셀룰라제 또는 비천연 헤미셀룰라제 조성물 또는 제조 제품 중 임의의 것과 접촉시키는 것을 포함하는 당화 방법을 제공한다.

당화된 바이오매스(예를 들어, 본 개시내용의 효소에 의해 처리되는 리그노셀룰로스 물질)는 예를 들어, 미생물 발효 및/또는 화학적 합성과 같은 공정을 통하여 수많은 바이오-기반의 제품으로 제조될 수 있다. 본 명세서에

사용되는 "미생물 발효"는 적절한 조건 하에서 발효 미생물을 성장시키고 수집하는 공정을 말한다. 발효 미생물은 바이오-기반의 제품의 제조를 위한 원하는 발효 공정에 사용하기에 적합한 임의의 미생물일 수 있다. 적절한 발효 미생물에는 제한 없이, 사상 진균, 효모 및 박테리아가 포함된다. 당화된 바이오매스는 예를 들어, 발효 및/또는 화학적 합성을 통해 그것이 연료(예를 들어, 바이오에탄올, 바이오부탄올, 바이오메탄올, 바이오프로판올, 바이오디젤, 제트 연료 등)로 제조될 수 있다. 또한, 당화된 바이오매스는 예를 들어, 발효 및/또는 화학적 합성을 통하여, 범용 화학물질(예를 들어, 아스코르브산, 아이소프렌, 1,3-프로판다이올), 지질, 아미노산, 단백질 및 효소로 제조될 수 있다.

전처리

당화 전에, 바이오매스(예를 들어, 리그노셀룰로스 물질)은 바람직하게는 하나 이상의 전처리 단계(들)로 처리되어, 자일란, 헤미셀룰로스, 셀룰로스 및/또는 리그닌 물질이 효소에 더욱 접근가능하거나 민감하게 하며, 이에 따라 본 개시내용의 효소(들) 및/또는 셀룰라제 또는 비천연 헤미셀룰라제 조성물에 의해 더욱 가수분해될 수 있게 한다.

예시적인 실시형태에서, 전처리는 바이오매스 물질을 반응기 내에서 강산 및 금속염의 희석 용액을 포함하는 촉매로 처리하는 것을 수반한다. 바이오매스 물질은 예를 들어, 원료 또는 건조 물질일 수 있다. 이러한 전처리는 셀룰로스 가수분해의 활성화 에너지 또는 온도를 낮추어, 궁극적으로 발효성 당의 수율이 보다 높아지게 할 수 있다. 예를 들어, 미국 특허 제6,660,506호; 제6,423,145호를 참조한다.

다른 예시적인 전처리 방법은 바이오매스 물질을 셀룰로스의 글루코스로의 상당한 해중합을 달성하지 않고 주로 헤미셀룰로스의 해중합을 유발하기 위해 선택된 온도 및 압력에서 수성 매질 중에서 제1 가수분해 단계로 처리함에 의한 바이오매스의 가수분해를 포함한다. 이러한 단계에 의해, 슬러리가 제공되며, 여기서, 액체 수상은 헤미셀룰로스의 해중합으로부터 야기되는 용해된 단당류를 함유하며, 고체상은 셀룰로스와 리그닌을 함유한다. 이어서, 슬러리를 상당 부분의 셀룰로스가 해중합되게 하는 조건 하에서 제2 가수분해 단계로 처리하여, 셀룰로스의 용해된/가용성 해중합 산물을 함유하는 액체 수상을 제공한다. 예를 들어, 미국 특허 제5,536,325호를 참조한다.

추가 예시적인 방법은 바이오매스 물질을 약 0.4% 내지 약 2%의 강산을 사용한 묽은 산 가수분해의 하나 이상의 단계로 처리하고; 이어서, 산에 의해 가수분해된 물질의 미반응된 고체 리그노셀룰로스 성분을 알칼리 탈리그닌화(alkaline delignification)로 처리하는 것을 포함한다. 예를 들어, 미국 특허 제6,409,841호를 참조한다.

다른 예시적인 전처리 방법은 전가수분해 반응기에서 바이오매스(예를 들어, 리그노셀룰로스 물질)를 전가수분해하고; 산성 액체를 고체 리그노셀룰로스 물질에 첨가하여, 혼합물을 제조하고; 혼합물을 반응 온도로 가열하고; 반응 온도를 리그노셀룰로스 물질을 리그노셀룰로스 물질로부터의 적어도 약 20%의 리그닌을 함유하는 가용성 부분 및 셀룰로스를 함유하는 고체 분획으로 분별시키기에 충분한 기간 동안 유지하고; 반응 온도에서 또는 반응 온도 근처에서, 고체 분획으로부터 가용성 부분을 분리하여, 가용성 부분을 제거하고; 가용성 부분을 회수하는 것을 포함한다. 고체 분획 중의 셀룰로스가 효소에 의해 더욱 분해될 수 있게 된다. 예를 들어, 미국 특허 제5,705,369호를 참조한다.

추가 전처리 방법은 과산화수소, H_2O_2 의 사용을 수반할 수 있다. 문헌[Gould, 1984, Biotech, and Bioengr. 26:46-52]을 참조한다.

또한, 전처리는 바이오매스 물질을 매우 낮은 농도에서 화학량론적 양의 수산화나트륨 및 수산화암모늄과 접촉시키는 것을 포함할 수 있다. 문헌[Teixeira *et al.*, 1999, Appl. Biochem. and Biotech. 77-79:19-34]을 참조한다.

또한, 전처리는 리그노셀룰로스를 약 9 내지 약 14의 pH에서, 적당한 온도, 압력 및 pH에서, 화학물질(예를 들어, 염기, 예를 들어, 탄산나트륨 또는 수산화칼륨)과 접촉시키는 것을 포함할 수 있다. 국제 특허 공개 제W02004/081185호를 참조한다.

예를 들어, 바람직한 전처리 방법에서 암모니아가 사용된다. 이러한 전처리 방법은 바이오매스 물질을 높은 고형분의 조건 하에서 낮은 암모니아 농도로 처리하는 것을 포함한다. 예를 들어, 미국 특허 공개 제20070031918호 및 국제 특허 공개 제W0 06110901호를 참조한다.

당화 공정

일부 태양에서, 바이오매스를 폴리펩티드로 처리하는 것을 포함하는 당화 공정이 본 명세서에 제공되며, 여기서 폴리펩티드는 셀룰라제 활성을 갖고, 상기 공정은 적어도 바이오매스를 약 50 wt.%(예를 들어, 적어도 약 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 또는 80 wt.%)로 발효성 당으로 전환시킨다. 일부 태양에서, 바이오매스는 리그닌을 포함한다. 일부 태양에서, 바이오매스는 셀룰로스를 포함한다. 일부 태양에서, 바이오매스는 헤미셀룰로스를 포함한다. 일부 태양에서, 셀룰로스를 포함하는 바이오매스는 자일란, 갈락탄, 또는 아라비난 중 하나 이상을 추가로 포함한다. 일부 태양에서, 바이오매스는 제한 없이, 종자, 낱알, 덩이줄기, 식품 가공 또는 생산 가공의 식물 폐기물 또는 부산물(예를 들어, 줄기), 옥수수(예를 들어, 옥수수 속대, 옥수수 대 등 포함), 목초(예를 들어, 인디안 그래스, 예컨대 소르카스트럼 누탄스; 또는, 스위치그래스, 예를 들어, 파니쿰 종, 예컨대 파니쿰 비르가툼), 다년생 줄기(예를 들어, 물대), 목재(예를 들어, 목재 칩, 가공 폐기물 포함), 종이, 펄프 및 재생지(예를 들어, 신문지, 인쇄 용지 등 포함), 감자, 대두(예를 들어, 평지씨), 보리, 호밀, 귀리, 밀, 비트 및 사탕수수 바가스를 포함한다. 일부 태양에서, 바이오매스를 포함하는 물질은 폴리펩티드로 처리하기 전에, 산 및/또는 염기로 처리된다. 일부 태양에서, 산은 인산이다. 일부 태양에서, 염기는 암모니아 또는 수산화나트륨이다. 일부 태양에서, 당화 공정은 바이오매스를 셀룰라제 및/또는 헤미셀룰라제로 처리하는 것을 추가로 포함한다. 일부 태양에서, 바이오매스는 전체 셀룰라제로 처리된다. 일부 태양에서, 당화 공정은 적어도 약 50 wt.%, 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%의 바이오매스를 당으로 전환시킨다. 일부 태양에서, 셀룰라제 조성물 또는 헤미셀룰라제 조성물은 하이브리드 또는 키메라 β -글루코시다제 효소인 폴리펩티드를 포함하며, 이는 적어도 2개의 β -글루코시다제 서열로 된 키메라이다.

일부 태양에서, 바이오매스를 폴리펩티드를 포함하는 조성물로 처리하는 것을 포함하는 당화 공정이 제공되며, 여기서 폴리펩티드는 서열 번호 60, 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 가지고, 상기 공정은 바이오매스를 적어도 약 50 wt.%(예를 들어, 적어도 약 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%)로 발효성 당으로 전환시킨다. 일부 태양에서, 서열 번호 60, 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) 서열 동일성을 갖는 폴리펩티드로 바이오매스를 처리하는 것을 포함하는 당화 공정은 바이오매스를 적어도 약 60 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%로 당으로 전환시킨다. 일부 태양에서, 서열 번호 60, 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79 중 어느 하나에 대하여 적어도 80%, 적어도 90%, 적어도 95%, 또는 적어도 97% 서열 동일성을 갖는 폴리펩티드로 처리하기 전에, 바이오매스를 포함하는 물질을 산 및/또는 염기로 처리한다. 일부 태양에서, 산은 인산이다.

일부 태양에서, 적어도 2개의 β -글루코시다제 서열로 된 키메라 또는 하이브리드인 β -글루코시다제를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하는 당화 공정이 제공된다.

일부 태양에서, 당화 공정은 적어도 2개의 β -글루코시다제 서열로 된 키메라를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 동일한 길이의 Fv3C의 아미노산 서열(서열 번호 60)로 된 서열에 대하여 약 60%(예를 들어, 약 65%, 70%, 75%, 또는 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 68, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79로부터 선택되는 아미노산 서열 중 동일한 길이의 하나의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 또는 80%) 서열 동일성을 포함한다. 일부 태양에서, 당화 공정은 적어도 2개의 β -글루코시다제 서열로 된 키메라를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기이고, 서열 번호 54, 56, 68, 62, 64, 66, 68, 70, 72, 74, 76, 78, 또는 79로부터 선택되는 아미노산 서열 중 어느 하나의 아미노산 서열로 된 동일한 길이의 서열에 대하여 약 60%(예를 들어, 약 65%, 70%, 75%, 또는 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 동일한 길이의 서열 번호 60의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 또는 80%) 서열 동일성을 포함한다. 일부 태양에서, 당화 공정은 적어도 2개의 β -글루코시다제 서열로 된 키메라를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하며, 여기서 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 136

내지 148의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 149 내지 156의 아미노산 서열 모티프 중 하나 이상 또는 모두를 포함한다. 특히, 둘 이상의 β -글루코시다제 서열 중 제1 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있으며, 서열 번호 164 내지 169의 아미노산 서열 모티프 중 적어도 2개(예를 들어, 적어도 2, 3, 4개 또는 모두)를 포함하는 것이며, 둘 이상의 β -글루코시다제 중 제2 서열은 길이가 적어도 50개의 아미노산 잔기로 되어 있으며, 서열 번호 170을 포함한다. 일부 실시형태에서, 제1 β -글루코시다제 서열은 하이브리드 또는 키메라 폴리펩티드의 N-말단에 존재하고, 제2 β -글루코시다제 서열은 하이브리드 또는 키메라 폴리펩티드의 C-말단에 존재한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 서로 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 특정 태양에서, 제1 또는 제2 β -글루코시다제 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 되어 있는 루프 서열을 포함한다. 일부 실시형태에서, 루프 서열은 하이브리드 또는 키메라 효소가 루프 서열의 부위 또는 루프 서열의 외측에 있는 잔기에서 단백질 가수분해에 의한 절단에 좌우되지 않도록 변경된다. 특정 실시형태에서, 제1 또는 제2 β -글루코시다제 중 어느 것도 루프 서열을 포함하지 않지만, 링커 도메인은 루프 서열을 포함한다. 일부 실시형태에서, 링커 도메인은 하이브리드 또는 키메라 폴리펩티드의 중앙에 위치한다. 일부 태양에서, 바이오매스를 포함하는 물질은 적어도 2개의 β -글루코시다제로 된 키메라를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 처리되기 전에, 산 및/또는 염기로 처리된다. 일부 태양에서, 산은 인산이다. 일부 태양에서, 염기는 암모니아 또는 수산화나트륨이다. 일부 태양에서, 당화 공정은 바이오매스를 헤미셀룰라제로 처리하는 것을 추가로 포함한다. 일부 태양에서, 바이오매스는 전체 셀룰라제로 처리된다. 일부 태양에서, 적어도 2개의 β -글루코시다제 서열로 된 키메라 또는 하이브리드를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하는 당화 공정 - 여기서, 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 60의 동일한 길이의 서열에 대하여 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79로부터 선택되는 아미노산 서열 중 어느 하나의 동일한 길이의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 약 70%, 약 75%, 또는 약 80%)의 서열 동일성을 포함함 - 은 바이오매스를 적어도 약 50 wt.%, 60 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%로 당으로 전환시킨다. 일부 태양에서, 적어도 2개의 β -글루코시다제 서열로 된 키메라 또는 하이브리드를 포함하는 비천연 셀룰라제 조성물 또는 헤미셀룰라제 조성물로 바이오매스를 처리하는 것을 포함하는 당화 공정 - 여기서, 제1 β -글루코시다제 서열은 길이가 적어도 약 200개의 아미노산 잔기로 되어 있고, 서열 번호 54, 56, 58, 62, 64, 66, 68, 70, 72, 74, 76, 78, 및 79로부터 선택되는 아미노산 서열 중 어느 하나의 동일한 길이의 서열에 대하여 약 60%(예를 들어, 약 65%, 약 70%, 약 75%, 또는 약 80%) 이상의 서열 동일성을 포함하며, 제2 β -글루코시다제 서열은 길이가 적어도 약 50개의 아미노산 잔기로 되어 있고, 서열 번호 60의 동일한 길이의 서열에 대하여 적어도 약 60%(예를 들어, 적어도 약 65%, 70%, 75%, 또는 80%)의 서열 동일성을 포함함 - 은 바이오매스를 적어도 약 50 wt.%, 60 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%로 당으로 전환시킨다. 일부 태양에서, 제1 β -글루코시다제 서열은 키메라 또는 하이브리드 β -글루코시다제 폴리펩티드의 N-말단에 존재하며, 제2 β -글루코시다제 서열은 그의 C-말단에 존재한다. 특정 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있거나 직접 연결되어 있다. 다른 실시형태에서, 제1 및 제2 β -글루코시다제 서열은 바로 인접해 있지 않지만, 링커 도메인을 통하여 연결되어 있다. 일부 태양에서, 제1 또는 제2 β -글루코시다제 서열은 루프 서열을 포함하며, 여기서 루프 서열은 FDRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 구성되고, 루프 서열의 변경은 향상된 안정성을 가져오며, 이는 보다 적은 정도의 하이브리드 또는 키메라 폴리펩티드의 절단 또는 분해에 의해 반영될 수 있다. 특정 실시형태에서, 향상된 안정성은 루프 서열 잔기에 서의 절단의 감소 또는 제거에 의해 반영된다. 일부 실시형태에서, 향상된 안정성은 루프 영역 외측의 잔기에

서의 절단의 감소 또는 제거에 의해 반영된다. 특정 실시형태에서, 제1 또는 제2 β -글루코시다제 서열 중 어느 것도 루프 영역을 포함하지 않는 반면에, 링커 도메인은 FDRRSPG의 서열(서열 번호 171), 또는 FD(R/K)YNIT의 서열(서열 번호 172)을 포함하는, 길이가 약 3, 4, 5, 6, 7, 8, 9, 10, 또는 11개의 아미노산 잔기로 된 루프 서열을 포함한다. 일부 실시형태에서, 당화 공정은 바이오매스를 적어도 약 50 wt.%, 60 wt.%, 70 wt.%, 75 wt.%, 80 wt.%, 85 wt.%, 또는 90 wt.%로 당으로 전환시킨다.

비즈니스 방법

본 개시내용의 셀룰라제 및/또는 헤미셀룰라제 조성물은 산업 및/또는 상업적 환경에서 추가로 사용될 수 있다. 따라서, 본 발명의 셀룰라제 및 비천연 헤미셀룰라제 조성물의 제조, 시판 또는 다르게는 상업화 방법도 또한 고려된다.

구체적인 실시형태에서, 본 발명의 셀룰라제 및 비천연 헤미셀룰라제 조성물은 특정 에탄올(바이오에탄올) 정제소 또는 기타 생화학물질 또는 바이오물질의 제조처에 공급되거나 판매될 수 있다. 제1 예에서, 비천연 셀룰라제 및/또는 헤미셀룰라제 조성물은 산업적 규모의 효소의 제조를 전문으로 하는 효소 제조 시설에서 제조될 수 있다. 이어서, 비천연 셀룰라제 및/또는 헤미셀룰라제 조성물은 포장되거나 효소 제조처의 고객에게 판매될 수 있다. 이러한 운영 전략은 본 명세서에서 "머천트(merchant) 효소 공급 모델"로 지칭된다.

다른 운영 전략에서, 본 발명의 비천연 셀룰라제 및/또는 헤미셀룰라제 조성물은 바이오에탄올 정제소 또는 생화학물질/바이오물질 제조처에 또는 그 부근에 위치한 소정의 위치("온-사이트")에서, 효소 제조처에 의해 구축된 최신식의 효소 생성 시스템에서 생성될 수 있다. 일부 실시형태에서, 효소 공급 협정은 효소 제조처 및 바이오에탄올 정제소 또는 생화학물질/바이오물질 제조처에 의해 시행된다. 효소 제조처는 본 명세서에 기재된 바와 같은 숙주 세포, 발현 및 생성 방법을 사용하여, 비천연 셀룰라제 및/또는 헤미셀룰라제 조성물을 제조하도록, 온-사이트 효소 생성 시스템을 설계하고, 제어하고, 운영한다. 특정 실시형태에서, 바람직하게는 본 명세서에 기재된 바와 같은 적절한 전처리로 처리되는 적절한 바이오매스는 바이오에탄올 정제소 또는 생화학물질/바이오물질 제조 시설에서 또는 그 근처에서, 당화 방법 및 본 명세서의 효소 및/또는 효소 조성물을 사용하여 가수분해될 수 있다. 생성된 발효성 당은 이어서 동일한 시설에서 또는 부근의 시설에서 발효될 수 있다. 이러한 운영 전략은 본 명세서에서 "온-사이트 바이오리파이너리(biorefinery) 모델"로 지칭된다.

온-사이트 바이오리파이너리 모델은 예를 들어, 머천트 효소 공급자로부터의 효소 공급에 대한 의존성을 최소로 하는 자급 자족 운영의 공급을 포함하여, 머천트 효소 공급 모델에 비해 소정의 이점을 제공한다. 이는 결국, 실시간 또는 거의 실시간의 요구에 기초하여, 바이오에탄올 정제소 또는 생화학물질/바이오물질 제조처가 더 나은 효소 공급 제어를 가능하게 한다. 특정 실시형태에서, 온-사이트 효소 생성 시설이 서로 근접하게 위치한 2개의 바이오에탄올 정제소 및/또는 생화학물질/바이오물질 제조처 간에 또는 2개 이상의 바이오에탄올 정제소 및/또는 생화학물질/바이오물질 제조처 중에 공유되어, 효소 운반 및 저장 비용을 저감시킬 수 있는 것으로 고려된다. 추가로, 이는 온-사이트 효소 생성 시설에서 더욱 즉각적인 "드롭-인(drop-in)" 기술 향상을 가능하게 하여, 효소 조성물의 향상 간의 시간 지연을 줄여, 보다 높은 수율의 발효성 당, 궁극적으로 바이오에탄올 또는 생화학물질에 이른다.

온-사이트 바이오리파이너리 모델은 바이오에탄올 및 생화학물질의 산업적 생성 및 상업화에서 더욱 일반적인 적용가능성을 갖는데, 이는 온-사이트 바이오리파이너리 모델이 본 개시내용의 셀룰라제 및 비천연 헤미셀룰라제 조성물뿐 아니라, 전분(예를 들어, 옥수수)을 처리하는 효소 및 효소 조성물을 제조, 공급, 및 생산하는데 사용하여, 전분의 바이오에탄올 또는 생화학물질로의 직접적인 전환을 더욱 효율적이며 효과적이게 할 수 있기 때문이다. 전분-처리 효소는 특정 실시형태에서, 온-사이트 바이오리파이너리에서 생성된 다음, 바이오에탄올 정제소 또는 생화학물질/바이오물질 제조 시설로 신속하게 용이하게 통합되어, 바이오에탄올을 생성할 수 있다.

따라서, 특정 태양에서, 본 발명은 또한, 특정 바이오에탄올, 바이오연료, 생화학물질 또는 기타 바이오물질의 제조 및 판매에서 본 명세서의 효소(예를 들어, 셀룰라제, 헤미셀룰라제), 세포, 조성물 및 공정을 적용하는 특정 비즈니스 방법에 관한 것이다. 일부 실시형태에서, 본 발명은 온-사이트 바이오리파이너리 모델에서의 이러한 효소, 세포, 조성물 및 공정의 응용에 관한 것이다. 다른 실시형태에서, 본 발명은 머천트 효소 공급 모델에서의 이러한 효소, 세포, 조성물 및 공정의 응용에 관한 것이다.

관련지어 말하자면, 본 개시내용은 상업적 환경에서 본 발명의 효소 및/또는 효소 조성물의 용도를 제공한다. 예를 들어, 본 개시내용의 효소 및/또는 효소 조성물은 효소 및/또는 조성물을 사용하는 전형적이거나 바람직한 방법에 대한 설명과 함께 적절한 시장에서 판매될 수 있다. 따라서, 본 개시내용의 효소 및/또는 효소 조성물

은 머천트 효소 공급 모델 내에서 사용되거나 상품화될 수 있으며, 여기서, 본 개시내용의 효소 및/또는 효소 조성물은 연료 또는 바이오제품의 생산 업계에서 바이오에탄올의 제조처, 연료 정제소, 또는 생화학물질 또는 바이오물질 제조처에 판매된다. 일부 태양에서, 본 개시내용의 효소 및/또는 효소 조성물은 온-사이트 바이오-리파이너리 모델(on-site bio-refinery model)을 사용하여 시판되거나 상품화될 수 있으며, 여기서, 효소 및/또는 효소 조성물은 연료 정제소에 있는 또는 그 근처의 시설에서, 또는 생화학물질/바이오물질 제조처의 시설에서 생성되거나 제조되며, 본 발명의 효소 및/또는 효소 조성물은 실시간으로 연료 정제소 또는 생화학물질/바이오물질 제조처의 특정 요구에 맞춤화된다. 더욱이, 본 개시내용은 이들 제조처에 효소 및/또는 효소 조성물을 사용하기 위한 기술 지원 및/또는 설명을 제공하여, 원하는 바이오제품(예를 들어, 바이오연료, 생화학물질, 바이오물질 등)이 제조되고 시판될 수 있게 하는 것에 관한 것이다.

본 발명은 하기 실시예를 참조하여 더욱더 이해될 수 있으며, 실시예는 본 발명을 이로 한정하는 것을 의미하는 것이 아니라 예로서 주어진 것이다.

실시예

실시예 1: 검정법/방법

하기의 검정법/방법을 일반적으로 하기에 기재되는 실시예에서 사용하였다. 하기에 제공된 프로토콜로부터의 임의의 변형은 구체적인 실시예에 나타나 있다.

A. 바이오매스 기질의 전처리

옥수수 속대, 옥수수 대 및 스위치그래스를 W006110901A호에 기재된 방법 및 처리 범위에 따른(달리 기재되지 않는 한) 효소 가수분해 전에 전처리하였다. 또한, 전처리에 대한 이들 참조문헌은 US-2007-0031918-A1호, US-2007-0031919-A1호, US-2007-0031953-A1호 및/또는 US-2007-0037259-A1호의 개시내용에 포함된다.

암모니아 섬유 폭발 처리(ammonia fiber explosion treated, AFEX) 옥수수 대를 미시간 바이오테크놀로지 인스티튜트 인터내셔널(Michigan Biotechnology Institute International, MBI)로부터 취득하였다. 옥수수 대의 조성을 국립재생에너지연구소(National Renewable Energy Laboratory, NREL) 절차, NREL LAP-002를 사용하여 MBI(문헌[Teymour, F et al. Applied Biochemistry and Biotechnology, 2004, 113:951-963])로 측정하였다. NREL 절차는 http://www.nrel.gov/biomass/analytical_procedures.html에서 이용가능하다.

B. 바이오매스의 조성 분석

문헌[Determination of structural carbohydrates and lignin in the biomass (National Renewable Energy Laboratory, Golden, CO 2008 <http://www.nrel.gov/biomass/pdfs/42618.pdf>)]에 기재된 2단계 산 가수분해 방법을 사용하여, 바이오매스 기질의 조성을 측정하였다. 이러한 방법을 사용하여, 효소 가수분해 결과를 기질의 출발 셀룰로스 및 자일란 함량으로부터의 이론적 수량에 대한 전환율로 환산하여 본 명세서에 기록하였다.

C. 총 단백질 검정법

BCA 단백질 검정은 분광광도계를 사용하여 단백질 농도를 측정하는 비색 분석이다. BCA 단백질 어세이 키트(Protein Assay Kit)(피어스 케미컬(Pierce Chemical))를 제조처의 권고에 따라 사용하였다. 효소 희석액을 50 mM 아세트산나트륨 pH 5 완충제를 사용하여 테스트 튜브에서 제조하였다. 효소 희석액(각각 0.1 mL)을 1 mL의 15% 트라이클로로아세트산(TCA)을 함유하는 2 mL 에펜도르프(Eppendorf) 원심분리용 튜브에 별도로 첨가하였다. 튜브를 볼텍싱(vortexed)시키고, 빙욕에 10분간 두었다. 튜브를 14,000 rpm으로 6분 동안 원심분리시켰다. 상청액을 따라내고, 펠렛을 개별적으로 1 mL 0.1 N NaOH에 재현탁시키고, 펠렛이 용해될 때까지 튜브를 다시 볼텍싱시켰다. BSA 표준용액을 2 mg/mL의 원액으로부터 제조하였다. 0.5 mL의 시약 B와 25 mL의 시약 A를 혼합하여, BCA 작업(working) 용액을 제조하였다. 재현탁된 효소 시료를 각각 0.1 mL의 부피로 3개의 에펜도르프 원심분리용 튜브에 첨가하였다. 2 mL의 피어스 BCA 작업 용액을 각 시료 및 BSA 표준물질의 튜브에 첨가하였다. 튜브를 37°C 수조에서 30분 동안 인큐베이션시켰다. 시료를 실온으로 냉각시키고(15분), 각 시료의 562 nm에서의 흡광도를 측정하였다.

각 표준물질에 대한 단백질 흡광도 평균값을 계산하였다. 단백질 표준물질 평균을 흡광도는 x-축에 그리고 농도(mg/mL)는 y-축에 두어 플로팅(plotted)하였다. 점들을 1차 방정식에 피팅시켰다: $y=mx+b$. 효소 시료의 원래 농도는 x-값에 흡광도를 대입함으로써 계산하였다. 총 단백질 농도는 희석 계수를 곱하여 계산하였다.

정제된 시료의 총 단백질은 A280에 의해 측정하였다(문헌[Pace, CN, et al. *Protein Science*, 1995, 4:2411-2421]).

2423]).

발효 산물의 총 단백질 함량은 종종 켈달법(Kjeldahl method) (rtech laboratories)을 사용하거나 DUMAS법(TruSpec CN) (문헌[Sader, A.P.O. et al., Archives of Veterinary Science, 2004, 9(2):73-79])을 사용하여, 방출된 질소의 연소, 포획 및 측정에 의해 총 질소로서 측정하였다. 복합 시료의 경우, 예를 들어, 발효 브로쓰, 평균 16% N 함량, 및 질소의 단백질 환산 계수 6.25를 계산을 위해 사용하였다. 경우에 따라서는, 방해 비단백질 질소를 설명하기 위해, 총 침강성 단백질을 측정하였다. 그러한 경우에는, 12.5% TCA 농도를 측정을 위해 사용하고, 단백질 함유 TCA 펠릿을 0.1 M NaOH에 재현탁시켰다.

경우에 따라서는, 베타 브래드포드 에세이(Better Bradford Assay) (미국 일리노이주 록퍼드에 소재하는 써모 사이언티픽(Thermo Scientific))로도 알려진 쿠마시 플러스(Coomassie Plus)를 제조처 권고에 따라 사용하였다. 다른 경우에는, 총 단백질을 캘리브레이터(calibrator)로서 소 혈청 알부민을 사용하여 바이크셀바움(Weichselbaum) 및 고날(Gornall)에 의해 변형된 뷰렛(Biuret) 방법을 사용하여 측정하였다(문헌[Weichselbaum, T. Amer. J. Clin. Path. 1960,16:40]; 문헌[Gornall, A. et al. J. Biol. Chem. 1949, 177:752]).

D. ABTS를 이용한 글루코스 정량

글루코스 정량을 위한 ABTS(2,2'-아지노-비스(3-에틸렌티아졸린-6)-설폰산) 분석은 과산화수소(H_2O_2)의 화학양론적 양을 생성하는 동안에, O_2 의 존재 하에 글루코스 산화효소가 글루코스의 산화를 촉진시키는 원리에 기초하였다. 이러한 반응 후에 ABTS의 서양고추냉이 과산화효소(HRP) 촉매 산화가 이어지며, 이는 H_2O_2 의 농도와 선형적으로 상관관계가 있다. 산화된 ABTS의 출현은 녹색의 발생으로 나타나며, 이는 405 nm의 OD에서 정량화된다. 2.74 mg/mL ABTS 분말(Sigma), 0.1 U/mL HRP(Sigma) 및 1 U/mL 글루코스 산화효소(옥시고(OxyGO)(등록 상표) HP L5000, 제넨코(Genencor), 다니스코 유에스에이(Danisco USA))의 혼합물을 50 mM 아세트산나트륨 완충제, pH 5.0에서 준비하여, 어두운 곳에 두었다. 글루코스 표준물질(0, 2, 4, 6, 8, 10 nmol)을 50 mM 아세트산나트륨 완충제, pH 5.0에서 준비하였다. 10 μ L의 표준물질을 96-웰 플랫 보텀(flat bottom) 마이크로타이터 플레이트에 3별로 개별적으로 첨가하였다. 10 μ L의 연속적으로 희석된 시료도 플레이트에 첨가하였다. 100 μ L의 ABTS 기질 용액을 각 웰에 첨가하여, 플레이트를 분광광도 플레이트 리더 상에 두었다. ABTS의 산화를 405 nm에서 5분간 리딩하였다.

교대로, 405 nm에서의 흡광도를 15 내지 30분간의 인큐베이션 후에 측정한 다음에, 50 mM 아세트산나트륨 완충제, pH 5.0, 및 2% SDS를 함유하는 켄칭 믹스(quenching mix)를 사용하여 반응물을 켄칭시켰다.

E. HPLC에 의한 당 분석

원심분리, 0.22 μ m 나일론 스피ن-엑스(Spin-X) 원심분리관 필터(미국 뉴욕주 코닝에 소재하는 코닝(Corning))를 통한 여과, 및 증류수를 사용한 가용성 당의 원하는 농도로의 희석을 이용하여 불용성 물질을 제거하여, 옥수수 속대 당화 가수분해로부터의 시료를 준비하였다. 6 x 50 mm SH-1011P 가드 컬럼(guard column; www.shodex.net)을 갖는 쇼덱스 슈거(Shodex Sugar) SH-G SH1011(8 x 300 mm)에서 단당류를 측정하였다. 사용된 용매는 0.01 N H_2SO_4 이고, 크로마토그래피 런을 0.6 mL/min의 유속으로 행하였다. 컬럼 온도를 50°C로 유지하고, 굴절률로 검출하였다. 교대로, 워터즈(Waters) 2410 굴절률 검출기를 구비한 바이오라드 아미넥스(Biorad Aminex) HPX-87H 컬럼을 사용하여, 당의 양을 분석하였다. 분석 시간은 약 20분이고, 주입량은 20 μ L이며, 이동상은 0.01 N 황산이고, 이는 0.2 μ m 필터를 통해 여과하여 탈가스하며, 유속은 0.6 mL/min이고, 컬럼 온도는 60°C로 유지되었다. 글루코스, 자일로스, 및 아라비노스의 외부 표준물질을 각 시료 세트를 사용하여 런하였다.

크기 배제 크로마토그래피를 사용하여, 올리고머 당을 분리하여 동정하였다. 토소 바이오셉(Tosoh Biosep) G2000PW 컬럼 7.5 mm x 60 cm을 사용하였다. 증류수를 사용하여, 당을 용리하였다. 0.6 mL/min의 유속을 사용하여, 컬럼을 실온에서 런하였다. 6탄당 표준물질은 스타키오스, 라피노스, 셀로비오스 및 글루코스를 포함하고; 5탄당 표준물질은 자일로헥소스, 자일로펜토스, 자일로테트로스, 자일로트리오스, 자일로비오스 및 자일로스를 포함하였다. 자일로-올리고머 표준물질을 구입하였다(메가자임(Megazyme)). 굴절률로 검출하였다. 피크 면적 단위 또는 상대 피크 면적율을 사용하여 결과를 기록하였다.

원심분리되어 필터로 정제된 시료(상기)의 가수분해에 의해 총 가용성 당을 측정하였다. 정제된 시료를 0.8 N H_2SO_4 를 사용하여 1:1로 희석하였다. 생성된 용액을 121°C에서 1시간 동안 캡핑된 바이알에서 오토클레이빙하였

다. 가수분해 동안에 당당류의 손실에 대한 보정을 행하지 않고 결과를 기록하였다.

F. 옥수수 속대로부터의 올리고머 제제 및 효소 분석

글루칸 + 자일란 g 당 8 mg 트리코테르마 리세이 Xyn3을 250 g 건조 중량의 회석 암모니아로 전처리된 옥수수 속대와 함께 50 mM pH 5.0 아세트산나트륨 완충제 중에서 인큐베이션하여, 옥수수 속대의 트리코테르마 리세이 Xyn3 가수분해로부터의 올리고머를 제조하였다. 반응물을 180 rpm으로 회전 진탕시키면서 48℃에서 72시간 동안 처리하였다. 상청액을 9,000 x G로 원심분리한 다음에, 0.22 μ m 날젠(Nalgene) 필터를 통해 여과하여, 가용성 당을 회수하였다.

G. 바이오매스 당화 분석

특정예가 특정 변형을 나타내지 않는 한, 본 명세서의 전형적인 예에서, 옥수수 속대 당화 분석을 하기 절차에 따라 마이크로타이터 플레이트 포맷에서 행하였다. 바이오매스 기질, 예를 들어, 회석 암모니아로 전처리된 옥수수 속대를 물에 희석시키고, 황산으로 pH 조절하여, pH 5, 7% 셀룰로스 슬러리를 형성시켜, 분석 시에 추가의 처리없이 사용하였다. 셀룰로스 g 당, 자일란 g 당, 또는 배합된 셀룰로스 및 자일란 g 당(통상적인 조성 분석법을 이용하여 측정됨, 상기 참조) 총 단백질 mg을 기준으로 하여, 효소 시료를 옥수수 속대 기질에 로딩하였다. 효소를 50 mM 아세트산나트륨, pH 5.0에 희석시켜, 원하는 로딩 농도를 얻었다. 40 μ l의 효소 용액을 웰당 7% 셀룰로스(웰당 최종 4.5% 셀룰로스에 상당함)로, 회석 암모니아로 전처리된 옥수수 속대 70 mg에 첨가하였다. 그 다음에 분석 플레이트를 알루미늄 플레이트 시일러로 커버하고, 실온에서 혼합하여, 50℃, 200 rpm에서 3일간 인큐베이션하였다. 인큐베이션 기간 종료 시에, 각 웰에 100 μ l의 100 mM 글리신 완충제, pH 10.0을 첨가하여 당화 반응물을 쉐킹하고, 플레이트를 3,000 rpm으로 5분간 원심분리하였다. 10 μ l의 상청액을 96-웰 HPLC 플레이트의 밀리큐(MilliQ) 물 200 μ l에 첨가하여, 가용성 당을 HPLC로 측정하였다.

H. 마이크로타이터 플레이트 당화 분석

정제된 셀룰라제 및 전체 셀룰라제 균주 무세포 산물을 기질 중의 g 셀룰로스 당 총 단백질(mg)을 기준으로 한 양으로 당화 분석에 도입하였다. 정제된 헤미셀룰라제를 기질의 자일란 함량을 기준으로 하여 로딩하였다. 예를 들어, 회석산으로 전처리된 옥수수 대(PCS), 암모니아 섬유로 팽창된(AFEX) 옥수수 대, 회석 암모니아로 전처리된 옥수수 속대, 수산화나트륨(NaOH)으로 전처리된 옥수수 속대, 및 회석 암모니아 스위치그래스를 비롯한 바이오매스 기질을 표시된 % 고형분 수준으로 혼합하여, 혼합물의 pH를 5.0으로 조절하였다. 플레이트를 알루미늄 플레이트 시일러로 커버하여, 50℃ 인큐베이터에 두었다. 2일간 진탕하면서 인큐베이션하였다. 100 μ l 100 mM 글리신, pH 10을 각각의 웰에 첨가하여 반응을 종료시켰다. 완전히 혼합한 후에, 플레이트를 원심분리하여, 상청액을 100 μ l 10 mM 글리신 완충제, pH 10을 포함하는 HPLC 플레이트로 10배 희석시켰다. 생성된 가용성 당의 농도를 셀로비오스 가수분해 분석(이하)에 대하여 기술한 바와 같이 HPLC를 사용하여 측정하였다. 글루칸 전환율은 $[\text{mg 글루코스} + (\text{mg 셀로비오스} \times 1.056 + \text{mg 셀로트리오스} \times 1.056)] / [\text{기질 중의 mg 셀룰로스} \times 1.111]$ 로 정의되고; 자일란 전환율(%)은 $[\text{mg 자일로스} + (\text{mg 자일로비오스} \times 1.06)] / [\text{기질 중의 mg 자일란} \times 1.136]$ 로 정의된다.

I. 셀로비오스 가수분해 분석

셀로비아제 활성을 문헌[Ghose, T.K. Pure and Applied Chemistry, 1987, 59(2), 257-268]의 방법을 사용하여 측정하였다. 셀로비오스 단위(고세(Ghose)에 기술한 바와 같이 유도됨)는 분석 조건 하에 0.1 mg 글루코스를 방출하는데 필요한 효소의 양으로 나누어진 0.815로 정의된다.

J. 클로로-니트로-페닐-글루코시드(CNPG) 가수분해 분석

200 μ l의 50 mM 아세트산나트륨 완충제, pH 5를 마이크로타이터 플레이트의 각각의 웰에 첨가하였다. 플레이트를 커버하여, 에펜도르프 서모믹서(Thermomixer)에서 37℃로 15분간 평형시켰다. 50 mM 아세트산나트륨 완충제, pH 5에 희석시킨 5 μ l의 효소도 각각의 웰에 첨가하였다. 플레이트를 다시 커버하여, 37℃로 5분간 평형시켰다. 밀리포어(Millipore) 물에서 제조한 20 μ l의 2 mM 2-클로로-4-니트로페닐-베타-D-글루코피라노시드(CNPG, 미국 캘리포니아주 에드먼턴에 소재하는 로즈 사이언티픽 리미티드(Rose Scientific Ltd.))를 각각의 웰에 첨가하고, 플레이트를 신속하게 분광광도계(스펙트라맥스(SpectraMax) 250, 몰레큘러 디바이시스(Molecular Devices))에 옮겼다. OD 405 nm에서 15분간 키네틱 리딩(kinetic read)을 행하여, 데이터를 V_{\max} 로 기록하였다. CNP의 흡광 계수를 사용하여, V_{\max} 를 OD/sec의 단위로부터 μ M CNP/sec의 단위로 변환시켰다. μ M CNP/sec를 분석 시에 사용된 효소 단백질 mg으로 나누어, 비활성도(μ M CNP/sec/mg 단백질)를 측정하였다.

K. 칼코플루오르 검정법

사용된 모든 화학물질은 분석용으로 이루어졌다. 아비셀(Avicel) PH-101을 FMC 바이오폴리머(BioPolymer) (미국 펜실베이니아주 필라델피아에 소재)로부터 구입하였다. 셀로비오스 및 칼코플루오르 화이트를 시그마(Sigma(미국 미주리주 세인트 루이스에 소재))로부터 구입하였다. 인산 팽윤된 셀룰로스(PASC)를 문헌 [Walseth, TAPPI 1971, 35:228 and Wood, Biochem. J. 1971, 121:353-362]의 개조된 프로토콜을 사용하여, 아비셀 PH-101로 제조하였다. 요컨대, 아비셀을 진한 인산에 가용화시킨 다음에, 냉각 탈이온수를 사용하여 침전시켰다. 셀룰로스를 수집하고, 더 많은 물로 세정하여, pH를 중화시킨 후에, 이를 50 mM 아세트산나트륨 pH 5 중에서 1% 고형분으로 희석시켰다.

모든 효소 희석액을 50 mM 아세트산나트륨 완충제, pH 5.0으로 제조하였다. GC220 셀룰라제(다니스코 유에스 인코포레이티드, 제넨코)를 2.5, 5, 10 및 15 mg(단백질)/g(PASC)으로 희석하여, 선형 보정 곡선을 산출하였다. 테스트할 시료를 보정 곡선 범위 내에 있도록, 즉, 0.1 내지 0.4의 분획 생성물의 반응을 얻도록 희석시켰다. 150 μ l의 냉각 1% PASC를 96-웰 마이크로타이터 플레이트 내의 20 μ l의 효소 용액에 첨가하였다. 플레이트를 커버하여, 이노바(Innova) 인큐베이터/진탕기에서 50°C, 200 rpm에서 2시간 동안 인큐베이션하였다. 반응물을 100 mM 글리신, pH 10 중의 100 μ l의 50 μ g/mL 칼코플루오르로 켄칭하였다. 여기 파장 Ex = 365 nm 및 발광 파장 Em = 435 nm에서 형광 마이크로플레이트 리더(SpectraMax M5(몰레칼러 디바이시스))로 형광을 리딩하였다. 결과를 하기 방정식에 따라 분획 생성물로 나타낸다:

$$FP = 1 - (F1 \text{ 시료} - F1 \text{ 완충제 w/ 셀로비오스}) / (F1 \text{ 제로 효소} - F1 \text{ 완충제 w/ 셀로비오스}),$$

여기서, FP는 분획 생성물이고, F1 = 형광 단위이다.

실시예 2: 트리코테르마 리세이의 통합 발현 균주의 구축

5가지 유전자를 공동-발현하는 트리코테르마 리세이의 통합 발현 균주를 구축하였다: 트리코테르마 리세이 β -글루코시다제 유전자 *bg11*, 트리코테르마 리세이 엔도자일라나제 유전자 *xyn3*, 푸사리움 베르티실리오이데스 β -자일로시다제 유전자 *fv3A*, 푸사리움 베르티실리오이데스 β -자일로시다제 유전자 *fv43D*, 및 푸사리움 베르티실리오이데스 α -아라비노푸라노시다제 유전자 *fv51A*.

이들 상이한 유전자 및 트리코테르마 리세이 균주의 형질전환을 위한 발현 카세트의 구축은 후술되어 있다.

A. β -글루코시다제 발현 벡터의 구축

고유 트리코테르마 리세이 β -글루코시다제 유전자 *bg11*의 N-말단 부분을 코돈 최적화시켰다(DNA 2.0, 미국 캘리포니아주 멘로 파크 소재). 이러한 합성된 부분은 이러한 효소의 암호화 영역의 제1의 447개 염기로 이루어졌다. 그 다음에 이러한 단편을 프라이머 SK943 및 SK941(이하)을 사용하여 PCR로 증폭시켰다. 고유 *bg11* 유전자의 나머지 영역을 SK940 및 SK942(이하)을 사용하여 트리코테르마 리세이 균주 RL-P37로부터 추출된 게놈 DNA 시료로부터 PCR로 증폭시켰다(문헌[Sheir-Neiss, G *et al.* Appl. Microbiol. Biotechnol. 1984, 20:46-53]). *bg11* 유전자의 이들 2개의 PCR 단편을 프라이머 SK943 및 SK942를 사용하여 융합 PCR 반응에서 함께 융합시켰다:

정방향 프라이머 SK943: (5'-CACCATGAGATATAGAACAGCTGCCGCT-3')(서열 번호 92)

역방향 프라이머 SK941: (5'-CGACCGCCCTGCGGAGTCTTGCCCAGTGGTCCCGACAG-3')(서열 번호 93)

정방향 프라이머 (SK940): (5'-CTGTCGCGGGACCACTGGGCAAGACTCCGAGGGCGGTCG-3')(서열 번호 94)

역방향 프라이머 (SK942): (5'-CCTACGCTACCGACAGAGTG-3')(서열 번호 95)

생성된 융합 PCR 단편을 게이트웨이(Gateway)(등록 상표) 엔트리(Entry) 벡터 pENTR(상표명)/D-TOPO(등록 상표)로 클로닝하고, 에스케리키아 콜라이 원 샷(One Shot)(등록 상표) TOP10 화학적 컴피턴트(Competent) 세포(인비트로젠(Invitrogen))로 형질전환시켜, 중간 벡터, pENTR TOPO-Bg11(943/942)을 생성하였다(도 55b). 삽입된 DNA의 뉴클레오티드 서열을 결정하였다. 정확한 *bg11* 서열이 있는 pENTR-943/942 벡터를 LR 클로나제(clonase)(등록 상표) 반응(인비트로젠이 요약한 프로토콜을 참조함)을 사용하여 pTrex3g와 재조합시켰다. LR 클로나제 반응 혼합물을 에스케리키아 콜라이 원 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)로 형질전환시켜, 발현 벡터, pTrex3g 943/942를 생성하였다(맵, 도 55c 참조). 또한, 벡터에는 트리코테르마 리세이의 형질전환을 위한 선택가능한 마커로서 아세트아미다제를 암호화하는 아스페르길루스 니둘란스 *amdS* 유전자가 포함되어 있었다. 발현 카세트를 프라이머 SK745 및 SK771(이하)을 사용하여 PCR로 증폭시켜, 형질전환을

위한 산물을 생성하였다.

정방향 프라이머 SK771: (5'-GTCTAGACTGGAACGCAAC-3')(서열 번호 96)

역방향 프라이머 SK745: (5'-GAGTTGTGAAGTCGGTAATCC-3')(서열 번호 97)

1) 엔도자일라나제 발현 카세트의 구축

고유 트리코테르마 리세이 엔도자일라나제 유전자 *xyn3*를 프라이머 *xyn3F*-2 및 *xyn3R*-2를 사용하여, 트리코테르마 리세이로부터 추출된 게놈 DNA 시료로부터 PCR로 증폭시켰다.

정방향 프라이머 *xyn3F*-2: (5'-CACCATGAAAGCAAACGTCATCTTGTGCCTCCTGG-3')(서열 번호 98)

역방향 프라이머 *xyn3R*-2: (5'-CTATTGTAAGATGCCAACATGCTGTTATATGCCG GCTTGGGG-3')(서열 번호 99)

생성된 PCR 단편을 게이트웨이(등록 상표) 엔트리 벡터 pENTR(상표명)/D-TOPO(등록 상표)로 클로닝하고, 에스케리키아 콜라이 원 샷(등록 상표) TOP10 화학적 컴피턴트 세포로 형질전환시켜, 도 55d에 나타난 벡터를 생성하였다. 삽입된 DNA의 뉴클레오티드 서열을 결정하였다. 정확한 *xyn3* 서열이 있는 pENTR/Xyn3 벡터를 LR 클로나제(등록 상표) 반응 프로토콜(인비트로젠)을 사용하여 pTrex3g와 재조합시켰다. 그 다음에 LR 클로나제(등록 상표) 반응 혼합물을 에스케리키아 콜라이 원 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)로 형질전환시켜, 최종 발현 벡터, pTrex3g/Xyn3를 생성하였다(도 55e 참조). 또한, 벡터에는 트리코테르마 리세이의 형질전환을 위한 선택가능한 마커로서 아세트아미다제를 암호화하는 아스페르길루스 니들란스 *amdS* 유전자가 포함되어 있다. 발현 카세트를 프라이머 SK745 및 SK822(이하)를 사용하여 PCR로 증폭시켜, 형질전환을 위한 산물을 생성하였다.

정방향 프라이머 SK745: (5'-GAGTTGTGAAGTCGGTAATCC-3')(서열 번호 100)

역방향 프라이머 SK822: (5'-CACGAAGAGCGGCGATTC-3')(서열 번호 101)

2) β -자일로시다제 Fv3A 발현 벡터의 구축

푸사리움 베르티실리오이데스 β -자일로시다제 *fv3A* 유전자를 프라이머 MH124 및 MH125를 사용하여 푸사리움 베르티실리오이데스 게놈 DNA 시료로부터 증폭시켰다.

정방향 프라이머 MH124: (5'-CACCCATGCTGCTCAATCTTCAG-3')(서열 번호 102)

역방향 프라이머 MH125: (5'-TTACGCAGACTTGGGTCTTGAG-3')(서열 번호 103)

PCR 단편을 게이트웨이(등록 상표) 엔트리 벡터 pENTR(상표명)/D-TOPO(등록 상표)로 클로닝하고, 에스케리키아 콜라이 원 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)로 형질전환시켜, 중간 벡터, pENTR-Fv3A를 생성하였다(도 55f 참조). 삽입된 DNA의 뉴클레오티드 서열을 결정하였다. 정확한 *fv3A* 서열이 있는 pENTR-Fv3A 벡터를 LR 클로나제(등록 상표) 반응 프로토콜(인비트로젠)을 사용하여 pTrex6g와 재조합시켰다. LR 클로나제(등록 상표) 반응 혼합물을 에스케리키아 콜라이 원 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)로 형질전환시켜, 최종 발현 벡터, pTrex6g/Fv3A를 생성하였다(도 55g 참조). 벡터에는 또한 고유 트리코테르마 리세이 아세토락테이트 신타제(*als*) 유전자, *alsR*의 클로리무론 에틸 내성 돌연변이체가 포함되어 있으며, 이는 국제 특허 공개 제W02008/039370 A1호에 기재된 방법에 따라, 트리코테르마 리세이의 형질전환을 위한 선택가능한 마커로서 그의 고유 프로모터 및 터미네이터와 함께 사용되었다. 발현 카세트를 프라이머 SK1334, SK1335 및 SK1299(이하)를 사용하여 PCR로 증폭시켜, 형질전환을 위한 산물을 생성하였다.

정방향 프라이머 SK1334: (5'-GCTTGAGTGATCGTGAAG-3')(서열 번호 104)

정방향 프라이머 SK1335: (5'-GCAACGGCAAAGCCCACTTC-3')(서열 번호 105)

역방향 프라이머 SK1299: (5'-GTAGCGGCCGCTCATCTCATCTCATCCATCC-3')(서열 번호 106)

3) β -자일로시다제 Fv43D 발현 카세트의 구축

푸사리움 베르티실리오이데스 β -자일로시다제 Fv43D 발현 카세트의 구축을 위하여, *fv43D* 유전자 산물을 프라이머 SK1322 및 SK1297(이하)를 사용하여 푸사리움 베르티실리오이데스 게놈 DNA 시료로부터 증폭시켰다. 엔도글루카나제 유전자 *eg11*의 프로모터의 영역을 프라이머 SK1236 및 SK1321(이하)를 사용하여 균주 RL-P37로부터 추출된 트리코테르마 리세이 게놈 DNA 시료로부터 PCR에 의해 증폭시켰다. 이들 PCR 증폭된 DNA 단편을 이후에 프라이머 SK1236 및 SK1297(이하)를 사용하여 융합 PCR 반응에서 융합시켰다. 생성된 융합 PCR 단편을 pCR-

Blunt II-TOPO 벡터(인비트로젠)로 클로닝시켜, 플라스미드 TOPO Blunt/Peg11-Fv43D를 생성하였다(도 55h 참조). 그 다음에 이러한 플라스미드를 사용하여, 에스케리키아 콜라이 윈 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)를 형질전환시켰다. 플라스미드 DNA를 몇몇 에스케리키아 콜라이 클론으로부터 추출하고, 이들의 서열을 제한효소 분해에 의해 확인하였다.

정방향 프라이머 SK1322: (5'-CACCATGCAGCTCAAGTTTCTGTC-3')(서열 번호 107)

역방향 프라이머 SK1297: (5'-GGTTACTAGTCAACTGCCCGTTCTGTAGCGAG-3')(서열 번호 108)

정방향 프라이머 SK1236: (5'-CATGCGATCGCGACGTTTTGGTCAGGTCG-3')(서열 번호 109)

역방향 프라이머 SK1321: (5'-GACAGAACTTGAGCTGCATGGTGTGGGACAACAAGAAGG-3')(서열 번호 110)

발현 카세트를 프라이머 SK1236 및 SK1297(상기)을 사용하여 TOPO Blunt/Peg11-Fv43D로부터 PCR로 증폭시켜, 형질전환을 위한 산물을 생성하였다.

4) α -아라비노푸라노시다제 발현 카세트의 구축

푸사리움 베르티실리오이데스 α -아라비노푸라노시다제 유전자 fv51A 발현 카세트의 구축을 위하여, fv51A 유전자 산물을 프라이머 SK1159 및 SK1289(이하)를 사용하여 푸사리움 베르티실리오이데스 게놈 DNA 시료로부터 증폭시켰다. 엔도글루카나제 유전자 eg11의 프로모터의 영역을 프라이머 SK1236 및 SK1262(이하)를 사용하여 균주 RL-P37(상기 참조)로부터 추출된 트리코데르마 리세이 게놈 DNA 시료로부터 PCR에 의해 증폭시켰다. 그 다음에 PCR 증폭된 DNA 단편을 프라이머 SK1236 및 SK1289(이하)를 사용하여 융합 PCR 반응에서 융합시켰다. 생성된 융합 PCR 단편을 pCR-Blunt II-TOPO 벡터(인비트로젠)로 클로닝시켜, 플라스미드 TOPO Blunt/Peg11-Fv51A를 생성하고(도 55i 참조), 에스케리키아 콜라이 윈 샷(등록 상표) TOP10 화학적 컴피턴트 세포(인비트로젠)를 이러한 플라스미드를 사용하여 형질전환시켰다.

정방향 프라이머 SK1159: (5'-CACCATGGTTCGCTTCAGTTCAATCCTAG-3')(서열 번호 111)

역방향 프라이머 SK1289: (5'-GTGGCTAGAAGATATCCAACAC-3')(서열 번호 112)

정방향 프라이머 SK1236: (5'-CATGCGATCGCGACGTTTTGGTCAGGTCG-3')(서열 번호 113)

역방향 프라이머 SK1262: (5'-GAACTGAAGCGAACCATGGTGTGGGACAACAAGAAGGAC-3')(서열 번호 114)

발현 카세트를 프라이머 SK1298 및 SK1289(상기)를 사용하여 PCR로 증폭시켜, 형질전환을 위한 산물을 생성하였다.

정방향 프라이머 SK1298: (5'-GTAGTTATGCGCATGCTAGAC-3')(서열 번호 115)

역방향 프라이머 SK1289: (5'-GTGGCTAGAAGATATCCAACAC-3')(서열 번호 112)

5) β -글루코시다제 및 엔도자일라나제 발현 카세트를 이용한 트리코데르마 리세이의 공동-형질전환

RL-P37로부터 유래되고(문헌[Sheir-Neiss, G et al. Appl. Microbiol. Biotechnol. 1984, 20:46-53]) 높은 셀룰라제 생성을 위해 선택된 트리코데르마 리세이 돌연변이체 균주를 PEG-매개 형질전환법(문헌[Penttila, M et al. Gene 1987, 61(2):155-64] 참조)을 사용하여, β -글루코시다제 발현 카세트(*cbh1* 프로모터, 트리코데르마 리세이 베타-글루코시다제 1 유전자, *cbh1* 터미네이터, 및 *amdS* 마커), 및 엔도자일라나제 발현 카세트(*cbh1* 프로모터, 트리코데르마 리세이 *xyn3*, 및 *cbh1* 터미네이터)로 공동-형질전환시켰다. 많은 형질전환체를 단리하고, β -글루코시다제 및 엔도자일라나제 생성에 대해 시험하였다. 트리코데르마 리세이 균주 #229로 지칭되는 하나의 형질전환체를 다른 발현 카세트를 사용한 형질전환을 위해 선택하였다.

6) 2개의 β -자일로시다제 및 α -아라비노푸라노시다제 발현 카세트를 이용한 트리코데르마 리세이 균주 #229의 공동-형질전환

트리코데르마 리세이 균주 #229를 예를 들어, 국제 특허 공개 제W02008153712A2호에 따라 전기천공법(electroporation)을 사용하여 β -자일로시다제 *fv3A* 발현 카세트(*cbh1* 프로모터, *fv3A* 유전자, *cbh1* 터미네이터, 및 *alsR* 마커), β -자일로시다제 *fv43D* 발현 카세트(*eg11* 프로모터, *fv43D* 유전자, 고유 *fv43D* 터미네이터), 및 *fv51A* α -아라비노푸라노시다제 발현 카세트(*eg11* 프로모터, *fv51A* 유전자, *fv51A* 고유 터미네이터)로 공동-형질전환시켰다. 형질전환체를 클로리무론 에틸(80 ppm)을 함유하는 보겔스(Vogels) 아가 플레이트 상에서 선택하였다.

50x 보겔스 스톡 용액(레시피)	20 mL
BBL 아가	20 g
탈이온 H ₂ O	980 mL까지 첨가

멸균-후 첨가:50% 글루코스 20 mL

50x 보겔스 스톡 용액, 리터당:

750 mL 탈이온 H₂O 중에, 연속하여 다음을 용해:

Na ₃ 시트레이트*2H ₂ O	125 g
KH ₂ PO ₄ (무수)	250 g
NH ₄ NO ₃ (무수)	100 g
MgSO ₄ *7H ₂ O	10 g
CaCl ₂ *2H ₂ O	5 g

보겔스 미량 원소 용액(하기의 레시피) 5 mL

d-비오틴	0.1 g
탈이온 H ₂ O,	1 L까지 첨가

보겔스 미량 원소 용액:

시트르산	50 g
ZnSO ₄ . *7H ₂ O	50 g
Fe(NH ₄) ₂ SO ₄ . *6H ₂ O	10 g
CuSO ₄ . 5H ₂ O	2.5 g
MnSO ₄ . 4H ₂ O	0.5 g
H ₃ BO ₃	0.5 g
Na ₂ MoO ₄ . 2H ₂ O	0.5 g

많은 형질전환체를 단리하고, β-자일로시다제 및 L-α-아라비노푸라노시다제 생성에 대해 시험하였다. 또한, 형질전환체를 실시예 1에 기재된 옥수수 속대 당화 검정법에 따라 바이오매스 전환 성능에 대해 스크리닝하였다. 본 명세서에 기재된 트리코테르마 리세이 통합 발현 균주의 예는 H3A, 39A, A10A, 11A, 및 G9A로부터 선택되고, 이는 베타-글루코시다제 1, Xyn3을 암호화하는 트리코테르마 리세이 유전자, 및 Fv3A, Fv51A, 및 Fv43D를 암호화하는 푸사리움 유전자를 상이한 비율로 발현하였다. 다른 H3A 균주에 비하여, 낮은 수준의 트리코테르마 리세이 Bgl1을 발현시킨 특정 H3A 균주, #5 ("H3A-5")를 이하 본 명세서에 기재된 실험에 사용하였다. 감소된 수준의 트리코테르마 리세이 Bgl1을 발현하는 또 하나의 H3A 균주를 실시예 5에 기재된 실험에 사용하였다. 그 중에서도, 웨스턴 블롯(Western Blot)으로 결정시, 트리코테르마 리세이 균주에는 과발현된 트리코테르마 리세이 Xyn3가 결여되어 있고; 다른 균주에는 Fv51A가 결여되어 있으며, 2개의 균주에는 Fv3A가 결여되어 있었다.

7) 트리코테르마 리세이 통합 균주 H3A의 조성

트리코테르마 리세이 통합 균주 H3A의 발효 및 조성 결정에 의해, 본 명세서의 도 3에 나타난 비로, 하기 유전자 산물의 존재를 동정하였다: 트리코테르마 리세이 Xyn3, 트리코테르마 리세이 Bgl1, Fv3A, Fv51A, 및 Fv43D.

8) HPLC에 의한 단백질 분석

액체 크로마토그래피(LC) 및 질량분석(MS)을 수행하여, 발효 브로쓰에 함유되어 있는 효소를 분리하고, 정량화하였다. 효소 시료를 먼저 스트렙토마이세스 플리카투스(*S. pllicatus*)로부터 재조합에 의해 발현되는 endoH 글리코시다제(예를 들어, NEB P0702L)로 처리하였다. EndoH를 시료 중의 총 단백질 μg 당 0.01 내지 0.03 μg 의 endoH의 양으로 사용하였다. 혼합물을 37°C, pH 4.5 내지 6.0에서 3시간 동안 인큐베이션시켜, HPLC 분석 전에 N-결합 글리코실화를 효소에 의해 제거하였다. 이어서, 약 50 μg 의 단백질에 대하여 35분간에 걸쳐 하이-투-로우(high-to-low) 염 기울기 및 HIC-페닐 컬럼을 사용하여 소수성 상호작용 크로마토그래피(아질런트(Agilent) 1100 HPLC)를 행하였다. 기울기는 고농도 염 완충제 A: 20 mM 인산칼륨을 함유하는 4 M 황산암모늄, pH 6.75; 및 저농도 염 완충제 B: 20 mM 인산칼륨, pH 6.75를 사용하여 달성하였다. 피크를 UV 222 nm에서 검출하였다. 분획을 수집하고, 질량 분석을 이용하여 분석하였다. 단백질 비를 시료의 통합된 총 면적에 대한 각 피크 면적의 백분율로 기록하였다.

9) 희석 암모니아로 전처리된 옥수수 속대의 당화에 대한, 트리코테르마 리세이 통합 균주 H3A의 발효 브로쓰로 정제된 단백질의 첨가의 효과

본 실험은 전처리된 바이오매스의 당화에 대한 다양한 효소(대부분 정제되었지만 정제되지 않은 효소도 포함)에 의해 부여된 이점을 평가하였다. 정제된 단백질 및 하나의 정제되지 않은 단백질을 스톡 용액으로부터 연속적으로 희석하여, 트리코테르마 리세이 통합 균주 H3A의 발효 브로쓰에 첨가하였다. 희석 암모니아로 전처리된 옥수수 속대를 20% 고형분(w/w) (웰당 약 5 mg의 셀룰로스), pH 5로 96-웰 마이크로타이터 플레이트 웰로 로딩하였다. H3A 발효 브로쓰를 20 mg(단백질)/g(셀룰로스)으로 각 웰에 첨가하였다. 각 희석 단백질의 10, 5, 2, 및 1 μl 의 부피(도 4a)를 각 웰에 첨가하고, 또한 각 웰에 첨가되는 액체가 총 10 μl 가 되도록 물을 첨가하였다. 기준 웰에는 10 μl 의 물 또는 추가의 H3A μl 의 희석액 중 어느 하나의 첨가를 포함시켰다. 마이크로타이터 플레이트를 포일(foil)로 밀봉하고, 3일 동안 이노바 인큐베이터 진탕기에서 200 rpm의 속도로 진탕하면서, 50°C에서 인큐베이션시켰다. 시료를 100 μl 의 100 mM 글리신 pH 10으로 켄칭하였다. 그 다음에 플레이트를 플라스틱 시일로 커버하고, 4°C에서 5분 동안 3,000 rpm으로 원심분리하였다. 5 μl 분취량의 켄칭된 반응 혼합물을 100 μl 의 물을 사용하여 희석하였다. 반응에서 생성되는 글루코스의 농도를 HPLC를 사용하여 측정하였다. 글루코스 수율을 20 mg/g의 H3A에 첨가되는 단백질 농도의 함수로서 측정하였다. 결과는 도 4b 내지 4e에 나타나 있다.

실시예 3: Fv3C의 클로닝, 발현 및 정제

A. Fv3C의 클로닝 및 발현

Fv3C 서열(서열 번호 60)을 브로드 인스티튜트(Broad Institute) 데이터베이스(<http://www.broadinstitute.org/>) 내의 푸사리움 베르티실리오이데스 게놈에서 GH3 β -글루코시다제 상동체를 검색함으로써 획득하였다. Fv3C 오픈 리딩 프레임을 주형으로서 푸사리움 베르티실리오이데스로부터의 정제된 게놈 DNA를 사용하여 PCR에 의해 증폭시켰다. 사용된 PCR 써모사이클러(thermocycler)는 DNA 엔진 테트라드(Engine Tetrad) 2 펠티에르 써멀 사이클러(Peltier Thermal Cycler)(바이오-래드 래보러터리즈(Bio-Rad Laboratories))였다. 사용된 DNA 중합효소는 PfuUltra II 퓨전(Fusion) HS DNA 중합효소(스트라타진(Stratagene))였다. 오픈 리딩 프레임을 증폭시키기 위해 사용된 프라이머는 하기와 같았다:

정방향 프라이머 MH234 (5'-CACCATGAAGCTGAATTGGGTCGC-3')(서열 번호 116)

역방향 프라이머 MH235 (5'-TTACTCCAACCTGGCGCTG-3')(서열 번호 117)

정방향 프라이머는 5'-말단에 4개의 추가의 뉴클레오티드(서열 - CACC)를 포함시켜, pENTR/D-TOPO(미국 캘리포니아주 칼스배드에 소재하는 인비트로젠)로의 방향성 클로닝을 용이하게 하였다. 오픈 리딩 프레임을 증폭시키기 위한 PCR 조건은 하기와 같았다: 단계 1: 94°C에서 2분. 단계 2: 94°C에서 30초. 단계 3: 57°C에서 30초. 단계 4: 72°C에서 60초. 단계 2, 3 및 4를 추가 29 사이클 동안 반복하였다. 단계 5: 72°C에서 2분. Fv3C 오픈 리딩 프레임의 PCR 산물을 퀴아퀵(Qiaquick) PCR 정제 키트(퀴아젠(Qiagen))를 사용하여 정제하였다. 정제된 PCR 산물을 초기에 pENTR/D-TOPO 벡터로 클로닝하고, TOP10 화학적 컴피턴트 에스케리키아 콜라이 세포(인비트로젠)로 형질전환시키고, 50 ppm 카나마이신을 함유하는 LA 플레이트 상에 플레이트링하였다. 플라스미드 DNA를 퀴아스핀(QIAspin) 플라스미드 제조 키트(퀴아젠)를 사용하여 에스케리키아 콜라이 형질전환체로부터 획득하였다. pENTR/D-TOPO 벡터 내에 삽입된 DNA에 대한 서열 확인을 M13 정방향 및 역방향 프라이머 및 하기의 추가의 서열 프라이머를 사용하여 획득하였다:

MH255 (5'-AAGCCAAGAGCTTGTGTCC-3')(서열 번호 118)

MH256 (5'-TATGCACGAGCTCTACGCCT-3')(서열 번호 119)

MH257 (5'-ATGGTACCCTGGCTATGGCT-3')(서열 번호 120)

MH258 (5'-CGGTCACGGTCTATCTTGGT-3')(서열 번호 121)

Fv3C 오픈 리딩 프레임의 정확한 DNA 서열이 있는 pENTR/D-TOPO 벡터(도 44)를 LR 클로나제(등록 상표) 반응 혼합물(인비트로젠)을 사용하여 pTrex6g(도 45a) 데스티네이션(destination) 벡터와 재조합하였다.

LR 클로나제(등록 상표) 반응의 산물을 이후에 TOP10 화학적 컴피턴트 에스케리키아 콜라이 세포(인비트로젠)로 형질전환시킨 다음에, 이를 50 ppm 카르베니실린을 함유하는 LA 플레이트 상으로 플레이팅하였다. 생성된 pExpression 구축물은 Fv3C 오픈 리딩 프레임 및 트리코테르마 리세이 돌연변이된 아세트락테이트 신타제 선택 마커(als)를 포함하는 pTrex6g/Fv3C(도 45b)이었다. Fv3C 오픈 리딩 프레임을 함유하는 pExpression 구축물의 DNA를 퀴아젠 미니프랩(miniprep) 키트를 사용하여 단리하고, 트리코테르마 리세이 포자의 바이올리스틱(biolytic) 형질전환을 위해 사용하였다.

적절한 Fv3C 오픈 리딩 프레임을 포함하는 pTrex6g 발현 벡터를 이용한 트리코테르마 리세이의 바이올리스틱 형질전환을 수행하였다. 구체적으로, *cbh1*, *cbh2*, *egl1*, *egl2*, *egl3* 및 *bg11*이 결실된 트리코테르마 리세이 균주(즉, 헥사-결실 균주, 국제 특허 공개 제WO 05/001036호 참조)를 제조처의 설명서에 따라 바이올리스틱(등록 상표) PDS-1000/he 입자 전달 시스템(바이오-래드)을 사용하여 헬륨-충격(helium-bombardment)에 의해 형질전환시켰다(제US 2006/0003408호 참조). 형질전환체를 새로운 클로리무론 에틸 선택 플레이트로 옮겼다. 안정한 형질전환체를 탄소원으로서 약 2% 글루코스/소포로스 혼합물; 100 g/L CaCl₂ 10 mL/L; 175 g/L 무수 시트르산, 200 g/L FeSO₄ · 7H₂O, 16 g/L ZnSO₄ · 7H₂O, 3.2 g/L CuSO₄ · 5H₂O, 1.4 g/L MnSO₄ · H₂O, 0.8 g/L H₃BO₃를 함유하는 400X 트리코테르마 리세이 미량 원소 용액 2.5 mL/L를 멸균 후 첨가하면서, 200 μ l/웰의 글리신 최소 배지(6.0 g/L 글리신; 4.7 g/L (NH₄)₂SO₄; 5.0 g/L KH₂PO₄; 1.0 g/L MgSO₄ · 7H₂O; 33.0 g/L PIPPS, pH 5.5 함유)를 함유하는 필터 마이크로타이터 플레이트(코닝)로 접종하였다. 형질전환체를 28°C 인큐베이터에 수용된 O₂가 풍부한 챔버에서 5일 동안 액체 배양액에서 성장시켰다. 필터 마이크로타이터 플레이트로부터의 상청액 시료를 진공 매니폴드(vacuum manifold) 상에서 수집하였다. 상청액 시료를 4 내지 12% 뉴페이지(NuPAGE) 겔에서 런하여, 심플리 블루 염색제(Simply Blue stain)(인비트로젠)를 사용하여 염색시켰다.

B. Fv3C의 정제

진탕 플라스크 농축물로부터의 Fv3C를 25 mM TES 완충제, pH 6.8에 대하여 하룻밤 투석하였다. 투석된 효소 용액을 pH 6.8에서 25 mM TES, 0.1 M 염화나트륨으로 사전-평형화된 SEC HiLoad 슈퍼덱스(Superdex) 200 프랩 그레이트(Prep Grade) 가교결합된 아가로스 및 텍스트란 컬럼(지이 헬스케어(GE Healthcare)) 상에 1 mL/분의 유속으로 로딩하였다. SDS-PAGE를 사용하여, SEC 분리로부터의 분획 중의 Fv3C의 존재를 확인하고 알아냈다. Fv3C를 함유하는 분획을 풀링(pooled)하고, 농축시켰다. 또한, SEC 정제를 사용하여 저 분자량 및 고 분자량 오염물질로부터 Fv3C를 분리하였다. 효소 제제의 순도를 쿠마지 블루로 염색된 SDS/PAGE를 사용하여 측정하였다. SDS/PAGE에 의해, 97 kDa에서 단일의 주요 밴드가 나타났다.

C. Fv3C의 선택적 번역

Fv3C 유전자의 발현을 위하여, 푸사리움 데이터베이스에 주석이 달린 ORF 함유 게놈 서열을 사용하였다. http://www.broadinstitute.org/annotation/genome/fusarium_group/MultiHome.html. 예측된 암호화 영역은 3개의 인트론을 함유하며, 제1 인트론은 신호 펩티드 서열에 개재된다(도 46a).

그러나, 제1 인트론은 그의 3' 부분에서, 또한 신호 펩티드를 암호화하는 것으로 예측되는 성숙 서열과 프레임 내에 있는 선택적 ORF를 포함한다(도 46b). 둘 모두의 번역에서, N-말단 서열 분석에 의해 결정시, 성숙 단백질에 대한 시작 부위(도 46b에 밑줄 그어짐)는 둘 모두의 추정된 신호 펩티드 절단 부위(화살표로 표시)로부터 다운스트림에서 시작하였다. Fv3C가 추정의 번역 시작으로서 ATG 중 어느 하나를 사용함으로써 효과적으로 발현될 수 있는 것으로 나타났다(도 46c).

실시예 4: 셀로비오스 및 CNPG에 대한 β -글루코시다제 활성

본 실험에서, 셀로비오스 및 CNPG에 대한 트리코테르마 리세이 Bgl1, 아스페르길루스 니게르 Bglu(An3A)(아일랜드 위클로우에 소재하는 메가자임 인터내셔널 아일랜드 리미티드(Megazyme International Ireland Ltd.)), Fv3C(서열 번호 60), Fv3D(서열 번호 58) 및 Pa3C(서열 번호 80)의 β -글루코시다제 활성을 시험하였다. 트리

코데르마 리세이 Bgl1, 아스페르길루스 니게르 Bglu("An3A"), Fv3C, Fv3C/Te3A/Bgl3(FAB) 키메라, Fv3C/Bgl3(FB) 키메라, 트리코데르마 리세이 Bgl3, 및 Te3A는 정제된 단백질이었다. Fv3D 및 Pa3C는 정제되지 않은 단백질이었다. 이들은 트리코데르마 리세이 핵사-결실 균주(상술함)에서 발견되었지만, 일부 백그라운드 단백질 활성이 여전히 존재하였다. 도 5a에 나타난 바와 같이, Fv3C는 셀로비오스에 대하여 트리코데르마 리세이 Bgl1의 활성의 대략 2배를 갖는 것으로 관찰된 한편, 아스페르길루스 니게르 Bglu는 트리코데르마 리세이 Bgl1보다 약 12배 더 활성이 있는 것으로 관찰되었다.

CNPG 기질에 대한 Fv3C의 활성은 트리코데르마 리세이 Bgl1의 활성과 대략 동일하였으나, 아스페르길루스 니게르 Bglu의 활성은 트리코데르마 리세이 Bglu1의 활성의 약 14%였다(도 5a). Fv3C와 유사하게 발견되는 다른 푸사리움 베타-글루코시다제인 Fv3D는 측정가능한 셀로비오스 활성을 갖지 않았으나, CNPG에 대한 그의 활성은 트리코데르마 리세이 Bgl1의 활성의 약 5배였다. 또한, 유사하게 생성되는 포도스포라 안세리나 베타-글루코시다제 상동체 Pa3C는 셀로비오스 또는 CNPG 기질에 대하여 측정가능한 활성을 갖지 않았다. 이들 연구에 의해, 셀로비오스 및 CNPG에 대한 Fv3C의 활성이 분자 그 자체에 기인하며, 백그라운드 단백질 활성에 기인하지 않는다는 것이 입증된다.

실시예 5: 다양한 바이오매스 기질에 대한 Fv3C 당화

A. PASC에 대한 Fv3C 당화 성능

본 실험에서, PASC 당화를 증진시키기 위한 트리코데르마 리세이 Bgl1, Fv3C, 및 몇몇 Fv3C 상동체의 능력을 시험하였다. 20 μ l의 각 베타-글루코시다제를 96-웰 HPLC 플레이트에서, 5 mg(단백질)/g(셀룰로스)의 양으로, 10 mg(단백질)/g(셀룰로스) 로딩의 트리코데르마 리세이 Bgl1-감소된 균주로부터의 전체 셀룰라제에 첨가하였다. 150 μ l의 PASC의 0.7% 고형분 슬러리를 각 웰에 첨가하고, 플레이트를 알루미늄 플레이트 시일리로 커버하고, 50°C로 설정된 인큐베이터에 진탕하면서 2시간 동안 두었다. 100 μ l의 100 mM 글리신 완충제, pH 10을 개별 웰에 첨가함으로써 반응을 종결시켰다. 완전한 혼합 후에, 플레이트를 원심분리하여, 상청액을 개별 웰에 100 μ l의 10 mM 글리신, pH 10을 함유하는 다른 HPLC 플레이트로 10배 희석하였다. 생성된 가용성 당의 농도는 HPLC를 사용하여 측정하였다(도 47).

Fv3C-함유 혼합물이 동일한 조건 하에서 트리코데르마 리세이 Bgl1-함유 혼합물보다 더 높은 비율의 글루코스를 생성하는 것이 관찰되었다. 이에 의해, Fv3C가 트리코데르마 리세이 Bgl1보다 더 높은 셀로비오스 활성을 갖는 것이 나타났다(또한, 도 5b 참조). Fv3G, Pa3D 및 Pa3G는 PASC 가수분해에 대하여 관찰가능한 영향을 나타내지 않았으며, 이는 PASC 가수분해에 대하여 핵사-결실 백그라운드(다양한 Fv3C 상동체가 클로닝되고 발견되는)의 원인 제공이 없음이 나타났다.

B. 묶은 산으로 전처리된 옥수수 대(PCS)에 대한 Fv3C 당화 성능

본 실험에서, 13% 고형분으로 PCS 당화를 증진시키는 트리코데르마 리세이 Bgl1, Fv3C 및 몇몇 Fv3C 상동체의 능력을 마이크로타이터 플레이트 당화 검정법(상기 참조)에 기재된 방법을 사용하여 시험하였다. 시험한 각 효소에 대하여, 5 mg(단백질)/g(셀룰로스)의 베타-글루코시다제를 10 mg(단백질)/g(셀룰로스)의 트리코데르마 리세이-Bgl1 감소된 균주 유래의 전체 셀룰라제에 첨가하였다.

구체적으로, 5 mg(단백질)/g(셀룰로스)의 각 베타-글루코시다제(Bgl1, Fv3C 및 상동체)를 10 mg(단백질)/g(셀룰로스)의 트리코데르마 리세이 Bgl1 감소된 균주 유래의 전체 셀룰라제에, 또는 8 mg(단백질)/g(셀룰로스)의 정제된 헤미셀룰라제 혼합물(도 6에 나타난 것의 성분)에 첨가하였다. 효소 혼합물을 50°C에서 2일 동안 기질과 함께 인큐베이션시킨 후에 글루칸 전환%를 측정하였다.

결과를 도 48b에 나타내었다. 또한 Fv3C는 글루칸 전환%에 관해, 트리코데르마 리세이 Bgl1에 비하여 명백한 이점을 주는 것으로 관찰되었다. 또한, Fv3C는 트리코데르마 리세이 Bgl1보다 더 높은 글루코스 및 당의 총 수율을 조장하였다.

결과에 의해, 있다면 숙주 세포 백그라운드 단백질로부터의 원인 제공이 제한적임이 나타났다.

C. 희석 암모니아로 전처리된 옥수수 속대에 대한 Fv3C 당화 성능

본 실험에서, 20% 고형분으로 암모니아로 전처리된 옥수수 속대의 당화를 증진시키는 트리코데르마 리세이 Bgl1, Fv3C, 및 아스페르길루스 니게르 Bglu(An3A)의 능력을 마이크로타이터 플레이트 당화 검정법(상기 참조)에 기재된 방법에 따라 시험하였다. 구체적으로, 5 mg(단백질)/g(셀룰로스)의 베타-글루코시다제(예를 들어, 트리코데르마 리세이 Bgl1, Fv3C, 및 상동체)를 희석 암모니아로 전처리된 옥수수 속대 기질에 첨가하고, 10

mg(단백질)/g(셀룰로스)의 트리코테르마 리세이 Bgl1-감소된 균주 유래의 전체 셀룰라제도 첨가하였다. 또한, Xyn3, Fv3A, Fv43D 및 Fv51A를 함유하는, 8 mg(단백질)/g(셀룰로스)의 정제된 헤미셀룰라제 믹스(도 6)도 혼합물에 첨가하였다. 효소 혼합물을 50℃에서 2일 동안 기질과 함께 인큐베이션시킨 후에 글루칸 전환%를 측정하였다.

결과를 도 49에 나타내었다. 또한 Fv3C가 트리코테르마 리세이 Bgl1(Tr3A)을 포함하는 다른 베타-글루코시다제보다 더 잘 수행하는 것을 보인 것으로 관찰되었다. 또한, 아스페르길루스 니게르 Bglu(An3A)를 2.5 mg/g(셀룰로스) 초과로 수준으로 효소 혼합물에 첨가하면, 당화가 지연되는 것이 관찰되었다.

D. 수산화나트륨(NaOH)으로 전처리된 옥수수 속대에 대한 Fv3C 당화 성능

Fv3C 성능에 대한 다양한 기질 전처리 방법의 영향을 시험하기 위하여, 트리코테르마 리세이 Bgl1(Tr3A로도 지칭됨), Fv3C, 및 아스페르길루스 니게르 Bglu(An3A)가 12% 고형분으로, NaOH로 전처리된 옥수수 속대의 당화를 증가시키는 능력을 마이크로타이터 플레이트 당화 검정법(상기 참조)에 기재된 방법에 따라 측정하였다. 옥수수 속대의 수산화나트륨 전처리를 하기와 같이 수행하였다: 1,000 g의 옥수수 속대를 약 2 mm 크기로 분쇄한 다음, 5% 수산화나트륨 수용액 4 L에 현탁시키고, 16시간 동안 110℃로 가열하였다. 암갈색 액체를 실험실 진공 하에서 고온 여과하였다. 더 이상 색상이 용출되지 않을 때까지 필터 상의 고체 잔류물을 물로 세척하였다. 고체를 실험실 진공 하에서 24시간 동안 건조시켰다. 100 g의 시료를 700 mL의 물에 현탁시키고, 교반하였다. 용액의 pH가 11.2인 것으로 측정되었다. 시트르산 수용액(10%)을 첨가하여, pH를 5.0으로 낮추고, 현탁액을 30 분 동안 교반하였다. 이어서, 고체를 여과하고, 물로 세척하고, 실온에서 진공 하에 24시간 동안 건조시켰다. 건조 후에, 86.2 g의 다당류가 농축된 바이오매스를 수득하였다. 이러한 물질의 함수율은 약 7.3 wt%였다. 수산화나트륨 처리 전후에, 탄수화물 분석을 위한 NREL 방법에 의해 측정되는 바와 같이, 글루칸, 자일란, 리그닌 및 총 탄수화물 함량을 측정하였다. 전처리에 의해, 바이오매스의 탈리그닌화가 야기되는 한편, 글루칸/자일란 중량비가 미처리 바이오매스에 대한 중량비의 15% 내로 유지되었다.

낮은 수준의 Bgl1 발현("H3A-5 균주")에 대해 특이적으로 선택된 통합된 트리코테르마 리세이 균주 H3A 유래의 전체 셀룰라제 8.7 mg(단백질)/g(셀룰로스)을 포함시키는 것 이외에도, 약 5 mg(단백질)/g(셀룰로스)의 베타-글루코시다제(Fv3C 및 상동체)를 NaOH로 전처리된 기질에 첨가하였다. 추가의 정제된 헤미셀룰라제(예를 들어, 도 6의 혼합물)를 본 실험에서 전체 셀룰라제 백그라운드에 첨가하지 않았다. 효소 혼합물을 50℃에서 2일 동안 기질과 함께 인큐베이션시킨 후에 글루칸 전환%를 측정하였다.

결과를 도 50에 나타내었다. Fv3C가 트리코테르마 리세이 Bgl1(Tr3A), An3A, 및 Te3A를 포함하는 다른 베타-글루코시다제보다 약간 더 잘 수행하는 것을 보인 것으로 관찰되었다. 또한, 아스페르길루스 니게르 Bglu(An3A)를 4 mg/g(셀룰로스) 초과로 수준으로 첨가하면, 보다 낮은 전환이 야기되는 것이 관찰되었다.

E. 회석 암모니아로 전처리된 스위치그래스에 대한 Fv3C 당화 성능

본 실험에서, 17% 고형분으로 회석 암모니아로 전처리된 스위치그래스의 당화를 증가시키는 트리코테르마 리세이 Bgl1, Fv3C 및 아스페르길루스 니게르 Bglu(An3A)의 능력을 마이크로타이터 플레이트 당화 검정법(상기 참조)에 기재된 방법에 따라 시험하였다. 회석 암모니아로 전처리된 스위치그래스를 듀폰(DuPont)으로부터 수득하였다. www.nrel.gov/biomass/analytical_procedures.html에서 이용가능한 국립재생에너지연구소(National Renewable Energy Laboratory(NREL)) 절차(NREL LAP-002)를 사용하여 조성을 결정하였다.

건조 중량에 기초한 조성은 글루칸(36.82%), 자일란(26.09%), 아라비난(3.51%), 산불용성 리그닌(24.7%) 및 아세틸(2.98%)이었다. 이러한 원료를 나이프(knife)로 분쇄하여, 1 mm 스크린을 통과시켰다. 분쇄된 물질을 6 wt%(건조 고형분 중) 암모니아의 존재 하에 약 160℃에서 90분 동안 전처리하였다. 초기 고형분 로딩은 약 50% 건조물이었다. 처리된 바이오매스를 사용 전에 4℃로 보관하였다.

본 실험에서, 5 mg(단백질)/g(셀룰로스)의 베타-글루코시다제(예를 들어, 트리코테르마 리세이 Bgl1, Fv3C, 및 상동체)를 낮은 β-글루코시다제 발현을 위해 선택되는 통합된 트리코테르마 리세이 균주(H3A) 유래의 전체 셀룰라제의 10 mg(단백질)/g(셀룰로스)의 존재 하에, 회석 암모니아로 전처리된 스위치그래스에 첨가하였다. 효소 혼합물을 50℃에서 2일 동안 기질과 함께 인큐베이션시킨 후에 글루칸 전환%를 측정하고, 결과를 도 51에 나타내었다.

Fv3C는 스위치그래스 기질과 함께, 트리코테르마 리세이 Bgl1 및 아스페르길루스 니게르 Bglu보다 더 잘 수행한 것으로 나타났다.

F. AFEX 옥수수 대에 대한 Fv3C 당화 성능

본 실험에서, 14% 고형분으로 AFEX 옥수수 대의 당화를 증가시키는 트리코테르마 리세이 Bgl1, Fv3C 및 아스페르길루스 니게르 Bglu의 능력을 마이크로타이터 플레이트 당화 검정법(상기 참조)에 기재된 방법에 따라 시험하였다. AFEX로 전처리된 옥수수 대를 미시간 바이오테크놀로지 인스티튜트 인터내셔널(MBI)로부터 취득하였다. www.nrel.gov/biomass/analytical_procedures.html에서 이용가능한 국립재생에너지연구소(NREL) 절차 LAP-002를 사용하여 옥수수 대의 조성을 결정하였다:

건조 중량에 기초한 조성은 글루칸(31.7%), 자일란(19.1%), 갈락탄 (1.83%), 및 아라비난(3.4%)이었다. 이러한 원료는 90℃, 60% 함수율, 1:1 바이오매스 대 암모니아 로딩에서, 30분 동안 18.9 리터(5 갤런) 압력 반응기(Parr)에서 처리된 AFEX였다. 처리된 바이오매스를 반응기로부터 제거하여, 폼 후드(fume hood)에 두어, 잔류 암모니아를 증발시켰다. 처리된 바이오매스를 사용 전에 4℃로 보관하였다.

본 실험에서, 약 5 mg(단백질)/g(셀룰로스)의 베타-글루코시다제(Fv3C 및 상동체)를 β-글루코시다제를 적게 발현하는 통합된 트리코테르마 리세이 균주 유래의 전체 셀룰라제 10 mg(단백질)/g(셀룰로스)의 존재 하에, 전처리된 기질에 첨가하였다(도 3 참조). 효소 혼합물을 50℃에서 2일 동안 기질과 함께 인큐베이션시킨 후에 글루칸 전환%를 측정하고, 결과를 도 52에 나타내었다.

Fv3C는 글루칸 전환에 있어서, 트리코테르마 리세이 Bgl1보다 더 잘 수행한 것으로 관찰되었다. 또한, 상기 조건 하에서 10 mg/g(셀룰로스)의 Fv3C 및 10 mg/g(셀룰로스)의 H3A 전체 셀룰라제가 완전하거나 외관상 완전한 글루칸 전환을 야기한 것을 주목하였다. 1 mg/g(셀룰로스) 미만의 수준에서는, 아스페르길루스 니게르 Bglu(An3A)가 Fv3C 및 트리코테르마 리세이 Bgl1보다 더 높은 글루코스 및 총 글루칸 전환을 제공하는 것으로 보이나, 2.5 mg/g(셀룰로스) 초과인 수준에서는, Fv3C 및 트리코테르마 리세이 Bgl1이 아스페르길루스 니게르 Bglu보다 더 높은 글루코스 및 글루칸 전환을 나타내는 것으로 관찰되었다.

실시예 6: 희석 암모니아로 전처리된 옥수수 속대 당화를 위한 Fv3C 대 전체 셀룰라제 비의 최적화

본 실험에서, Fv3C 대 전체 셀룰라제의 비를 달라지게 하여, 헤미셀룰라제 조성물 중의 Fv3C 대 전체 셀룰라제의 최적비를 측정하였다. 희석 암모니아로 전처리된 옥수수 속대를 기질로 사용하였다. 헤미셀룰라제 조성물 중의 베타-글루코시다제(예를 들어, 트리코테르마 리세이 Bgl1, Fv3C, 아스페르길루스 니게르 Bglu) 대 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제의 비는 0 내지 50%로 다양하였다. 혼합물을 20 mg(단백질)/g(셀룰로스)으로 첨가하여, 암모니아로 전처리된 옥수수 속대를 20% 고형분으로 가수분해시켰다. 결과는 도 53a 내지 53c에 나타내었다.

트리코테르마 리세이 Bgl1 대 전체 셀룰라제의 최적비는 광범위하나, 약 10%에 집중되어 있으며, 50% 혼합물은 전체 셀룰라제 단독의 동일한 로딩과 유사한 성능을 제공하였다. 대조적으로, 아스페르길루스 니게르 Bglu는 약 5%에서 최적에 도달하였으며, 피크는 보다 뾰족하였다. 피크/최적 수준에서, 아스페르길루스 니게르 Bglu는 트리코테르마 리세이 Bglu를 포함하는 최적의 믹스보다 높은 전환을 제공하였다.

Fv3C 대 전체 셀룰라제의 최적비는 약 25%인 것으로 측정되었으며, 혼합물은 20 mg(총 단백질)/g(셀룰로스)에서, 96% 초과인 글루칸 전환을 제공하였다. 따라서, 전체 셀룰라제 중의 효소의 25%는 단일의 효소, Fv3C로 대체되어, 향상된 당화 성능이 야기될 수 있다.

실시예 7: 상이한 효소 배합물에 의한 암모니아로 전처리된 옥수수 속대의 당화

트리코테르마 리세이 통합 균주(H3A) 혼합물로부터의 25% Fv3C/75% 전체 셀룰라제를 용량 반응 실험에서 다른 고 성능 셀룰라제 혼합물과 비교하였다. 트리코테르마 리세이 통합 균주(H3A) 단독 유래의 전체 셀룰라제, 25%의 Fv3C/75%의 트리코테르마 리세이 통합 균주(H3A)로부터의 전체 셀룰라제 혼합물, 아셀레라제(등록 상표) 1500 + 멀티펙트(등록 상표) 자일라나제를 20% 고형분에서 희석 암모니아로 전처리된 옥수수 속대에 대한 그들의 당화 성능에 대하여 비교하였다. 효소 배합물을 반응에 2.5 내지 40 mg(단백질)/g(셀룰로스)으로 투여하였다. 결과를 도 54에 나타내었다.

25%의 Fv3C/75%의 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제 혼합물은 아셀레라제(등록 상표) 1500 + 멀티펙트(등록 상표) 자일라나제 배합물보다 훨씬 더 잘 수행되었으며, 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제와 비교하여 실질적인 향상을 보였다. 각 효소 믹스로부터의 70, 80 또는 90% 글루칸 전환에 필요한 용량은 도 7에 열거되어 있다. 70% 글루칸 전환에서, 25%의 Fv3C/75%의 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제 혼합물은 아셀레라제(등록 상표) 1500 + 멀티펙트(등록 상표) 자일라나제

배합물과 비교시 3.2배 용량 감소를 제공하였다. 70, 80 또는 90% 글루칸 전환에서, 25%의 Fv3C/75%의 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제 혼합물은 트리코테르마 리세이 통합 균주(H3A) 유래의 전체 셀룰라제 단독보다 약 1.8배 적은 효소를 필요로 하였다.

실시예 8: 아스페르길루스 니게르 균주에서의 Fv3C의 발현

아스페르길루스 니게르에서 Fv3C를 발현시키기 위하여, pENTR-Fv3C 플라스미드를 게이트웨이 LR 재조합 반응(인비트로젠)을 사용하여 미국 특허 제7459299호에 기재된 바와 같이 데스티네이션 벡터 pRAXdest2와 재조합시켰다. 발현 플라스미드는 아스페르길루스 니게르 글루코아밀라제 프로모터 및 터미네이터의 제어 하의 Fv3C 게놈 서열, 선택 마커로서 아스페르길루스 니돌란스 pyrG 유전자 및 진균 세포에서의 자율 복제를 위한 아스페르길루스 니돌란스 ama1 서열을 포함하였다. 생성된 재조합 산물을 에스케리키아 콜라이 맥스 에피션시(Max Efficiency) DH5 α (인비트로젠)로 형질전환시키고, 발현 구축물 pRAX2-Fv3C(도 55a)를 함유하는 클론을 16 g/L의 박토 트립톤(Bacto Tryptone)(디프코(Difco)), 10 g/L의 박토 이스트 익스트랙트(Bacto Yeast Extract)(디프코), 5 g/L의 NaCl, 16 g/L의 박토 아가(Bacto Agar)(디프코) 및 100 μ g/mL의 앰피실린으로 제조된 2xYT 아가 플레이트 상에서 선택하였다.

약 50 내지 100 mg의 발현 플라스미드를 아스페르길루스 니게르 변종 아와모리 균주로 형질전환시켰다(미국 특허 제7459299호 참조). 내인성 글루코아밀라제 *glcA* 유전자를 이러한 균주로부터 결실시켰으며, 이는 *pyrG* 유전자 내의 돌연변이를 지니며, 이는 우리딘 원형양성(prototrophy)에 대하여 형질전환체를 선택하는 것을 가능하게 하였다. 아스페르길루스 니게르 형질전환체를 37°C에서 4 및 5일 동안 MM 배지(트리코테르마 리세이 형질전환에 사용된 것과 동일하나, 질소원으로서 아세트아미드 대신에 10 mM NH_4Cl 을 사용한 최소 배지) 상에서 성장시키고, 상이한 형질전환 플레이트로부터의 전체 포자 집단(약 10^6 개 포자/mL)을 사용하여, 하기의 생성 배지를 함유하는 진탕 플라스크에 접종하였다(1L 당): 12 g의 트립톤; 8 g의 소이톤(soyton); 15 g의 $(\text{NH}_4)_2\text{SO}_4$; 12.1 g의 $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$; 2.19 g의 $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$; 1 g의 $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 1 mL의 트윈(Tween) 80; 150 g의 말토스; pH 5.8. 200 rpm에서의 진탕 및 30°C에서의 3일의 발효 후에, 형질전환체에서의 Fv3C의 발현을 SDS-PAGE로 확인하였다.

실시예 9: 트리코테르마 리세이 BGL3(Tr3B)의 성능

PASC 및 PCS 상에서의 전체 셀룰라제/트리코테르마 리세이 Bgl3 배합물을 사용한 당화

RL-P37로부터 유래되고(문헌[Sheir-Neiss, G. *et al.* Appl. Microbiol. Biotechnol. 1984, 20:46-53]), 높은 셀룰라제 생성을 위해 선택된 트리코테르마 리세이 돌연변이체 균주 유래의 정제된 전체 셀룰라제 발효 브로쓰를 이러한 실험의 백그라운드에 사용하였다. 전체 셀룰라제 및 정제된 트리코테르마 리세이 Bgl3(Tr3B)을 기질 중의 g(셀룰로스) 당 mg(총 단백질)에 기초하여 당화 검정에 로딩하였다. 정제된 트리코테르마 리세이 Bgl3을 0 내지 100% Bgl3의 수준으로 전체 셀룰라제와 배합하였다. 혼합물을 20 mg(단백질)/g(셀룰로스)으로 로딩하였다. 각 시료에 대하여 3별로 시험하였다.

인산 팽윤된 셀룰로스(PASC)를 문헌[Walseth, TAPPI 1971, 35:228] 및 문헌[Wood, Biochem. J. 1971, 121:353-362]의 변형된 프로토콜을 사용하여 아비셀 PH-101로부터 제조하였다. 약술하면, 25 아비셀을 진한 인산에서 용해시킨 다음, 차가운 탈이온수를 사용하여 침전시켰다. 셀룰로스를 수집하고, 더 많은 물로 세척하여, pH를 중화시킨 후에, 이를 50 mM 아세트산나트륨 완충제, pH 5.0 중의 1% 고형분으로 희석시켰다. 20 μ L의 희석된 효소 혼합물을 플랫 보텀 마이크로타이터 플레이트의 개별 웰에 첨가하였다. 리피터(repeater) 피펫을 사용하여, 웰당 150 μ L의 기질을 첨가하여, 플레이트를 2개의 알루미늄 플레이트 시일러로 커버하였다.

물은 산으로 전처리된 옥수수 대(상기 참조)를 50 mM 아세트산나트륨 pH 5 완충제 중의 7% 셀룰로스로 희석시켜, 혼합물의 pH를 5.0으로 조절하였다. 리피터 피펫을 사용하여, 150 μ L의 기질을 플랫 보텀 마이크로타이터 플레이트의 각 웰에 첨가하였다. 20 μ L의 희석된 효소 혼합물을 개별 웰에 첨가하여, 플레이트를 2개의 알루미늄 플레이트 시일러로 커버하였다.

이들 플레이트를 700 rpm으로 혼합하면서 37°C 또는 50°C에서 인큐베이션하였다. PASC를 2시간 동안 인큐베이션하여, PCS 플레이트를 48시간 동안 인큐베이션하였다. 100 μ L의 100 mM 글리신 완충제, pH 10을 개별 웰에 첨가하여, 반응을 종결시켰다. 완전한 혼합 후에, 플레이트의 내용물을 여과하여, 상청액을 100 μ L의 10 mM 글리신, pH 10을 함유하는 HPLC 플레이트로 6배 희석하였다. 그 다음에 생성되는 가용성 당의 농도를, 85°C로 유지되는 탈회분화(de-ashing)/가드 컬럼(바이오래드 #125-0118) 및 아미넥스 HPX-87P 탄수화물 컬럼을 구비한

HPLC(아질런트 1100 시리즈)를 사용하여 측정하였다. 이동상은 0.6 mL/min 유속의 물이었다. 글루칸 전환율은 본 명세서에서 $100 \times [\text{mg 글루코스} + (\text{mg 셀로비오스} \times 1.056)] / [\text{기질 중의 mg 셀룰로스} \times 1.111]$ 로 정의된다. 따라서, 전환율(%)은 가수분해수에 대하여 보정되었다. 50°C에서의 PASC의 당화에서의 전체 셀룰라제:트리코테르마 리세이 Bgl3 혼합물의 성능 결과는 도 64a에 나타낸다. 37°C에서의 PASC의 당화에서의 전체 셀룰라제:트리코테르마 리세이 Bgl3 혼합물의 성능 결과는 도 64b에 나타낸다. 50°C에서의 산으로 재처리된 옥수수 대의 당화에서의 전체 셀룰라제:트리코테르마 리세이 Bgl3 혼합물의 성능 결과는 도 64c에 나타낸다. 37°C에서의 산으로 재처리된 옥수수 대의 당화에서의 전체 셀룰라제:트리코테르마 리세이 Bgl3 혼합물의 성능 결과는 도 64d에 나타낸다.

B. PASC에 대한 전체 셀룰라제 백그라운드를 이용한 Bgl3의 용량 반응

RL-P37로부터 유래되고(문헌[Sheir-Neiss, G *et al.* Appl. Microbiol. Biotechnol. 1984, 20:46-53]), 셀룰라제 생성을 위해 선택된 트리코테르마 리세이 돌연변이체 균주 유래의 정제된 전체 셀룰라제 발효 브로쓰를 이러한 실험의 백그라운드에 사용하였다.

전체 셀룰라제 및 정제된 트리코테르마 리세이 Bgl3을 기질 중의 g(셀룰로스) 당 mg(총 단백질)에 기초하여 당화 검정에 로딩하였다. 정제된 트리코테르마 리세이 Bgl3을 0 내지 10 mg(단백질)/g(셀룰로스)의 양으로 로딩하였다. 일정한 수준의 10 mg(전체 셀룰라제 단백질)/g(셀룰로스)도 각 시료에 첨가하였다. 각 시료를 3별로 시험하였다.

인산 팽윤된 셀룰로스 기질을 50 mM 아세트산나트륨 pH 5 완충제 중의 1% 셀룰로스로 희석하고, pH를 5.0으로 조절하였다. 20 μL 의 희석된 효소 혼합물을 플랫 보텀 마이크로타이터 플레이트의 개별 웰에 첨가하였다. 리피터 피펫을 사용하여, 150 μL 의 기질을 개별 웰에 첨가하여, 플레이트를 2개의 알루미늄 플레이트 시일러로 커버하였다. 그 다음에 플레이트를 700 rpm으로 1시간 동안 혼합하면서 50°C에서 인큐베이션하였다.

100 μL 의 100 mM 글리신 완충제, pH 10을 개별 웰에 첨가하여, 반응을 종결시켰다. 완전한 혼합 후에, 플레이트의 내용물을 여과하여, 상청액을 100 μL 의 10 mM 글리신, pH 10을 함유하는 HPLC 플레이트로 6배 희석하였다. 그 다음에 생성되는 가용성 당의 농도를, 85°C로 유지되는 탈회분화/가드 컬럼(바이오래드 #125-0118) 및 아미넥스 HPX-87P 탄수화물 컬럼을 구비한 HPLC(아질런트 1100 시리즈)를 사용하여 측정하였다. 이동상은 0.6 mL/min 유속의 물이었다.

글루칸 전환율은 본 명세서에서 $100 \times [\text{mg 글루코스} + (\text{mg 셀로비오스} \times 1.056)] / [\text{기질 중의 mg 셀룰로스} \times 1.111]$ 로 정의된다. 따라서, 전환율(%)은 가수분해수에 대하여 보정되었다. 인산 팽윤된 셀룰로스의 당화에서의 트리코테르마 리세이 Bgl1 및 트리코테르마 리세이 Bgl3의 용량 반응 비교를 도 65a에 나타낸다. 인산 팽윤된 셀룰로스의 당화에서의 트리코테르마 리세이 Bgl1 및 트리코테르마 리세이 Bgl3에 의해 생성된 셀로비오스 및 글루코스의 비교를 도 65b에 나타낸다.

실시예 10: 키메라 β -글루코시다제

A. 트리코테르마 리세이에서의 발현

야생형 Fv3C C-말단 서열 부분을 트리코테르마 리세이 β -글루코시다제, Bgl3(Tr3B) 유래의 C-말단 서열로 교체하였다. 구체적으로, Fv3C의 잔기 1 내지 691을 나타내는 연속 스트레치를 Bgl3의 잔기 668 내지 874를 나타내는 연속 스트레치와 융합하였다. Fv3C/Bgl3 키메라/융합 폴리펩티드를 암호화하는 유전자의 개략도를 도 60a에 나타낸다. 융합/키메라 폴리펩티드 Fv3C/Bgl3를 암호화하는 아미노산 서열 및 폴리뉴클레오티드 서열을 도 60b 및 60c에 나타낸다.

키메라/융합 분자를 융합 PCR을 사용하여 구축하였다. 게놈 Fv3C 및 Bgl3 암호화 서열의 pENTR 클론을 PCR 주형으로서 사용하였다. 엔트리 클론 둘 모두를 pDonor221 벡터(인비트로젠)에서 구축하였다. 융합 산물을 2 단계로 조립하였다. 먼저, Fv3C 키메라 부분을 주형으로서의 pENTR Fv3C 클론 및 하기 올리고뉴클레오티드 프라이머를 사용하여 PCR 반응에서 증폭시켰다:

pDonor 정방향: 5'-GCTAGCATGGATGTTTCCAGTCACGACGTGTGAAAACGACGGC-3' (서열 번호 122)

Fv3C/Bgl3 역방향: 5'-GGAGGTTGGAGAACTTGAACGTCGACCAAGATAGACCGTGA CCGAAC TCGTAG 3' (서열 번호 123)

Bgl3 키메라 부분을 하기 올리고뉴클레오티드 프라이머를 사용하여 pENTR Bgl3 벡터로부터 증폭시켰다:

pDonor 역방향: 5'-TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG-3' (서열 번호 124)

Fv3C/Bg13 정방향: 5'-CTACGAGTTCGGTCACGGTCTATCTTGGTCGACGTTCAAGTTC TCCAACCTCC-3'(서열 번호 125)

제2 단계에서, 등몰량의 PCR 산물(각각, 약 1 μ l 및 0.2 μ l의 초기 PCR 반응물)을 하기와 같이 네스티드(nested) 프라이머 세트를 사용하여 이후의 융합 PCR 반응을 위한 주형으로서 첨가하였다:

Att L1 정방향: 5' TAAGCTCGGGCCCCAAATAATGATTTTATTTTACTGATAGT 3'(서열 번호 126)

AttL2 역방향: 5'GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTACTGATA 3'(서열 번호 127)

PCR 반응을 하이 피델리티 퓨전(high fidelity Phusion) DNA 중합효소(핀자임즈 오와이(Finnzymes OY))를 사용하여 수행하였다. 생성된 융합된 PCR 산물은 양쪽 말단에 무순상 게이트웨이-특이적 attL1, attL2 재조합 부위를 함유하여, 게이트웨이 LR 재조합 반응(인비트로젠)을 통하여 최종 데스티네이션 벡터로 직접적으로 클로닝되게 하였다.

0.8% 아가로스 겔 상에서의 DNA 단편의 분리 후에, 단편을 뉴클레오스핀(Nucleospin)(등록 상표) 엑스트랙트(Extract) PCR 클린-업(clean-up) 키트(마슈레이-나겔 게엠베하 운트 코 카게(Macherey-Nagel GmbH & Co. KG))를 사용하여 정제하고, 각 단편 100 ng을 pTTT-pyrG13 데스티네이션 벡터 및 LR 클로나제(상표명) II 효소믹스(인비트로젠)를 사용하여 재조합시켰다. 생성된 재조합 산물을 에스케리키아 콜라이 맥스 에피션시 DH5 α (인비트로젠)로 형질전환시키고, 키메라 β -글루코시다제를 함유하는 발현 구축물 pTTT-pyrG13-Fv3C/Bg13 융합(도 61)을 함유하는 클론을 16 g/L의 박토 트립톤(디프코), 10 g/L의 박토 이스트 엑스트랙트(디프코), 5 g/L의 NaCl, 16 g/L의 박토 아가(디프코), 및 100 μ g/mL의 앰피실린을 사용하여 제조된 2xYT 아가 플레이트 상에서 선택하였다. 100 μ g/ml 앰피실린을 함유하는 2xYT 배지에서 박테리아를 성장시켰다. 그 후에, 플라스미드를 단리시켜, BglI 또는 EcoRV 중 어느 하나로 제한 효소로 소화시켰다. 얻어진 Fv3C/Bg13 영역을 확인을 위해 ABI3100 서열 분석기(어플라이드 바이오시스템즈(Applied Biosystems))를 사용하여 시퀀싱하였다. 확인된 제한 패턴 및 정확한 서열을 갖는 플라스미드를 하기와 같이, 추가의 PCR 반응에서 주형으로 사용하고, 하이 피델리티 퓨전 DNA 중합효소(핀자임즈 오와이) 및 프라이머를 사용하여 DNA 단편을 생성하였다:

Cbh1 정방향: 5' GAGTTGTGAAGTCGGTAATCCCGCTG 3'(서열 번호 128)

AmdS 역방향: 5' CCTGCACGAGGGCATCAAGCTCACTAACCG 3'(서열 번호 129)

생성된 단편은 *cbh1* 프로모터 및 터미네이터의 제어 하에서 Fv3C/Bg13 암호화 영역을 포함하였다. 구체적으로, 0.5 내지 1 μ g의 이러한 단편을 후술되는 약간의 변형을 가한 PEG-프로토플라스트(Protoplast) 방법을 사용하여 트리코테르마 리세이 핵사-겔실 균주(상기 참조)로 형질전환시켰다. 프로토플라스트 제조를 위해, 포자를 트리코테르마 최소 배지 MM - 20 g/L 글루코스, 15 g/L KH₂PO₄, pH 4.5, 5 g/L (NH₄)₂SO₄, 0.6 g/L MgSO₄·7H₂O, 0.6 g/L CaCl₂·2H₂O, 1 mL의 1000 X 트리코테르마 리세이 미량 원소 용액(5 g/L FeSO₄·7H₂O, 1.4 g/L ZnSO₄·7H₂O, 1.6 g/L MnSO₄·xH₂O, 3.7 g/L CoCl₂·x6H₂O를 함유함)을 함유함 - 에서 150 rpm으로 진탕하면서 24°C에서 16 내지 24시간 동안 성장시켰다. 발아 포자를 원심분리에 의해 수집하고, 50 mg/mL의 글루카넥스(Glucanex) G200 (노보자임즈 아게(Novozymes AG)) 용액으로 처리하여, 진균 세포벽을 용해시켰다. 추가의 프로토플라스트 제조를 문헌[Penttilä *et al.* Gene 61(1987)155-164]에 기재된 방법에 따라 수행하였다.

총 부피 200 μ l 중에 약 1 μ g의 DNA 및 1-5 x 10⁷ 프로토플라스트를 함유한 형질전환 혼합물을 각각 2 mL의 25% PEG 용액으로 처리하고, 1.2 M 소르비톨/10 mM 트리스, pH 7.5, 10 mM CaCl₂의 2개의 부피로 희석하여, 5 mM 우리딘 및 20 mM 아세트아미드를 함유하는 3% 선택적 탐 아가로스 MM과 혼합하였다. 얻어진 혼합물을 우리딘 및 아세트아미드를 함유하는 2% 선택적 아가로스 플레이트 상에 부었다. 단일 형질전환체를 우리딘 및 아세트아미드를 함유하는 새로운 MM 플레이트 상에 다시 취하기 전에, 플레이트를 추가로 28°C에서 7 내지 10일간 인큐베이션하였다. 독립 클론으로부터의 포자를 사용하여, 96-웰 마이크로타이터 플레이트 또는 진탕 플라스크에서 발효 배지를 접종시켰다.

4.7 g/L (NH₄)₂SO₄, 33 g/L 1,4-피페라진비스(프로판설폰산), pH 5.5, 6.0 g/L 글리신, 5.0 g/L KH₂PO₄, 1.0 g/L CaCl₂·2H₂O, 1.0 g/L MgSO₄·7H₂O, 2.5 ml/L의 400X 트리코테르마 리세이 미량 원소 용액, 20 g/L 글루코스, 및 6.5 g/L 소포로스를 함유하는 글리신 생성 배지 250 μ l를 함유하는 96 웰 필터 플레이트(코닝)를 Fv3C/Bg13 하이브리드(웰당 10⁴ 개 초과 포자)를 발현하는 트리코테르마 리세이 형질전환체의 포자 현탁액을 사용하여 접종시켰다. 플레이트를 28°C에서 약 80% 습도에서 6 내지 8일간 인큐베이션하였다. 배양 상청액을 진공 여과에

의해 수집하여, 하이브리드의 성능 및 이의 발현 수준을 시험하는데 사용하였다. 전체 브로쓰 시료의 단백질을 프로파일을 PAGE 전기영동으로 측정하였다. 20 μ l의 배양 상청액을 환원제 없이 8 μ l의 4X 시료 로딩 완충제와 혼합하였다. 시료를 MES SDS 러닝(Running) 완충제(인비트로젠)를 사용하여 뉴페이지(등록상표)노박스(Novex) 10% 비스-트리스 겔 상에서 분리하였다.

이것에 의해, 트리코테르마 리세이에서 발현될 때에나 저장 동안에 프로테아제 분해에 덜 민감한 Fv3C/Bg13(FB) 키메라 β -글루코시다제를 얻었다. 마이크로타이터 플레이트에서의 8일간의 발효 후에, 동등한 조건 하에서 Fv3C β -글루코시다제와 비교하여, 발현된 β -글루코시다제의 분해가 Fv3C/Bg13(FB) 키메라에서 훨씬 적게 관찰되었다.

B. 크리스스포리움 톡크노웬스 숙주 세포에서의 Fv3C 및 FAB의 발현

발현 카세트의 구축

트리코테르마 리세이(pTrex6g/Fv3c, 실시예 3, 도 45b) 및 아스페르길루스 니게르(pRAX2-Fv3C, 실시예 8, 도 55a)에 대하여 기재된 Fv3C 발현 벡터를 사용하여, 크리스스포리움 톡크노웬스에서 Fv3C 또는 FAB를 발현시켰다. 고유 Fv3C 신호 서열을 사용하였다. 벡터 pRAX2-Fv3C는 아스페르길루스 니게르 글루코아밀라제 프로모터 및 터미네이터 서열의 제어 하의 fv3c 유전자 서열, 선택 마커로서 아스페르길루스 니돌란스 pyrG 유전자, 및 진균 세포에서의 자율 복제를 위한 아스페르길루스 니돌란스 *ama1* 서열을 포함하였다. 벡터 pTrex6g/Fv3c는 트리코테르마 리세이 *cbhI* 프로모터 및 터미네이터 서열의 제어 하의 Fv3C 오픈 리딩 프레임, 및 그의 고유 프로모터 및 터미네이터를 갖는 트리코테르마 리세이 돌연변이된 아세트락테이트 신타제 선택 마커(*als*)를 함유하였다. 대안적으로, 플레오마이신 또는 하이그로마이신 내성과 같은 선택 마커, 또는 영양 선택 마커 아세트아미다제(*amdS*)도 사용될 수 있다.

크리스스포리움 톡크노웬스의 형질전환

크리스스포리움 톡크노웬스 숙주 세포를 예를 들어, 미국 특허 제6,573,086호에 기재된 것과 같은 당업계에 공지된 변형을 가한 문헌[Penttilä *et al.* *Gene* 61(1987)155-164]에 기재된 원형질체 융합에 의해 pTrex6g/Fv3C로 형질전환시켰다. 그 다음에 내성 형질전환체는 새로운 클로리무론 에틸 플레이트 상에서 선택될 수 있다. 대안적으로, pyrG-(우리딘 영양 요구성) 크리스스포리움 톡크노웬스 숙주 세포는 실시예 8(상기 참조)에 기재된 바와 같이, 원형질체 융합에 의해 pRAX2-Fv3C로 형질전환되어 우리딘 원형양성을 위해 선택될 수 있다.

단백질 생성을 위한 크리스스포리움 톡크노웬스의 배양

Fv3C 및 FAB를 예를 들어, WO 98/15633에 기재된 배지에서 크리스스포리움 톡크노웬스 형질전환체를 약 5일간 진탕하면서 27 내지 40°C, pH 5 내지 10에서 배양시켜 생성하고, 셀룰로스 또는 락토스를 사용하여, CBHI 프로모터를 유도하거나, 말토스, 말트린 또는 전분을 사용하여 글루코아밀라제 프로모터를 유도하였다.

실시예 11: 키메라 베타-글루코시다제

SDS-PAGE 및 펩티드 맵핑 분석에 의해, Fv3C/Bg13 키메라가 트리코테르마 리세이에서 생성될 때에 2개의 단편으로 클립되는 것으로 나타났다. N-말단 시퀀싱은 Fv3C의 전장의 잔기 674와 683 사이의 클립 부위를 나타내었다.

제2 키메라 β -글루코시다제가 구축되었으며, Fv3C 유래의 N-말단 서열, 탈라로마이세스 에메르소니이 Te3A로부터의 제2 β -글루코시다제의 서열 유래의 루프 영역, 및 트리코테르마 리세이 Bg13(또는 Tr3B) 유래의 C-말단 부분 서열을 포함하였다. 이것은 Fv3C/Bg13 키메라의 루프 영역을 교체하여 달성되었다(상기 실시예 10 참조). 구체적으로, Fv3C/Bg13 키메라의 Fv3C 잔기 665 내지 683(RRSPSTDGKSSPNN TAAPL의 서열(서열 번호 157)을 가짐)을 Te3A 잔기 634 내지 640(KYNITPI(서열 번호 158))로 교체하였다. 이러한 하이브리드 분자를 상기 실시예 10에 기재된 바와 같이, 융합 PCR 접근법을 사용하여 구축하였다.

2개의 N-글리코실화 부위, 즉 S725N 및 S751N을 Fv3C/Bg13 골격에 도입하였다. 이들 글리코실화 돌연변이를 주형으로서 pTTT-pyrG13-Fv3C/Bg13 융합 플라스미드(도 61)를 사용하여, 상술한 융합 PCR 증폭 기술을 이용하여 Fv3C/Bg13 골격에 도입하여, 초기 PCR 단편을 생성하였다. 하기 프라이머 쌍을 분리된 PCR 반응에서 첨가하였다:

Pr CbhI 정방향: 5' CGGAATGAGCTAGTAGGCAAAGTCAGC 3'(서열 번호 130) 및

725/751 역방향: 5'-CTCCTTGATGCGCGAACGTTCTTGGGGAAGCCATAGTCCTTAA GGTTCCTTGCTGAAGTTGCCAGAGAG 3'(서열 번호 725/751)

호 131)

725/751 정방향: 5'-GGCTTCCCCAAGAACGTTTCGCCGATCAAGGAGTTTATCTACC CCTACCTGAACACCACTACCTC 3' (서열 번호 132), 및

Ter CbhI 역방향: 5' GATACACGAAGAGCGGCGATTCTACGG 3'(서열 번호 133).

다음에, PCR 단편을 Pr CbhI 정방향 및 Ter CbhI 프라이머를 사용하여 융합하였다. 생성된 융합 산물은 2개의 원하는 글리코실화 부위를 포함할 뿐만 아니라, 무손상 attB1 및 attB2 부위를 함유하여, 게이트웨이 BP 재조합 반응(인비트로젠)을 사용하여 pDonor221 벡터와 재조합되게 하였다. 이것에 의해, pENTR-Fv3C/Bg13/ S725N S751N 클론을 생성시킨 다음에, 삼원(triple) 하이브리드 분자 Fv3C/Te3A/Bg13를 구축하기 위한 골격으로서 사용하였다.

잔기 665 내지 683에서의 Fv3C/Bg13 하이브리드의 루프를 Te3A 유래의 루프 서열로 교체하기 위해, 일차 PCR 반응을 하기 프라이머 세트를 사용하여 수행하였다:

세트 1: pDonor 정방향: 5'-GCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAA ACGACGGC 3'(서열 번호 122) 및

Te3A 역방향: 5'-GATAGACCGTGACCGAACTCGTAGATAGGCGTGATGTT GTACTTGTCGAAGTGACGGTAGTCGATGAAGAC 3'(서열 번호 160);

세트 2: Te3A2 정방향: 5'-GTCTTCATCGACTACCGTCACTTCGACAAGTACAACATCAC GCCTATCTACGAGTTCGGTCACGGTCTATC-3'(서열 번호 161); 및

pDonor 역방향: 5' TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG 3'(서열 번호 124)

그 다음에, 일차 PCR 반응에서 얻어진 단편을 하기 프라이머를 사용하여 융합하였다:

Att L1 정방향: 5' TAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGT 3'(서열 번호 126) 및

AttL2 역방향: 5'GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTGACTGATA 3'(서열 번호 127).

생성된 PCR 산물은 말단에 무손상 게이트웨이-특이적 attL1, attL2 재조합 부위를 함유하여, 게이트웨이 LR 재조합 반응(인비트로젠)을 사용하여 최종 데스티네이션 벡터로 직접적으로 클로닝되게 하였다.

Fv3C/Te3A/Bg13 암호화 유전자의 DNA 서열은 서열 번호 83에 열거되어 있다. Fv3C/Te3A/Bg13(FAB) 하이브리드의 아미노산 서열은 서열 번호 135에 열거되어 있다. Fv3C/Te3A/Bg13 키메라를 암호화하는 유전자 서열은 상기 실시예 10에 기재된 바와 같이, pTTT-pyrG13 벡터에 클로닝되어, 트리코테르마 리세이 수용 균주에서 발현되었다.

실시예 12: 키메라 베타-글루코시다제의 향상된 안정성

본 실험은 시차주사 열량 측정법(DSC)을 사용하여 다양한 베타-글루코시다제의 열변성 온도를 측정하였다. 구체적으로, 열전이 온도를 정제된 효소 Fv3C/Te3A/Bg13 키메라, Fv3C, 및 트리코테르마 리세이 Bg11에 대하여 측정하였다. 효소를 50 mM 아세트산나트륨 완충제, pH 5.0에서 500 ppm으로 희석하였다. DSC 96-웰 마이크로타이터 플레이트(MicroCal)를 500 μ l의 각 희석된 효소 시료로 로딩하였다. 물 및 완충제 블랭크를 또한 포함하였다. DSC(Auto VP-DSC, MicroCal) 파라미터를 25°C 초기 온도, 및 110°C 최종 온도에서 90°C/h의 주사 속도로 설정하였다. 서모그램은 도 63에 나타나 있다. Fv3C 및 Fv3C/Te3A/Bg13 키메라에 대한 T_m 은 유사하게 나타났으며, 아마도 트리코테르마 리세이 Bg11의 그것보다 약간 낮게 나타날 것이다.

실시예 13: 희석 암모니아로 전처리된 옥수수 속대의 당화에서의 아스페르길루스 니게르로 발현된 FV3C의 활성

통합 균주 H3A-5(낮은 β -글루코시다제 생성자), 아스페르길루스 니게르에서 생성된 Fv3C(실시예 8 참조), 및 정제된 트리코테르마 리세이 Bg11(본 명세서에서 "트리코테르마 리세이 Bglu1" 또는 "Tr3A"로도 명명됨)을 기질 중의 mg(총 단백질)/g(셀룰로스)에 기초한 당화 검정으로 로딩하였다. 베타-글루코시다제를 0 내지 10 mg(단백질)/g(셀룰로스)으로부터 로딩하였다. 일정한 수준의 10 mg/g H3A-5를 각 시료에 첨가하였다. 각각의 시료를 5별로 검정하였다.

희석 암모니아로 전처리된 옥수수 속대 기질을 50 mM 아세트산나트륨 pH 5 완충제 중에서 7% 셀룰로스로 희석하여, pH를 5.0으로 조절하였다. 기질을 96-웰 마이크로타이터 플레이트에 주입하였다(웰당 65 mg). 웰당 30 μ l의 적절히 희석된 효소 믹스를 96-웰 플레이트에 첨가하였다. 효소 믹스의 첨가 후에, 5% 셀룰로스를 함유하고

록 기질을 계산하였다. 플레이트를 2개의 알루미늄 플레이트 시일러로 커버하였다. 그 다음에 모든 플레이트를 50℃ 및 200 rpm에서 48시간 동안 인큐베이터에 두었다.

각 웰에 100 μ l의 100 mM 글리신 완충제, pH 10을 첨가하여, 반응을 종결하였다. 완전한 혼합 후에, 플레이트의 내용물을 원심분리하여, 상청액을 100 μ l의 10 mM 글리신, pH 10을 함유하는 HPLC 플레이트로 11배 희석하였다. 그 다음에 생성되는 가용성 당의 농도를 HPLC로 측정하였다. 아질린트 1100 시리즈 HPLC는 85℃로 유지되는 탈회분화/가드 컬럼(바이오래드 #125-0118) 및 아미넥스 납계 탄수화물 컬럼(아미넥스 HPX-87P)을 구비하였다. 이동상은 0.6 ml/min 유속의 물이었다.

글루칸 전환율은 $100 \times [\text{mg 글루코스} + (\text{mg 셀로비오스} \times 1.056)] / [\text{기질 중의 mg 셀룰로스} \times 1.111]$ 로 정의된다. 이렇게 하여, 가수분해수에 대하여 보정된 전환율(%)은 도 62에 나타나 있다.

실시예 13: Fv3C, FAB 및 트리코테르마 리세이 Bgl1의 기질 결합의 비교

본 실험은 특정 전형적인 바이오매스 기질에 대한 Fv3C, 키메라 β -글루코시다제 분자 FAB, 및 트리코테르마 리세이 Bgl1 각각의 결합을 비교한다.

리그닌 - 페닐프로판노이드의 복합 바이오폴리머 - 은 식물의 세포벽을 경화 및 강화시키도록 셀룰로스 섬유에 결합하는 목재의 비탄수화물 주성분이다. 다른 세포벽 성분에 가교결합되기 때문에, 리그닌은 셀룰로스 분해 효소에 대한 셀룰로스 및 헤미셀룰로스의 접근성을 최소화시킨다. 그러므로, 리그닌은 일반적으로 모든 식물 바이오매스의 소화율 감소와 관련되어 있다. 특히 리그닌에 대한 셀룰라제의 결합은 셀룰라제에 의한 셀룰로스의 분해를 감소시킨다. 리그닌은 소수성을 나타내며, 겔보기엔 음전하를 띈다. FAB, Bgl1, 및 Fv3C 중에서, Fv3C는 최저 pI을 가지며, 최소로 양전하를 띈 반면에, Bgl1은 최고 pI을 가지며 최대로 양전하를 띈다, 리그노셀룰로스 기질에 대한 그들의 결합을 조사하였다.

아셀레라제를 100 mg/g의 셀룰로스 및 8 mg 멀티펙트 자일라나제/g 셀룰로스로 함유하는 당화 혼합물을 사용하여 희석 암모니아로 전처리된 옥수수 속대(DACC) 또는 옥수수 대(DACS) 또는 산으로 전처리된 옥수수 대(PCS 또는 whPCS)의 광범위한 당화를 행하여 리그닌을 회수하였다. 당화 후에 비특이적 세린 프로테아제 첨가에 의해 셀룰라제의 가수분해를 수행하였다. 0.1N HCl을 혼합물에 첨가하여, 프로테아제를 불활성화한 다음에, 아세트이트 완충제(50 mM 아세트산나트륨 pH 5)로 반복 세정하여, 시료를 pH 5로 되게 하였다.

100 μ l의 DACS(약 5% 글루칸에서), DACC(약 5% 글루칸에서), whPCS(약 5% 글루칸에서), DACC로 제조된 리그닌(5% 글루칸에서와 같이), PCS로 제조된 리그닌(5% 글루칸에서와 같이), 또는 50 mM 아세트산나트륨 pH 5 완충제 대조군을 마이크로타이터 플레이트에서 100 μ l의 150 μ g/mL FAB, 트리코테르마 리세이 Bgl1, 또는 Fv3C와 배합한 다음에, 밀봉하여, 50℃에서 44시간 동안 인큐베이션하였다. 마이크로타이터 플레이트를 고속으로 원심분리하여, 가용성 물질을 불용성 물질과 분리하였다. 가용성 분획 중의 효소 활성을 측정하였다. 간단히 말하면, 상청액을 5배로 희석한 다음에, 20 μ l를 80 μ l의 2 mM 2-클로로-4-니트로페닐 β -D-글루코피라노시드(CNPG)에 첨가하여, 실온에서 6분간 인큐베이션하였다. 100 μ l의 500 mM Na₂CO₃ pH 9.5를 첨가하여, 반응물을 쉐킹하였다. OD405를 리딩하였다. 비결합 베타-글루코시다제의 비율을 리그닌 및 바이오매스 기질의 부재 하에 동일한 방법으로 인큐베이션된 대조 시료의 OD405로 나눈 가용성 분획 중의 베타-글루코시다제 활성의 OD405를 사용하여 계산하였다.

결합 및 비결합 β -글루코시다제의 총 활성을 측정하였다. 마이크로타이터 플레이트를 다시 혼합하고, 20 μ l의 분취량을 각각 80 μ l의 아세트산나트륨 완충제 pH 5에 첨가하며, 20 μ l의 희석된 믹스를 80 μ l의 2 mM 2-클로로-4-니트로페닐 β -D-글루코피라노시드(CNPG)에 첨가하여 실온에서 6분간 인큐베이션하고, 100 μ l의 500 mM Na₂CO₃ pH 9.5를 첨가하여 반응물을 쉐킹하였다. 반응 혼합물을 침강시키고, 100 μ l의 상청액을 새로운 마이크로타이터 플레이트에 옮겼다. OD405를 측정하였다. 바이오매스 또는 리그닌의 존재 하에서의 상대적 총 β -글루코시다제 활성을 리그닌 및 바이오매스 기질의 부재 하에 동일한 방법으로 인큐베이션된 대조 시료의 OD405로 나눈 총 믹스의 OD405를 사용하여 계산하였다.

결합 베타-글루코시다제가 측정 시간 프레임에서 해리하지 않았다는 것을 확인하기 위해, 20 μ l 분취량을 리믹스된 마이크로타이터 플레이트로부터 새로운 마이크로타이터 플레이트의 80 μ l의 아세트산나트륨 완충제 pH 5로 꺼내어, 플레이트를 베타-글루코시다제가 바이오매스 또는 리그닌으로부터 해리되도록 반시간 동안 진탕시키면서 실온에서 인큐베이션하였다. 그 다음에 플레이트를 원심분리하여, 상청액 중의 베타-글루코시다제 활성을 상술한 바와 같이 측정하였다. 게다가, 비결합 베타-글루코시다제를 계산하였다.

Fv3C는 바이오매스 기질 또는 리그닌에 대한 최소 결합을 나타내는 반면에, FAB 및 트리코테르마 리세이 1 둘 모두는 바이오매스 기질 및 리그닌에 대한 고 결합 수준을 나타내었다(도 71a). 이들 세가지의 β -글루코시다제 중 어느 것도 DACC에 결합되지 않으나, 트리코테르마 리세이 및 FAB 둘 모두는 DACC의 완전 당화로부터 제조되는 리그닌에 결합되었다. 의외로, 결합 FAB 또는 트리코테르마 리세이 Bg11은 유리 FAB 또는 Bg11에 비하여, 여전히 약 50 내지 80% 활성을 나타내었다(도 71b). 또한 결합 FAB가 바이오매스 또는 리그닌으로부터 해리되지 않으나, 약 20% Bg11이 30분간의 인큐베이션 기간 동안에 결합 상태에서 비결합 상태로 해리되지 않는 것으로 관찰되었다(도 71c).

도면

도면1a

서열 번호	NT 또는 AA	명칭
1.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH3 효소, Fv3A의 뉴클레오티드 서열
2.	아미노산	Fv3A의 단백질 서열
3.	뉴클레오티드	페니실리움 푸니쿨로숨 유래의 GH43 효소, Pf43A의 뉴클레오티드 서열
4.	아미노산	Pf43A의 단백질 서열
5.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43E의 뉴클레오티드 서열
6.	아미노산	Fv43E의 단백질 서열
7.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH39 효소, Fv39A의 뉴클레오티드 서열
8.	아미노산	Fv39A의 단백질 서열
9.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43A의 뉴클레오티드 서열
10.	아미노산	Fv43A의 단백질 서열
11.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43B의 뉴클레오티드 서열
12.	아미노산	Fv43B의 단백질 서열
13.	뉴클레오티드	포도스포라 안세리나 유래의 GH51 효소, Pa51A의 뉴클레오티드 서열
14.	아미노산	Pa51A의 단백질 서열
15.	뉴클레오티드	지베렐라 제아에 유래의 GH43 효소, Gz43A의 뉴클레오티드 서열
16.	아미노산	Gz43A의 단백질 서열
17.	뉴클레오티드	푸사리움 옥시스포룸 유래의 GH43 효소, Fo43A의 뉴클레오티드 서열
18.	아미노산	Fo43A의 단백질 서열
19.	뉴클레오티드	아스페르길루스 푸미가투스 유래의 GH43 효소, Af43A의 뉴클레오티드 서열
20.	아미노산	Af43A의 단백질 서열
21.	뉴클레오티드	페니실리움 푸니쿨로숨 유래의 GH51 효소, Pf51A의 뉴클레오티드 서열
22.	아미노산	Pf51A의 단백질 서열
23.	뉴클레오티드	아스페르길루스 푸미가투스 유래의 GH11 효소, AfuXyn2의 뉴클레오티드 서열
24.	아미노산	AfuXyn2의 단백질 서열
25.	뉴클레오티드	아스페르길루스 푸미가투스 유래의 GH11 효소, AfuXyn5의 뉴클레오티드 서열
26.	아미노산	AfuXyn5의 단백질 서열

도면1b

서열 번호	NT 또는 AA	명칭
27.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43D의 뉴클레오티드 서열
28.	아미노산	Fv43D의 단백질 서열
29.	뉴클레오티드	페니실리움 푸니쿨로sum 유래의 GH43 효소, Pf43B의 뉴클레오티드 서열
30.	아미노산	Pf43B의 단백질 서열
31.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH51 효소, Fv51A의 뉴클레오티드 서열
32.	아미노산	Fv51A의 단백질 서열
33.	뉴클레오티드	케토미움 글로보sum(C. globosum) 유래의 GH51 효소, Cg51B의 뉴클레오티드 서열
34.	아미노산	Cg51B의 단백질 서열
35.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43C의 뉴클레오티드 서열
36.	아미노산	Fv43C 단백질 서열
37.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH30 효소, Fv30A의 뉴클레오티드 서열
38.	아미노산	Fv30A 단백질 서열
39.	뉴클레오티드	푸사리움 베르티실리오이테스 유래의 GH43 효소, Fv43F의 뉴클레오티드 서열
40.	아미노산	Fv43F 단백질 서열
41.	뉴클레오티드	트리코테르마 리세이 유래의 GH10 자일라나제, Xyn3의 뉴클레오티드 서열
42.	아미노산	Xyn3 단백질 서열
43.	아미노산	트리코테르마 리세이 유래의 GH11 자일라나제, Xyn2의 단백질 서열
44.	아미노산	트리코테르마 리세이 유래의 GH3 β -자일로시다제, Bx11의 단백질 서열
45.	아미노산	트리코테르마 리세이 유래의 GB β -글루코시다제, Bg11의 단백질 서열
46.	뉴클레오티드	Pa51A의 추정된 cDNA
47.	뉴클레오티드	Pa51A에 대한 코돈 최적화된 cDNA
48.	뉴클레오티드	성숙 Gz43A를 암호화하는 게놈 DNA의 CBH1 신호 서열 업스트림의 암호화 서열
49.	뉴클레오티드	성숙 Fo43A를 암호화하는 게놈 DNA의 CBH1 신호 서열 업스트림의 암호화 서열
50.	뉴클레오티드	Pf51A를 암호화하는 코돈 최적화된 DNA의 CBH1 신호 서열 업스트림의 뉴클레오티드 서열
51.	뉴클레오티드	트리코테르마 리세이 유래의 엔도글루카나제, Eg4의 뉴클레오티드 서열

도면1c

서열 번호	NT 또는 AA	명칭
52.	아미노산	Eg4의 단백질 서열
53.	뉴클레오타이드	포도스포라 안세리나 유래의 GH3 β -글루코시다제, Pa3D의 뉴클레오타이드 서열
54.	아미노산	Pa3D의 단백질 서열
55.	뉴클레오타이드	푸사리움 베르티실리오이테스 유래의 GH3 β -글루코시다제, Fv3G의 뉴클레오타이드 서열
56.	아미노산	Fv3G의 단백질 서열
57.	뉴클레오타이드	푸사리움 베르티실리오이테스 유래의 GH3 β -글루코시다제, Fv3D의 뉴클레오타이드 서열
58.	아미노산	Fv3D의 단백질 서열
59.	뉴클레오타이드	푸사리움 베르티실리오이테스 유래의 GH3 β -글루코시다제, Fv3C의 뉴클레오타이드 서열
60.	아미노산	Fv3C의 단백질 서열
61.	뉴클레오타이드	트리코테르마 리세이 유래의 GH3 β -글루코시다제, Tr3A의 뉴클레오타이드 서열
62.	아미노산	Tr3A의 단백질 서열
63.	뉴클레오타이드	트리코테르마 리세이 유래의 GH3 β -글루코시다제, Tr3B의 뉴클레오타이드 서열
64.	아미노산	Tr3B의 단백질 서열
65.	뉴클레오타이드	트리코테르마 리세이에서의 발현을 위해 코돈 최적화된, 탈라로마이세스 에베르소니이 유래의 GH3 β -글루코시다제, Te3A의 뉴클레오타이드 서열
66.	아미노산	Te3A의 단백질 서열
67.	뉴클레오타이드	아스페르길루스 니게르 유래의 GH3 β -글루코시다제, An3A의 뉴클레오타이드 서열
68.	아미노산	An3A의 단백질 서열
69.	뉴클레오타이드	푸사리움 옥시스포룸 유래의 GH3 β -글루코시다제, Fo3A의 뉴클레오타이드 서열
70.	아미노산	Fo3A의 단백질 서열
71.	뉴클레오타이드	지베렐라 제아에 유래의 GH3 β -글루코시다제, Gz3A의 뉴클레오타이드 서열
72.	아미노산	Gz3A의 단백질 서열
73.	뉴클레오타이드	넥트리아 해마토코카 유래의 GH3 β -글루코시다제, Nh3A의 뉴클레오타이드 서열
74.	아미노산	Nh3A의 단백질 서열
75.	뉴클레오타이드	버티실리움 달리아에 유래의 GH3 β -글루코시다제, Vd3A의 뉴클레오타이드 서열
76.	아미노산	Vd3A의 단백질 서열

도면1d

서열 번호	NT 또는 AA	명칭
77.	뉴클레오티드	포도스포라 안세리나 유래의 GH3 β -글루코시다제, Pa3G의 뉴클레오티드 서열
78.	아미노산	Pa3G의 단백질 서열
79.	아미노산	썬모도가 네아폴리타나 유래의 GH3 β -글루코시다제, Tn3B의 단백질 서열
80.	아미노산	포도스포라 안세리나 유래의 GH3 효소, Pa3C의 단백질 서열
81.	뉴클레오티드	포도스포라 안세리나 유래의 Pa3C의 뉴클레오티드 서열
82.	뉴클레오티드	Fv3C/트리코테르마 리세이 Bg13 융합/키메라의 뉴클레오티드 서열
83.	뉴클레오티드	Fv3C/Te3A/트리코테르마 리세이 Bg13 융합 키메라의 뉴클레오티드 서열
84.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 1
85.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 2
86.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 3
87.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 4
88.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 5
89.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 6
90.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 7
91.	아미노산	GH61 패밀리 엔도글루카나제의 단백질 서열 모티프 8
135.	아미노산	Fv3C/Te3A/Bg13 (FAB) 키메라의 단백질 서열
159.	아미노산	Fv3C/Bg13 (FB) 키메라의 단백질 서열
162.	뉴클레오티드	트리코테르마 리세이 Xyn2의 뉴클레오티드 서열
163.	뉴클레오티드	트리코테르마 리세이 Bx11의 뉴클레오티드 서열

도면2

효소	Tn3B	Fv3G	Fv3D	Tr3A	Pa3D	Te3A	An3A	Tr3B	Nh3A	Gz3A	Fv3C	Fo3A	Pa3G	Vd3A
기질														
상호작용	D58	D101	D111	D92	D87	D92	D92	D99	D106	D106	D119	D119	D101	D107
	R64	R107	R117	R98	R93	R98	R98	R105	R112	R112	R125	R125	R107	R113
	L116	L150	L160	L141	L136	L141	L141	L148	L155	L155	L168	L168	L150	L156
	R130	R165	R175	R156	R151	R156	R156	R163	R170	R170	R183	R183	R165	R171
	K163	K198	K208	K189	K184	K189	K189	K196	K203	K203	K216	K216	K198	K204
	H164	H199	H209	H190	H185	H190	H190	H197	H204	H204	H217	H217	K199	H205
	R174	R209	R219	R200	R195	R200	R200	R207	R214	R214	R227	R227	R209	R215
	M207	M237	M266	M232	M227	M242	M245	M252	M259	M259	M272	M272	M254	M260
	Y210	Y240	Y269	Y235	Y230	Y245	Y248	Y255	Y262	Y262	Y275	Y275	Y257	Y263
전핵체	D242	D272	D301	D267	D262	D277	D277	D287	D294	D294	D307	D307	D289	D295
기질														
상호작용	W243	W273	W302	W268	W263	W278	W278	W288	W295	W295	W308	W308	W290	W296
	S370	S455	S472	S415	S406	S447	S451	S457	S464	S464	S477	S477	S458	S465
산/염기	E458	E509	E534	E472	E463	E505	E509	E516	E523	E523	E536	E536	E517	E524

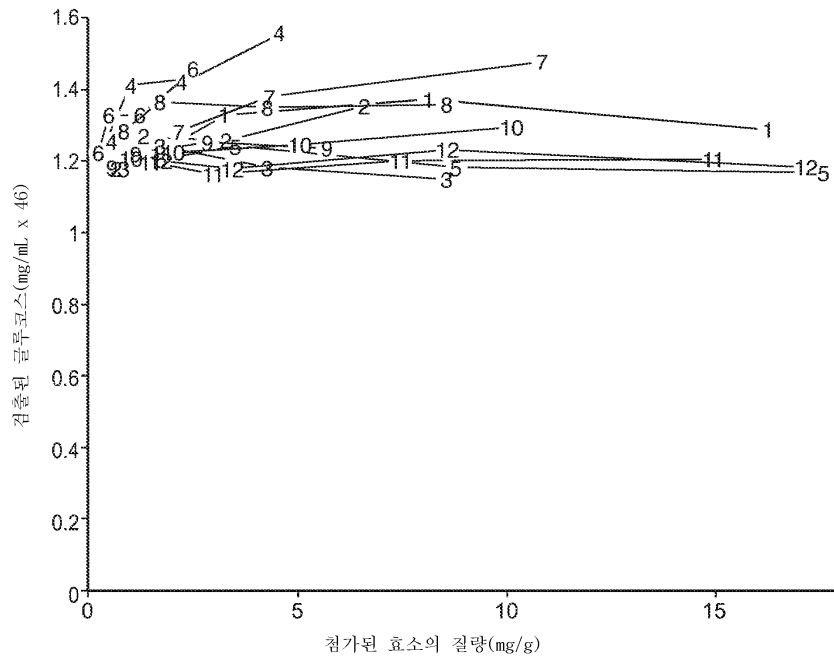
도면3

트리코테르마 리세이 통합 균주 H3A의 단백질 조성	
단백질	전체 영역의 %
Fv3A	9.6
Fv51A+Fv43D	14.8
Xyn 3	12.6
Bgl 1	7.5
CBH1	36.4
EGLs	5.6
CBH2	9.5
기타	4.0

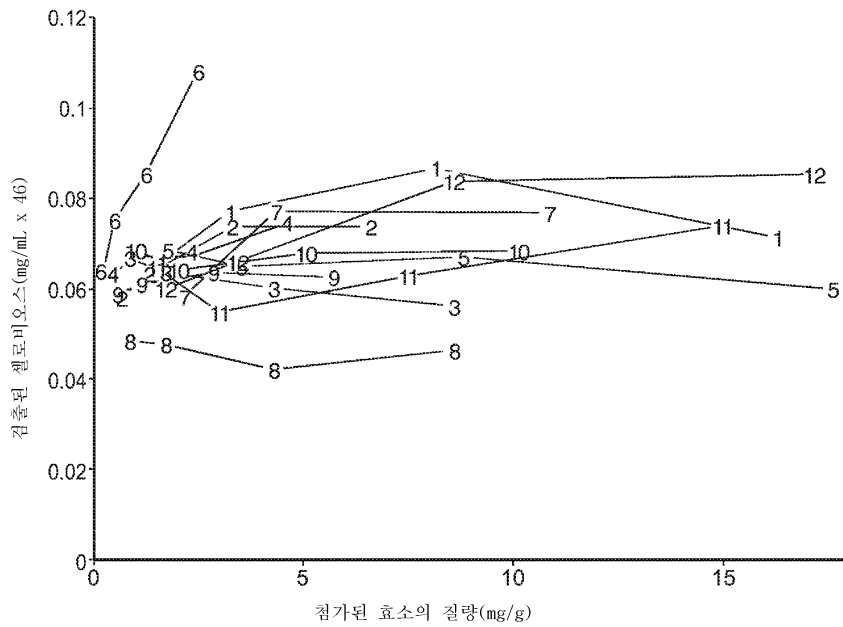
도면4a

트리코테르마 리세이 통합 균주 H3A에 첨가되는 단백질		
	단백질	스톡 단백질 농도 (mg/mL)
1	정제된 트리코테르마 리세이 CBH1	7.4
2	정제된 트리코테르마 리세이 CBH2	3.0
3	정제된 트리코테르마 리세이 EGL	3.9
4	정제되지 않은 Fv3C	2.1
5	물	
6	정제된 트리코테르마 리세이 EG4	1.1
7	H3A UF 농축물	102.8
8	정제된 트리코테르마 리세이 Bgl1	3.9
9	정제된 트리코테르마 리세이 Xyn2	2.6
10	정제된 트리코테르마 리세이 Xyn3	4.6
11	정제된 푸사리움 베르티실리오이데스 Fv43D	6.8
12	정제된 푸사리움 베르티실리오이데스 Fv51A	7.8

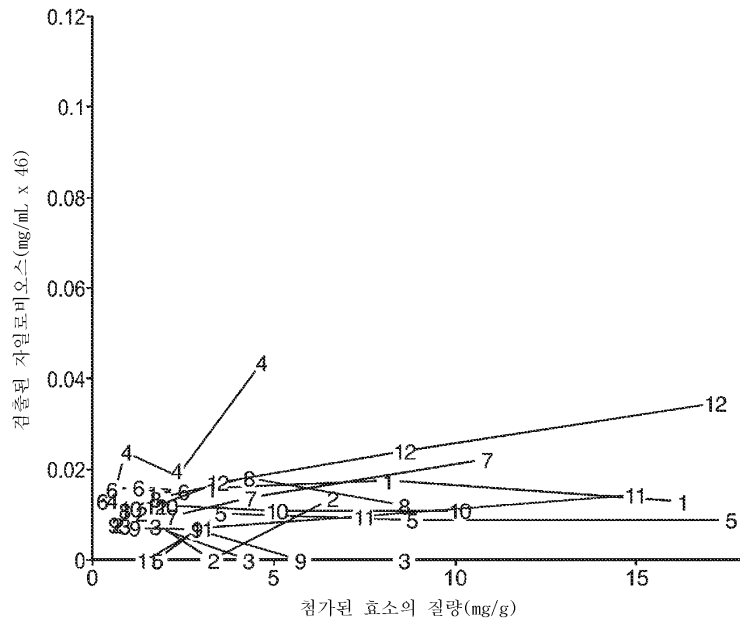
도면4b



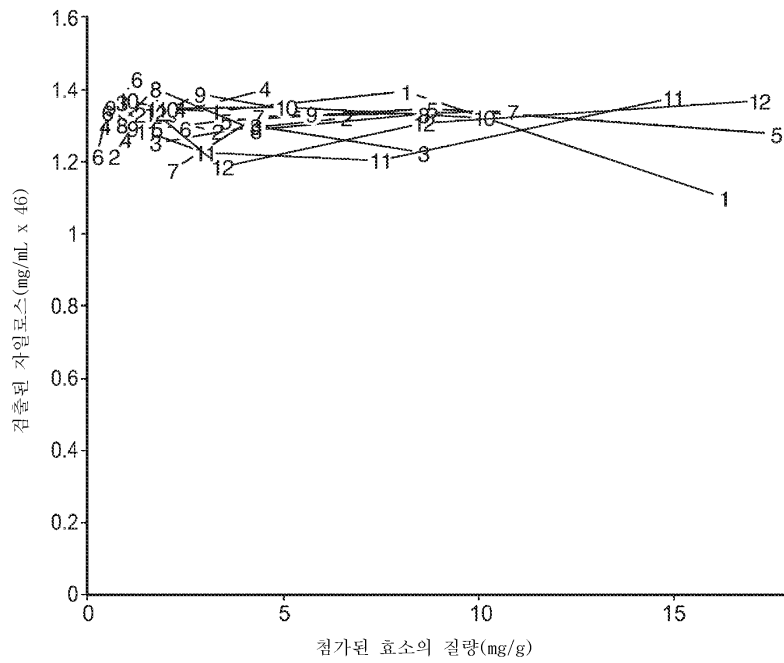
도면4c



도면4d



도면4e



도면5a

효소	농도 (mg/mL)	셀로비아제 활성		CNPg nM(CNP)/초/mg (단백질)
		U/ml	U/mg	
정제된 트리코테르마 리세이 Bglu1	2.3	19.4	8.4	1242
Fv3C 진탕 플라스크	2.4	42.7	18	1156
Fv3D 진탕 플라스크	2.9	0.0	0.0	6221
Pa3C 진탕 플라스크	1.9	0.0	0.0	2
정제된 아스페르길루스 니게르 Bglu1	2.4	244	102	168

도면5b

효소	트리코테르마 리세이 Bgl1에 대한 셀로비아제 활성	트리코테르마 리세이 Bgl1에 대한 CNPg 활성
정제된 트리코테르마 리세이 Bgl1	1	1
정제된 Fv3C	1.8	1
정제되지 않은 Fv3D	0	5
정제되지 않은 Pa3C	0	0
정제된 아스페르길루스 니게르 Bglu	12.1	0.1
정제된 FAB	1.3	1
정제된 FB	2.4	1.3
정제된 트리코테르마 리세이 Bgl3	5.6	1.5
정제된 Te3A	0.7-1.4	4.2-7.7

도면6

효소	w/w%
Xyn 3	45.0
Fv3A	15.0
Fv43D	5.0
Fv51A	35.0

도면7

효소 믹스	글루칸 전환을 위한 용량(mg(단백질)/ g(셀룰로스)):		
	70%	80%	90%
아셀레라제 1500 + 멀티펙트 자일라나제	38	-	-
트리코테르마 리세이 통합 균주(IRA) 유래의 전체 셀룰라제	21	28	36
75%의 트리코테르마 리세이 통합 균주(IRA) 유래의 전체 셀룰라제/25%의 Fv3C	12	15	19

도면8a

서열 번호 1

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 효소, Fv3A에 대한 뉴클레오티드 서열

atgctgctcaatcttcagggtcgctgccagcgctttgtcgctttctcttttaggtggattggctgaggetg
ctacgccatataaccttccggactgtaccaaaggacctttgagcaagaatggaatctgcgatacttcgtt
atctccagctaaaagagcggtgctctagtgtgctctgaagccgaagagaaggtgggcaatctggtc
aggtaaaatataccccccccataatcactattcggagattggagctgacttaacgcagcaatgcaactg
gtgcaccaagaatcggaacttccaaggtacaactgggtggaacgaagcccttcattggcctcgctggatctcc
aggtggctcgctttgcccgaactctcctacgcagcgccacatcatttcccatgcctcttctcatggcc
gctgctttcgacgatgatctgatccacgatatcggaacgtcgctggcaccgaagcgctgctgcttacta
acggcggttggcgcgagtcgacttctggacaccccaacgtcaaccccttttaaagatcctcgctggggtcg
tggctccgaaactccaggtgaagatgcccttcatgtcagcggtatgctcgctatatcgctcaggggtctc
gaaggcgataaagggaacacagctattgttgctacctgcaagcactatgctggaacgactttgaggact
ggggaggtctcagcgctcagactttgatgcaagattactcctcaggacttggctgagtactacgtcag
gcttttccaggagtgccacgtgatgcaaggttgggtccatcatgtgcgctacaatgcgctgaacggc
atcccgcatgcgcaactcgtatctgcaggagacgatcctcagagggcactggaactggaacgcgcgata
acaactggatgactaggtgtggcgccatgcaggatatctggcagaatcacaagtatgtcaagaccaa
cgctgaaggtgccaggtagcttttgagaacggcatggattctagctgcgagtatactactaccagcgat
gtctccgattcgtacagcaaggcctcttgactgagaagctcatggatcgttctggaagcgcttttcg
aagggtctgttcatactggtttctttgaacgggtgcaaaagcgcaatggaactcgctcagttttgaggatgt
caacaccaaggaggtcaggatcttgactcagatctgctgtggagggtgctgttcttctaagaatgac
ggcactttgctctgaagctcaagaagaaggatagtgtgcaatgatcggttctgggccaacgatactt
ccaagctgcaggggtgttacagtggaagctgctcgttctccacagcccgctttatgcagctgagaagct
tggctctgacaccaaactggtggtgggtcgacactgcagaacagctcatctcatgataactggaccacc
aagtgtgtgtgctgcggcgagaagctgtgattacattctcactttggtgtgttcttgacgcctctgctgctg
gcgaggacagagatcgtgagaaccttgactggcctgagagccagctgaccccttctcagaagctctctag
tctcggaagccactggttgttatccagcttggtgatcaagtcgatgacacgcctcttttgaagaacaag
aagattaacagctattctttgggtcaattaccctgggtcaggatggcgccactgcagtcattgacactgctca
ctggacgaaagagctcgtgctggcgactacccgtcacgcaatatccagtaaatatacactgagcagattgg
catgactgacatggacctcagacctaaccagctgttgccagggagaacttatcgctggtactcaactcca
gttcttccctacggctttggcctccactacaccaagttccaagcccaagttcaagtcacaagttgacgt
ttgacatccagaagcttctcaagggtgcagtgctcaatactccgatacttgcgcgctgcccccatcca
agttagtgtcaagaacacggcgccattacctcgcactttgtctctgtgttcttatcaagagtgaagtt
ggacctaaagccttaccctctcaagaccttgccggttatggctgcttgcatgatgtcgcgcttcatcga
cgaaggatatctcactggagtggaagcttgataacattgcgcgacggggagagaatggtgatttgggtgt
ttatctgggaacttacctctgtgtgctgagcctacgcaagccaagatccaggttacgctgactgga
aagaaggtctattttggataaagtgccctcaagacccaagctcgcgtaa

도면8b

서열 번호 2

Fv3A의 단백질 서열

mlnlqlqvaasalsisllgqlaeaatpytlpdctkgplskngicdtslspakraaalvaaltpeekvgnlv
snATGAPRIGLPRYNWNEALHGLAGSPGGRFADTPPYDAATSFPMPLLMAAAFDDDLIHDIGNVVGTEA
RAFTNGGWRGVDFWTPNVNPFKDPWRGRSETPGEDALHVSRYARYIVRGLEGDKQRRIVATCKHYAGN
DFEDWGGFTRHDFDAKITPQDLAEYYVRPFQECTRDARVGSIMCAYNAVNGIPACANSYLQETILRGHWN
WTRDNNWITSDCGAMQDIWQNHKYVKTNAEGAQVAFENGMDssceytttsdvsdsykqgllteklmdrsl
krlfeglvhtgffdgakagwnslsfadvntkeaqdlalrsavegAVLLKNDGTLPLKLLKKDSVAMIGFW
ANDTSKLQGGYSGRAPFLHSPLYAAEKLGLD TNVAVGPTLQNSSSHDNWNTNAVAAAKKSDYILYFGGLD
ASAAGEDRDRENLDWPESQLTLLQLSLGKPLVVIQLGDQVDDTALLKNKKINSILWVNYPGQDGGTAV
MDLLTGRKSPAGRLPVTOYPSKYTEQIGMTDMDLRPTKSLPGRTYRWYSTPVLPGFGFLHYTKfqakfks
nkltfdiqkllkqcsaqysdtcalppiqlsvskntgritsdfvslvfiksevgpkpykltlaaygrlhdv
apsttkdislewtldniarrgengdlvvygptytllldeptqakigvtltgkkailldkwpqdpksa

도면9a

서열 번호 3

페니실리움 푸니쿨로sum 유래의 GH43 패밀리 효소, Pf43A에 대한 뉴클레오티드 서열

atgcttcagcgatttgccttatatttaccactggetctattgagtggttgagtgaaagccgacacccct
ttgtgcagagcatctacaccgctgatccggcaccgatggtatacaatgaccgctttatgtcttcatgga
ccatgacacaccggagctacactacacatgacagactggcatctgtctcgtcagcagatatggcg
aattggcaagatcatggcattccaatgagcctggccaatttcacctgggccaacgcgaatgctggggcc
cgcaagtcacccctcgcaacggccaattctacttttatgctcctgtccgacacacagatggttctatggc
tatcggtgtgggagtgagcagcaccatcagaggtccataccatgatgctatcggcaaacccgctagtagag
aacaacgagattgatccaccggtgttcacgacgatgacggtcaggcatacctgtactgggaaatccag
acctgtggtacgtcaaatgaaaccaagatatgatatcgtacagcgggagccctactcagattccactcac
cacggctggatttggctactcgaacgggcaatgctcaacggccgaccacttttgaagaagctccatgggta
tacaacgcacacggcatctactatatcgctatgcagcggattgtgttctgaggatattcgctactcca
cggaacacagtgccactggtccgctggacttatcgaggcgtcatcatcgccacccaaggtagcagcttcac
caatcacgagggtattatcgacttccagaacaactcctactttttctatcacaacggcgctcttcccggc
ggaggcggtctaccaacgacgtctgtatgtgtggagcaattcaatacaatgacagtggaaccattccgacga
tcgaaatgaccaccgcgggtccagctcaaatgggactctcaacccctacgtgcgacaggaagcgaac
ggcgcatggtcttcaggcatcactacggaggtttgtagcgaaggcggaattgacgtcgggtttatcaac
aatggcgattacatcaaaagttaaaggcgtagctttcggttcaggagccattcttctcagcgcggttg
cttctgcaaatagcggcggaactattgcaatacacctcggaagcacaactggtacgctcgtgggcacttg
tactgtccccagcactggcggttggcagacttggaactaccgttacctgttctgtcagtgggcactctggg
accaggtatgtgattttgttttcggtggtagcggaacagagataacctgttcaactttgattattggcagt
tcgcataa

도면9b

서열 번호 4

Pf43A의 단백질 서열

mlqrfayilplallsqvkadnpfvqsiytadpampvndrvyvmfmdhntgatyyntdwhlfssadma
nwqdhgipmslanftwananawapviprngqfyfyapvrhndgsmaigvsvsstitgpyhdaigkplve
nneidptvfiddgqaylywgnpdlwyvklngdmisysgsptqipilttagfgtrtgnarpttfeeapwv
ykngiiyyiayaadccsedirystgtsatgpwtvrgvimptqgssftnhegiidfqnnsyfffyhngalpg
gggyqrsvcveqfkydnadgtiptiemttagpaqigtlnpyvrqEAETAWSSGITTEVCSEGGIDVGFIN
NGDYIKVKGVAFGSGAHSFARVASANSGGTIAIHLGSTTGLVGTCTVPSTGGWQTWTTVTCSVSGASG
TQDVYFVFGSGGTGYLFNFYDWFa

도면10a

서열 번호 5

푸사리움 베르티실리오이레스 유래의 GH43 패밀리 효소, Fv43E에 대한 뉴클레오티드 서열

atgaaggtatactggctcgtggcggtgggccaactctttgacgcggcactggctggcttgattggacacc
gtcgcggccaccacctcaacaatcctatcatctactcagactttccagataacgatgtattcctcggtcc
agataactactactacttctctgcttccaacttccacttcagcccaggagcaccgcttttgaaagtctaaa
gatctgctaaactgggatctcatcgccattcaattccccgctgaactttggcgacggctatgatcttc
ctcctggctcagcttattaccggtggaggtacttgggcacatccctcagatacagaaagagcaatggaca
gtggtactggtacggctgcactcaacttctggcagacactgggtatacaactgcctcatcgccggaaggtcca
tggtacaacaagggaacttcgggtgataacaattgctactacgacaatggcatactgatcgatgacgatg
ataccatgtatgtcgtatagcgttccggtgaggtcaaagtatctcaactatctcaggacggattcagcca
ggtcaaatctcaggtagttttcaagaacactgatattgggtccaagacttgagggttaaccgcatgtac
aagatcaacgggctctactatatcctaaacgatagcccaagtggcagtcagacctggatttggagtcga
atcaccctggggcccttatgagtcctcctcgcgacaaagtcaccccgctatctctggtggttaa
ctcgccgcatcagggttagtctcctcaaaagactcccaatggtggtggtactcattcattcacttgggc
tactcctgcgggcgctcttcgggttcttgcaacgattacgtggggtagcgatgggtttcccaattctgtca
agggtgctaattggcggtatggggtatcattacccaacacttctcggcagcgatggtgtgacaaagaattg
gacaaggactgataccttcggcggaacctcacttgcctcgtcgtggagtggaaccataatccggacgtc
aactccttcaactgtcaacaacggcctgactctcgcactgctagcattacgaaggatatttaccaggcga
ggaaacagctatctcaccgaactcattggtgatcctcaacaggaatagtgaaagattgatttctctccgat
gaaggacggcgaccgggcgggctttcagcgtttcgagaccaagtgacatacctcggtattcatcgagat
aacggaaggttcacaatcgctacgaagcatgggatgaatatggatgagtggaacggaacaacaacagacc
tgggacaaataaaagccacagctaatgtccttctggaaggaccaagatctggctgagacttcaacttga
taccaaccagcaggaaactggcaacactatctttcttacagttgggatggagtcaagatgaacactg
ggtcccaacttcaactgtacaatggttgggcattctttattgcttacggattcggcatcttcaacttcg
ccgagacggccttagaggctcgatcaagggttagtctttcacagctgcatag

도면10b

서열 번호 6
Fv43E의 단백질 서열

mkvywlvawatsltpalaglighrrattfnnpiiysdfpdndvflgpdnyyyfsasnfhfspgapvlksk
dlnwdlghsiprlnfgdgydlpggsryrrgtwasslryrksngqwywigcinfwqtwyvtasspegp
wynkgnfgdnncyydngilidddtmyvvygsgvkvslsqdgsqvkssqvfkntdigvqdlegrmy
kinglyyilndspsgsqtwiwwkskspwgyeskladkvtpisggnsphqgslktpnggwyfmsftwa
ypagrlpvlapitwgsdgfpilvkganggwssyptlpqtdgvtknwtrtdtfrgtslapswewnhnpdv
nsftvnngltlrltasitkdiygarnltlshrthgdhptgivkidfspmkgddraglsafdrqsayigihrd
ngkftiatkhhgmmdewngttdlqgikatanvpsgrtkiwlrlqltdnpgagtntifsyswdgvkyetl
gpnfklyngwaffiayrfgifnfaetalggsikvesftaa

도면11a

서열 번호 7
푸사리움 베르티실리오이레스 유래의 GH39 패밀리 효소, Fv39A에 대한 뉴클레오타이드 서열

atgcactacgcctaccctcaccacttttggtgctggtcttgaccaccaacgtcgctgcacagcaaggcacag
caactgtcgacctctccaaaaatcatggaccggcggaaggcccttggttcaggcttcataacggctggcc
tgacaacgggaacaagcgctcgacacctccataccagatttcttggttaactgacatcaaattcaactcaaac
cgcgcggttgccgcccataccatcactgggttgccgagaggtggtatgaaggatacctcgccgct
tcaactcaacttatccaactatcgccaccgcgcaagtataacgctgactttatcttgttgctcatga
ccctctgggggtcggtatggcgggcaggggttcaaaactcccggttctcgccgacaatggcaattggactgag
atggagttattctggaatcagcttgtgtctgacttgaaggctcataatatgctggaaggctcttgattg
atgtttggaatgagcctgatatgtatcttttgggacgcccgtggtcgagttcttgagtattacaa
tcgcgcgaccaaactacttcgggtgagtctactactgatccatacgtattttacagtgagctgactggcgga
attagaaaaacacttcccaaaactctctcagtgcccgccagccatggcacattctccattctgtccgatg
ataaatggcatacctggcttcaatcagtagcggttaacaagacagtcctctgatatttactcctggcatca
gattggcgcttggaacgtgagccggacagcactatccccgactttaccacottgcggcgcaatatggc
gttcccgagaagccaattgacgtcaatgagtagcgtgcacgcgatgagcaaaatccagccaactccgtct
actacctctctcaactagagcgctcataaccttagaggtcttcgcgcaaaactggggtagcggtctgacct
ccacaactggatgggcaacttgattacagcactaccggtacctcgagggggacttactaccctaattggt
gaatggcaggttacaagtactatggcgccatggcagggcagagacttgtgaccaaagcatcgctggact
tgaaagtttgatgtctttgcccactaagcaagggcgtaagattaaagattatagccggcagcaggaaccgttca
agcaaaagtataacatcaaaatcagcggttggaaagtagcaggacttccataagatgggtacggtaaaaggtc
cggaacttatcggttcgactgggctgggcgaatggaaaggttgacgggcctgttgatttgggggagaaga
agtatacttattcgcccaatcaggtgagcagccctctacttga

도면11b

서열 번호 8
Fv39A의 단백질 서열

mhyatlttlvlalttnvaagqgtatvdlsknhgpakalgsgfiygpndngtsvdtssipdfivtdikfn
rgggaqipslgwarggyegylgrfnstlsnyrttrkynadfillphdlwgadggggsnspfpdngnwte
melfwnqlvsdlkahnmlglvidwnepdidifwdrpwsqfleyynratkllrktlptkllsgpamahs
pilsddkwhwtlqsvagnktvpdiyswhqigawerepdstipdfttlraqygvpekpivneyaardeqn
pansvyyilsqlerhnlrglranwsgsdlnhwmgnliysttgtsegtyypngewqaykyyaamagqrlvt
kassdlkfdvfatkgqrkikilagttrtvqakynikisglevaglpkmgvtkvrttyrfdwagpngkvdgqv
dlgekkytysantvsspst

도면12a

서열 번호 9

푸사리움 베르티실리오이테스 유래의 GH43 패밀리 효소, Fv43A에 대한 뉴클레오타이드 서열

atgtggctgacctccccattgctgttgcgcagcaccctcctgggctcactggcggttgcctagcagaca
accccatcggtccaagacatctacacgcgagaccagcaccatgggtctacaatggcgcggtctacctctt
cacagggccatgacaacgcagcggtctaccgacttcaacatgacagactggcggtctctctcgtcagcagac
atggccaactggcagcaccatgggtgtcccatgagcttaaagaccttcagctgggccaacagcagagcct
gggctgggtcaagtcgttgcgcgaaacggaagttttactttctatgttccgtccgtaatgccaagacggg
tggaatggctattgggtgcgtgttagtaccacaacatccttgggcccactgatgccttggaaagcca
ttggctcgagaacaatgagatcgacccaactgtctacatcgacactgatggccagccctatctctactggg
gcaacctggattgtactacgtcaagctcaaccaagacatgctctcctacagtggtagcatcaacaaagt
atcgctcacaacagctggattcggcagccgcgcgaacaacgcgcagcgctcctactactttcgaggaagga
ccgtggctgtacaagcgtggaaatctctactacatgatctacgcagccaactgctgttccgaggacatto
gctactcaactggaccagcgccactggaccttggacttaccgcggtgctgctgatgaacaaggcgggtcg
aagcttcaccaaccatccttggcatcatcgactttgagaacaactcgacttcttttaccacaatggcgct
cttgatggaggttagcggttatactcgtgtgtggtgtcgagagcttcaagtatgggttcggacggtctga
tccccgagatcaagatgactacgcaaggccagcgcagctcaagtctctgaaccatattgtcaagcagga
ggccgagactatcgctcgtgtgagggatcgagactgaggtctgcagcgaagggtggtctcaacgttgct
ttcatcgacaatgggtgactacatcaaggtcaaggagtcgactttggcagcacccggtgcaaagacgttca
gcgcccgtgttgcctccaacagcagcggaggcaagattgagcttcgacttggtagcaagaccggttaagt
ggttggtaacctgcacggtaacgactacgggaaactggcagacttataagactgtggattgccccgtcagt
ggtgctactggtacgagcgatctattcttctcttccacgggctctgggtctggtctctgttcaacttca
actggtggcagtttagctaa

도면12b

서열 번호 10

Fv43A의 단백질 서열

mwltsp11fastllq1tgvaiadnpiqvdiytadpampvyngrvylftghdndgstdfnmtdwrlfssad
mvnwqhghgvpmslktfswansrawagqvvarngkfyfyvpvrnaktggmaigvgvstnilgpytdalgkp
lvenneidptvyidtdgqaylywgnpglyyvklnqdm1sysgsinkvslttagfgsrpnaqrpttfeeg
pwlykrnlyymiyaanccsedirystgpsatgpwtyrgvmmnkagrsftnhpgiidfennsyffyhnga
ldggsgytrsvavesfkygsdgl¹ipeikmttqgpaqlsINPYVKQEAETIAWSEGIETEVCSEGGLNVA
FIDNGDYIKVKGVDFGSTGAKTF SARVASNSSGGKIELRLGSKTGKLVGTCTVTTTGNWQTYKTVDCPVS
GATGTSDLFFVFTGSGSGSLFNFNWWQFs

도면13a

서열 번호 11

푸사리움 베르티실리오이테스 유래의 GH43 패밀리 효소, Fv43B에 대한 뉴클레오타이드 서열

atgogcttctcttggctattgtgcccccttctagcgatgggaagtgcctcttctgaaacgaagacggatg
tttgcacatacaaccaacctgtccttccaggatggcactcgatccatcggtatccagaaagatggcct
ctttctctgctcacttcaacattcatctccttcccagggtcttcccgctctatgctcaagggatctagtc
aactggcgtctcatcagccatgtctggaaccgcgagaaacagttgcctggcattagctggaagacggcag
gacagcaacagggaatgtatgcaccaaccattcgataccacaagggaacatactacgtcatctgcgaata
cctgggcggttgagatattattggtgtcatcttcaagaccaccaatccgtgggacgagagtagctggagt
gacctgtttacottcaagccaaatcacatcgaccccgatctgttctgggatgatgaaggaaagggtttatt
gtgctacccatggcatcactctgcaggagattgatttggaaactggagagcttagcccgagcttaatat
ctggaacggcacaggaggtgtatggcctgaggggtcccatatctacaagcgcgacgggtactactatctc
atgattgcccagggtggaactgcccgaagaccacgctatcacatcgctcgggcccgaagatcaccggcc
cctatgaagcctacaataacaaccaatcttgaccaaccgcgggacatctgagtaacttccagactgtcgg
tcaecggtgatctgttccaagataccaagggaactggtgggtctttgtcttgcactcgcatacacagca
cagggagtttcaaccatggcgctgaagctgtttgttcaatggcactggaacaagggcgaatggccca
agttgcaaccagtagcagggtgcctgcctggaacacctctccaaagccgacgcgaaacgttcccgga
tgggcccttcaacgctgacccagacaactacaacttgaagaagactaagaagatccctcctcatttgtg
caccatagagttccaaagagacggtgcctctctcttcttgccttccaaggtctgcacatcgctcctagtcaa
acaacgttaccggtagtgtgttgcaggagatgagattgagctatcaggacagcgagggtctagcttcat
cggaacgcgcgcaaaactcacactctgttcaaatatagttgtgatctgacttcaagcccaagtcgatgat
caggaagctggaatcacggttttccgcacgcagttcgaccatctgatcttggcattgttctgtcttcta
caaaccaaggcagcaacaagaatctaagcttgcttccgattccgggcccacaggagctcagaatgttcc
tgacccgaaggttagtaccggtcccgatggctgggagaaggcgtaatcagttcatatctgagggcagcc
aacgcgacgcactacaaccttgagcttccagccacagggcaagactctcgacatcgcgacagcatcag
caagttctgtgagtgagggaacgggttcatttgttggtagtttgcctggaccttatgtaacctgcaacgg
caaggtatctggagtggaatgtccaaggaggtgatgtctatgtgaaccaatggacttataagcccggtg
gcacaagagattgatcatggtgtttttgtgaaatcagaattgttag

도면13b

서열 번호 12

Fv43B의 단백질 서열

mrfswwllcp1lamgsalpetktdvstytnpvlpgwhsdpsciqkdglflcvststfifpglpvyasrdlv
nwrlishvwnrekqlpgiswktagqqgmyaptiryhkgtyyviceylgvvgdiigvifkttnpwdessws
dpvtfkpnhidpdlfwdddgkvycathgitlqeiddletgelspelniwngtggvwppegphiykrdgyyyl
miaeggtaedhaitiararkitgpyeaynnnpiltnrgtseyfqtvgghd1fqdtkgnwwg1clatrita
qgvspmgreav1fngtwnkgewpklqpvrgrmpgnllpkptrnvpdgdgpfnadpdnynlkktkkipphfv
hhrvprdgafslsskglhivpsrnnvtgsvlpqdeielsgqrglafigrqrqthtlfkysvddidfkpsdd
qeagitvfrtqfdhidlgivrlptnqgsnkksklafrfratgaqnvapkvvpvdpdgwekqvishieaa
nathynlgasshrgrktldiatasaslvsggtgsfvgsllgpyatcngksgvecpkggdvyy
tqwtypvqaeidhgvfvkse1

도면14a

서열 번호 13

포도스포라 안세리나 유래의 GH51 패밀리 효소, Pa51A에 대한 뉴클레오타이드 서열

atgatccacacctcaagccagccctcgccggtgtgtggcgctgtcgacgcaatgtgtggctattgatttgt
ttgtcaagctcttoggggggaataagacgactgatcatgtatgggtttatgcacgaggtatgtgttt
gcgagatctccctttgttttgcgcactgctgacatggagactgcaaacaggatatcaacaactccggc
gacggcggcatctacgcccagctaatctccaaccgcccgttccaaggagtgagaagtcccccaccaacc
tcgacaactggagcccgctcggtggcgctacccttacccttcagaagcttgccaagccctttctctgc
gttgcccttactccgtcaatgttgccaaccccaaggagggcaagggaagggaacccaagggaag
aagggttggtcttgcccaatgctgggttttgggtatggatgtcaagaggcagaagtacactggtagcttcc
acgttactgggtgagtacaagggtgactttgaggttagcttgcgcagcgcgattaccggggagaccttgg
caagaagggtggtgaagggtgggagtaagaagggaagtggaaccgagaaggagtttgagttggtgccttc
aaggatgcgcccacacgacaacaacaccttgtgtgacgtggatgcgaggtatgtgcttctttgat
tggtgagatagaagttgggttgacatgatgtgtgacgggcgcaaggacggatctttggatctcaact
tgatcagcttgttccctccgacattcaagggaagggaagaatgggtgagaattgatcttgcgcagacgat
ggttgagctcaagccggttaagctctcttagtcagaaaagtagagcctttgttaacgcttgacagacct
cttgccgttccccggtggcaacatgctcgagggttaacaccttggaacacttggtggaagtgttacgagacc
attggccctctgaaggatcgcccgggcatggctggtgtgtgggagtaccagcaaaccttggcttgggtc
tggtcgagtacatggagtgggccgatgacatgaacttgagcccgatgtgatccattttctggagtg
acttctcttgctaacgtatccacagttgtcggtgtcttcgctggtcttgccctcgatggctcgcttc
ccgaatccgagatgggagtggtgcatccaacaggtcttcgacgaaatcgagttccctcactggcgatgctaa
gaccaccaaatgggtgctcgctccgcccgaagcttgggtcaccccaagccttggaagggtcaagtgggttgag
atcggttaacgaggttggttgcgggacgccctgctggtctogagtcgtacatcaactaccgcttcccca
tgatgatgaaggccttcaacgaaaagtaacccgacatcaagatcatcgctcgccctccatcttcgacaa
catgacaatccccgggtgctgcccgtgacaccccgtaacctgaactccgatgagttcgttgagcga
ttcgccaagttcgataacttgagcaaggataacgtgacgctcatcgccgaggtgcgtcgacgcatccta
acggtggtatcgcttgggagggagatctcatgccttgccttgggtggggcgagtggtgctgaggtat
cttcttgatcagcactgagagaaaagggtgacaagatcatcggtgctacttacgcgcctggtcttcgcagc
ttggaccgctggcaatggagcatgaactgggtgcagcatgcgcgcgacccggccctcaccactcgctcga
ccagttggtatgtctggagaatcttcgcccaccacatcatccgtgagagcgtcccggtcgatgcccggc
cggaagcccaactttgacccctctgttctacgttgccggaaagagcgagagtggtgacccggtatctcaag
gctgcccgtctacaactcgactgaatcgatcccggtgtcggtgaagtttgatggtctcaacgagggagcgg
ttgccaacttgacggtgcttactggtgcccggaggtccgtatggatacaacgacccctcactggtatcaa
tggtgtcaaggagaagaccaccttcatcaaggccggaaggcgccgaagtcaaccttaccctgccgggc
ttgagtggtgctgtgttgagagcggccgacgcggtcaagggtggcaagggaagggaagggaagggaag
agggtaactga

도면14b

서열 번호 14

Pa51A의 단백질 서열

mihlkpalaallalstgcvaidlfrvkssggnkttidmyglmhedinnsgdgggiyaelisnrafqgsekfp
snldnswpvvgatitlqklakplssalpyvsnvanpkegkkgkdkdkgkkgvlganagfwgmdvkrqkytg
sfhvtgeykgdfevslrsaitgetfgkvvkkggskkgkwtekefelvpfkdapnsnntfvtqwdagakd
gsldlnlislfpptfkgkrknglridlaqtmvelkptflrfpggnmlegntldtwkwyetigplkdrpgm
agvweyqqtllglglveymewaddmnllepivgvfaglalldgsfvpesemgwiqgaldeiefltdakttk
wgavraklghpkpwkvkveignedwlagrpagfesyinyrfpmmmkafnekypdikiiaspsifdnmti
pagaagdhpyltpdefverfakfdnliskdnvtlgeaasthpnnggiawegdlmplpwggsvaeafli
sterngdkiiigatyapglrsldrwqswmtwvqhaadpalttrstswyvrilahhiiretlpvdapagkp
nfdplfyvagksesgtgikfaavynstesipvslkfdglnegavanltvltgpedpygndpftginvvk
ekttfikagkggkftftlpglsvavletadavkkgkkgkkgkgn

도면15a

서열 번호 15

지베렐라 제아에 유래의 GH43 패밀리 효소, Gz43A에 대한 뉴클레오티드 서열

atgaagtccaagttgttattcccactcctctctttcgttggtcaaagtcttggccaccaacgacgactgtc
ctctcatcactagtagatggactgcggatccttcggctcatgtctttaacgacaccttgggtctaccc
gtctcatgacatcgatgctggatttgagaatgatcctgatggaggccagtagcccatgagagattaccat
gtctactctatcgacaagatctacggttcctcgcggtcgatcaggtacggccctgtcagtgaggatg
tccctcgggctctcgcacagatgtgggctcctgacgctgccacaagaacggcaatactacctatactt
cctgccaaagacaaggatgatattcttcagaatcggcgttgctgtctaccaacccccggcgaccattc
gtccccgacaagagttggatccctcacactttcagcatcgaccccgccagtttctcgatgatgatgaca
gagcctacttggcatggggtggtatcatgggtggccagcttcaacgatggcaggataagaacaagtacaa

도면15b

서열 번호 16

Gz43A의 단백질 서열

mksklflfpllsfvqgslatnddeplitsrwtadpsahvfndtlwlypshdidagfendpdggqyamrdyh
vysidkiygsipvdhgtalsvedvpwasrqmwapdaahkngkyylyfpakdkddifrigvavspptggpf
vpdkswiphtfsidpasfvdddraylawggimggqlqrwqdknkynesgtepgngtaalspqiaklskd
mhtlaekprdmliildpktgkpllsededrffegpwihrnkiiyltystgtthylvyatsktpygypty
qgrilepvdgwtthssivkyqgqwlffyhdaktsgkdylrqvkakkiwydskgkiltkpk

도면16a

서열 번호 17

푸사리움 옥시스포룸 유래의 GH43 패밀리 효소, Fo43A에 대한 뉴클레오티드 서열

atgcagctcaagtttctgtcttcagcattgctgttctctctgaccagcaaatgcgctgcgcaagacacta
atgacattcctccctgatcacgacctctggctccgcagatccctcggctcatgtttcgaaggcaagct
ctgggtttaccatctcacgacatcgaagccaatgttgtaacggcacaggaggcgtcaatacgccatg
agggattaccatacctactccatgaagacatctatggtaaagatccggttgctgaccacggcgctcgtc
tctcagtcgatgacgttcctcggcgcaagcagcaaatgtgggctcctgacgcagctcataagaacggcaa
atattatctgtaottccccgccaaaggacaaggatgagatcttcagaattggagttgctgtctccaacaag
ccagcggctcctttcaaggccgacaagagctggatccctggcacgtacagtatcgatcctgctagctacg
tcgacactgataacgaggcctacctcatctggggcggtatctggggcgccagctccaagcctggcagga
taaaagaactttaacgagtcgtggattggagacaaggctgctcctaacggcaccatgcctatctcct
cagatcgccaagctaaagcaaggacatgcacaagatcacggaacaccccgcgactcgtcattctcgccc
ccgagacaggcaagcctcttcaggctgaggacaacaagcgaagattcttcgaggggccttggatccaaa
gcgcggcaagctttactacctcatgtactccacgggtgatacccacttcttctgctacgctacttccaag
aacatctacggctccttatacctaccggggcaagattcttgatcctgttgatgggtggactactcatggaa
gtattgttgagtataagggacagtggtggcttttcttctgtgatgcgcatacgtctggtaaggattacct
tcgacaggtgaaggcgaggaaagatctggtatgacaagaacggcaagatcttgcctcacccgtccttag

도면16b

서열 번호 18

Fo43A의 단백질 서열

mglkflssallfsitkskaagqndndipplitdlwsadpsahvfegklwvypshdieanvvnngtggagqyam
rdyhtysmkisiyqkdpvvdhgvalsvddvpwakqgmwapdaahkngkyylyfpakdkdeifrigvavsnk
psgpfkadmkipgtysidpasvdtneayliwggigggqlqawqdkknfneswigdkaapngtnalsp
qiaklskdmhkitetprdlvilapetgkplqaednkrrffegpwihrnklylmystgdthflvyatsk
niygyptyrgkildpvdgwtthgsiveykgqwwlffadahtsgkdylrqvkarkiydngkillhrp

도면17a

서열 번호 19

아스페르길루스 푸미가투스 유래의 GH43 패밀리 효소, Af43A에 대한 뉴클레오타이드 서열

```
atggcagctccaagtttatccctacccacaggtatccaatcgatataccaatcctctcttccctgggtggc
actccgatacccagctgtgcctacgtagcggagcaagacacctttttctgctgacgtccactttcattgc
cttccccgggtcttctcttttatgcaagccgagatctgcagaactggaaactggcaagcaatatttcaat
cgccccagccagatccctgatcttcgcgtcacggatggacagcagtcgggtatctatgcgccactctgc
gctatcatgagggccagttctacttgatcggttctgtaacctgggcccgcagactaagggcttgctgttcac
ctgctctgatccgtacgacgatgccgcgtggagcgatccgctcgaattcgcgggtacatggcatcgaccg
gatattcttctgggatcacgacgggacgggtctatgtcacgtccgcggaggaccagatgattaagcagtaca
cactcgatctgaagacggggcgattggcccgggttgactacctctggaacgggcacggaggagtctggcc
cgagggcccgacatttacaagagagacggatactactacctcatgatcgagagggagggtaccgagctc
ggccactcggagaccatggcgcatctagaacccggacaggtccctgggagccataccgcacaaatccgc
tcttgctgaacaagggcacctcggagtacttccagactgtgggcatgaggacttggtccaggatgggaa
cggaactgggtggccggtggcggttgagcaccgcgatcagggctgcaggaagaactatcccatgggtcgg
gagacggtgctgcgcccgccgcttgggagaagggtgagtgccctgtcattcagcctgtgagaggccaaa
tgcaggggcccgtttccaccaccaataagcgagttcctcgcggcgagggcggtatggatcaagcaaccga
caaagtggatttcaggccccgatcgaagataccggcgcaacttccagtactggcgatatcccaagacagag
gattttaccgtctccctcggggccaccgaataactcttccggtcacaccctcttttacaacctcaccg
gaactgcggaacttcaagccgatgatggcctgtcgttgttatgcgcaaacagacgcacacttggtcac
gtacactgtggacgtgtcttttgaccccaaggttgcgatgaagaggcggtgtgactgttttcttacc
cagcagcagcacatcgatcttggtattgtctcttccagacaaccgaggggctgtcgttgtcttccggt
tcgcgttggaaggcccggttaactacgaaggctctcttccagaagccaccgtgctgttcccaagggaatg
gtgtggacagaccatccggttgagattcaggccgtgagtgacacgagatgtcttttgcggctgccccg
gctcggcacccctgcacagaggcaaatcatcagccgcgccaactcgttgattgtcagtggtgatacgggac
ggtttactggctcgttgttggcggtgatgccacgtcgaacgggggtgcgggatccacgcccgcatatat
cagcagatggagatacgaaggacggggccagatgattgattttggtcagtggtcccgagctactga
```

도면17b

서열 번호 20

Af43A의 단백질 서열

```
maapslsyptgiqsytnplfpgwhsdpscayvaeqdtffcvststfiafppliyasrdlqnwklaasnifn
rpsqipdlrvtdgqqsgiyaptlryheggfyliivsylvpqtqglftssdpyddaawsdplefavhgidp
difwdhdgtvyvtsaedqmikqytlldlktgaigpvdylwngtggvwppegphiykrdgyyylmiaeggtel
ghsetmarsrtrtgpweypyhnpillsnkgtsyfqtvgadlfdqgngnwwavalstrsqpawknypmqr
etvlapaawekgewpviqpvrgmqmfpppnkrvrprgeggwikpdkvdfprgskipahfqywrypkte
dftvsprghpntlrlltspfyntltgtadfkpddglslvmrkqtdlftytvdvsfdpkvadeeagvtvflt
qqqhldlgivllqtteqlslsfrfrvegrgnyegplpeatvpvpkewcgqtirleiqavsdteyvfaaap
arhpaqrqiisranslivsgdtgrftgslvgvyatsnggagstpayisrwrlyegrqgmidfgrvvpsy
```


도면18a

서열 번호 21

페니실리움 푸니쿨로숨 유래의 GH51 패밀리 효소, Pf51A에 대한 뉴클레오티드 서열

atgggaaagatgtggcattcgatcttggttgtgtgggttatgtctgtcgggcatgccatcactatca
acgtgtcccaagtgccggcaataagaccagtccttttgaatatggctctgatgttcgaggtaatecttct
cttataccacataataaagtgtcgctcatttctaaagacaagtcaggacataaatcacggcggatgagcg
gtctgtatgcagagcttgttcgaaccgagcattccaaggtagcacgctctatccagcaaacctcgatgg
atacgaactcggtcaatggagcaatcctagcgttcagaatttgacaaacctctatcacctccatgcct
agctctctcaacgtcgcccaagggtccaacaatggaagcatcggtttcgcaaatgaaggctggtgggga
tagaagtcaagcgcgaagatacgcgggtcattctacgtccaggggactatcaaggagatttcgacat
ctctctcagtcgaaattgacacagaagtcttcgcaacggcaaaagtcaggctcctcgggcaaacacgag
gactgggttcaatacaagtagcaggttgggtgccaaaaaggcagcatcaaacaccaataaactctgacca
ttacttttgactcaaaaggtatgttaaattttgggtttagttcgatgtcttgcaattgtcttacgagaaac
gtagggtatgaaagacggtatccttgaaacttcaacttgatcagcctatttcccccaacttacacaatcgg
cccaatggcctaagaatcgacctggttgagctatggctgaactagaggggtaagctcttacaaatcaa
ctttatctttacgaagactaatgtgaaaacttagaaatttctcggtttccaggcggtagcgatgtggaa
ggtgtacaagctccttactggtataagtgaatgaaacggtaggagatctcaaggaccttatagtaggc
ccagtgcatggacgtacgaagaaagcaatggaattggcttgattgagtacatgaattggtgtgatgacat
ggggcttgagccgagtgagtgattccattcagcgtcaaattccagtggttctaatcatacacatcagttct
tgcgtatgggatggacattaccttcgaacgaagtgatatcggaacacgatttgcagccatatatcgac
gacacctcaaccaactggaattcctgatgggtgcccagatacgcataatggtagttggcgtgcgtctc
tgggtatccgaagcgtggacgattaaactacgtcgagattggaacgaagacaatctatacgggggact
agaaacatacatcgctacccggttccagcattacacgctataacagctaaatatccccatatgacg
gtcatggaatctttgacggagatgcttggtcggcgccgctgcaagcgattaccatcaatatctactc
ctgatgggtttgtttccagttcaactactttgatcagatgccagtcactaatagaacactgaacggtat
gaaaaccccccttttttaaatatgcttttaattggtatttaaccatctttcataggagagattgcaacct
ttatccaaataatcctagtaattcgggtggcctggggaagccattcccccttgatccttggtggattggg
tcggttcgagaagctgttttccataattggtgaagagaggaattcgccaaagataatcggtgctagctacg
tacggaattctacttttcgagattttaacattggataagaaggactaacctcaatacaggtccaatgtt
cagaaatatcaacaattggcagtggtctccaacactcatcgcttttgacgctgactcgtcggtacaagt
cgttcaacaagctgcatgtgatcaaggtatgctaattttctcctcattcaaacccgcagatgtgagct
aactttccgaagcttctctcgacaaacaaaatcacgcaaaatttaccacagacttgagtgccggtgaca
taggtccattatactgggtagctggacgaaacgacaatacaggatcgaacataattcaaggccgctgttta
caacagcacctcagacgtccctgtcaacgttcaatttgcaggatgcaacgcaagagcgcaaatttgacc
atcttgcctacgacgatccgaacgcacgaactacctggggggcccgaaagttgtgaagactgagatcc
agtctgtcactgcacaaatgctcatggagcatttgagttcagctctccgaacctaaagtgtggtgttctcaa
aacggagtaa

도면18b

서열 번호 22

Pf51A의 단백질 서열

mqkmwhsilvvlqllsvghaitinvsqsggnktsplqyglmfedinhggdgglyaelvrnrafqgstvyp
anldgydsvngailalqnltnplspmpsslnvakgsnngsigfanegwwgievkpqryagsfyvqgdyq
gdfdislqskltqevfatakvrssgkhedwvqkyelvpkkaasntnntltitfidskglkdgslnfnlis
lfpptynnrpnlridlveamaelegkflrfpggsdvegvpqpywykwnetvgdlkdrysrsawtyees
ngiglieymnwddmgilepilavwdghylsnevisendlqpyiddtlnqlleflmgapdtpygswraslgy
pkpwtinyveignednlyggletyiayrfqayydaitakyphmtvmesltempgpaaaasyhdyqystpdg
fvsqfnfydqmptvnrtnlgeiatvypnnpsnsvawgspfpplypwwigsvaeavfligeernspkiigas
yapmfrninnwqwsptliafdadssrtsrstswhviklstnkitqnlpttwsggdigplywvagrndnt
gsnifkaavynstsdvptvtvfagcnaksanltlssddpnasnypggpevvkteiqsvtanahgafefs
lpnlsvavlkte

도면19a

서열 번호 23

아스페르길루스 푸미가투스 유래의 GH11 패밀리 효소, AfuXyn2에 대한 뉴클레오타이드 서열

```
atggttttttttctactacgtgctggtggtgctccgcatgtggagctctggtgcccccgtagaaccg
agaccacgtcgttcaatgagactgctcttcatgagttcgctgagcgcgccggcaccaccaagctccaccg
ctggaacaaacggtactactactccttctggactgatggcgcgcgacgtgacctacaccaatggcgcc
ggtggctcgtactccgtcaactggaggaacgtgggcaactttgtcggtggaagggtggaacctggaa
gcgctaggtaccgagctttgtcaacgtcggtatgtgcagacctgtggctgacagaaagtagaacatcaact
acggaggcagcttcaacccacggcgaatggctacgtggtgctgctacggctggaccaccaaccccttgat
tgagtactacgttgttgagtcgtatggtacatacaaccccgcgagcggtaccttcaggggcactgtc
aacacggacgggtggcacttacaacatctacacggccgttcgctacaatgctccctccatcgaaggcacca
agaccttcaacccagtagtggctgtgctgcacctccaagcgtagccggcgactgtcaccatggccaacca
cttcaacgcctggagcagactgggcatgaacctgggaactcacaactaccagattgtcgccactgagggt
taccagagcagcggtatctgcttccatcactgtctactag
```

도면19b

서열 번호 24

AfuXyn2의 단백질 서열

```
mvsfsylllacsaiqalaapvepettsfnetalhefaeragtpsstgwnngyyysfwtgddvtytnga
ggsysvnrnrvnfvvgkgwnpgsartinyggsfnpsngylavvgwttnplieyyvvesygtynpgsgg
tfrgtvntdggtyinytavrnyapsiegtktftqywsrvrtskrtggtvtmanhfnawsrlgmnlgtthyq
ivategyqssgsasitvy
```

도면20a

서열 번호 25

아스페르길루스 푸미가투스 유래의 GH11 패밀리 효소, AfuXyn5에 대한 뉴클레오타이드 서열

```
atgatctccatttctcgtcagctttggactcgcgctatcgccggcgcatatgctcttccgagtgaca
aatccgtcagcttagcggaacgtcagacgatcacgaccagccagacaggcacaaacaatggctactacta
ttccttctggaccaacggtgcccggatcagtgcaatatacaaatgggtggtggcgcaatatagtgtgacg
tgggcgaaaccagaacggtggtgactttacctgtgggaagggtggaatccaggagtgaccagtaggcaa
cgcccgagaactatagaaggagcagcaaaagaaagcactaaactctctactagtgcattaccttctctgg
cagcttcaatccttccggaatgcttacctgtcgtgtatggatggactaccaacccctagtcgaatac
tacatcctcgagaactatggcagttacaatcctggtcggcgatgacgcacaagggcacccgtcaccagcg
atggatccacctacgacatctatgagcaccacaggtcaaccagccttcgatcgtcggcagcgccacctt
caaccaatactggtccatccgccaacaaagcgatccagcgccacagtcaccacccggaatcacttcaag
gcctgggctagtctgggatgaacctgggtaccataactatcagattgtttccactgagggatatgaga
gcagcggtacctcgaccatcactgtctcgtctggtggttcttcttctggtggaagtgggtggcagctcgtc
tactacttccctcaggcagctccctactggtggctccggcagtgtaagtcttcttccataggtgtgtggc
tttatgtgtattctgactgtgatagtgtctgttctgtggggccagtgcggtggaattggctggtctggt
cctacttgtgtctctcgggcaacttgccaggtttcgaaactcgtactactcaccagtgctttagtagtacctt
ttgcagggttatatccaagtga
```

도면20b

서열 번호 26

AfuXyn5의 단백질 서열

```
MISISSLSFGLAAIAGAYALPSDKSVSLAERQTITTSQTGTNNGYYYSFWTNGAGSVQYTNGAGGEYSVT
WANQNGGDFTCGKGWNPGRSDHDITFSGSFNPSGNAYLSVYGWTTNPLVEYYILENYGSYNPGSGMTHKGT
VTSDBGSTYDIYEHQQVNQPSIVGTATFNQYWSIRQNRSSGTVTTANHFKAWSLGMNLGTHNYQIVSTE
GYESSGSTTITVSSGSSSSGGSGSSSTSSGSSPTGGSGSCSALWQCQGGIGWSGPTCCSSSGTCQVSNS
YYSQCL
```


도면21a

서열 번호 27

푸사리움 베르티실리오이레스 유래의 GH43 패밀리 효소, Fv43D에 대한 뉴클레오티드 서열

atgcagctcaagtttctgtcttcagcattgttgcgtgtctttgacccggcaattgcgctgcgcaagacacta
atgatatccctcctctgatcacccgacctctggctctcggatccctcggctcatgttttcgagggcacaact
ctgggtttaccatctcagacatcgaagccaatgtcgtcaacggcaccggaggcgctcagtacgccatg
agagattatcacacctattccatgaagaccatctatgaaaagatcccggtatcgaccatggcgctcgtc
tgtcagtcgatgatgtcccatgggccaagcagcaaatgtgggctcctgacgcagcttacaagaacggcaa
atattatctctacttccccgccaaaggataaagatgagatcttcagaattggagttgctgtctccaacaag
ccagcgggtcctttcaaggccgacaagagctggatccccggtacttacagtatcgatcctgctagctatg
tcgacactaatggcgaggcacaacctcatctggggcggtatctggggcgccagcttcaggcctggcagga
tcacaagacctttaatgagtcgtggctcggcgacaaaagctgctcccaacggcaccacacgccttatctct
cagatcgccaaagctaagcaaggacatgcacaagatcacccgagacaccccgcatctcgtcatcctggccc
ccgagacaggcaagccccctcaagcagaggacaataagcgacgatttttcgaggggcccctgggttcacaa
gcgccgcaagctgtactacctcatgtactctacggcgacacgcacttctcgtctacgcgacttccaag
aacatctacggctccttatacctatcagggcaagattctcgacctgttgatgggtggactacgcatggaa
gtattgttgagtacaaggacagtggtggtgttctttcgcgatgcgcatacttctggaaaggattatct
gagacagggttaaggcgagggaagatctggtatgacaaggatggcaagattttgcttactcgtcctaagatt
tag

도면21b

서열 번호 28

Fv43D의 단백질 서열

mqlkflssalllsitgncaaqdndipplitdlwsadpsahvfegklwvypshdieanvvnvgtggaqyam
rdyhtysmkttiyqkdpvidhgvalsvddvpwakqmqmwapdaaykngkylylyfpakdkdeifrigvavnsk
psgpfkadtswipgtysidpasyvdtngaeayliwgggiwggqlqawqdhkftfneswlgdkaapngtnalsp
qiaklskdmhkitetprdlvilapetgkplqaednkrrffegpwwhkrkgllylmystgdthflvyatsk
niygpytyqgkildpvdgwtthgsiveykgqwwlffadahtsgkdylrqvkarkiwkydkdkilltrpki

도면22a

서열 번호 29

페니실리움 푸니쿨로숨 유래의 GH43 패밀리 효소, Pf43B에 대한 뉴클레오티드 서열

atgagtcgcagcatccttccgtacgcctctgttttccgctcctggggcgggctatgcgcgaaccgtttt
tggttctcaatagcgattttcccgatccagctctcatagagacatccagcggatactatgcattcggtac
cacccgaaacggagtcgaatgcgcaggttgcttcttccaccagactttaatacctggactttgctttccggc
acagatgccctcccggaacctttccgtcatggtagcttctgtctccacaaatctggcgccagatgttt
tggttaaggatgttcttatggaataacagtttttaggagtaggtcagccaggatattgacaaaattataa
taggcccagtggtacctatgtcatgtacttttcggcatctgctgcgagtgactcgggcaaacactgcgttg
gtgcgcgaactgcgcacctcaccggaaggaccttacaccccggtcgatagcgctgttgctgtccattaga
ccagggaggagctattgatgccaatggattttattgacaccgacggcactatatacgttgtatacaaaatt
gatggaacacagtctagacggtgatggaaccacacatcctacccccatcatgcttcaacaaatggaggcag
acgggaacaaccccaacccggcagcccaatccaactcattgacggatccgacctcgacggacctttgatcga
ggctcctagtttgctcctctccaatggaatctactacctcagtttctcttccaaactactacaacactaat
tactacgacacttccatacgcctatgcctcgtcgattacttggtccttggaacaaacaatctgcgccttatg
cacccttgttggttactggaaccgagactagcaatgacggcgcatgtgagcgccctggtggtgccgattt
ctccgtcgatggcacaagatgttggtccacgcaaacctcaatggacaagatatctcgggcggaacgcgcc
ttatttgcgtcgaattactgaggccaagcagatgtggttacattgcagtag

도면22b

서열 번호 30

Pf43B의 단백질 서열

mrsilpvasvfaliggaiaepflvlnsdfpdpслиetssgyyafgttgngvnaqvasspdfntwtllsg
tdalpgpfpsswvasspqiwadpvlvkadgtvmyfsasaasdsghkcvgaatatspegpytpvdsavacp
ldcggaidangfidtdgtiyyvykidgnsldgdgtthptpimlqqmeadgttptgspiglidrsdldgpl
ieapslllslngiyyisfssnnyntnyydsyayassitgppwtksapyaplvtgtetsndgalsapgga
dfsvdgtkmlfhanlngqdisggralfaaasiteasdvvtlq

도면23a

서열 번호 31

푸사리움 베르티실리오이레스 유래의 GH51 패밀리 효소, Fv51A에 대한 뉴클레오티드 서열

atgggttcgcttcagttcaatcctagcggctgcggcttgccttcgtggctgttgagtcagtcacatcaagg
tcgacagcaagggcgaaacgctactagcggtcaccaatatggcttccttcacgaggttggtattgacac
accactggcgatgattgggatgctaaacttgagctagatatcaacaattccgggtgatgggtggcatctac
gctgagctcatccgaatcgtgcttccagctacagcaagaaataccctgtttctctatctggctggagac
ccatcaacgatgctaaagctctccctcaaccgtctcgacactcctctctccgacgctctcccggtttccat
gaacgtgaagcctggaaggcgcaaggccaaggagattggtttccctcaacgaggggttactggggaatggat
gtcaagaagcaaaagtacactggctcttctctgggttaaggcgcttacaaggggcaactttacagctctct
tgcgatctaacccttacgacgatgtcttggcagcgtcaaggccaaggtccaagggccaacaagaagcagtg
gggtgagcatgagtttgcttactcctaacaagaatgccctaacagcaacaacacttttgctatcacc
tacgatcccaaggtgagtaacaatacaaaactgggacgtgatatactgacaatttgtagggcgctgatg
gagctcttgacttcaacctcattagcttggctccctccacctaacaagggcgcaagaacgggtcttcgagt
tgatcttgccgaggtctctgaaggtctccaccccgtaagggttaccgtctcacgtgtatcgtgaacagtc
gctgacttgtagaaaagagcctgctgcgcttcccggtggtaacatgctcgagggcaacaccaacaagac
ctgggtgggactggaaggataacctcgacacctctccgcaaccgtctggtttcgaggggtgctggaactac
cagcagacccatggctcttggaatcttgagtaacctccagtgggctgagggacatgaaccttgaaatcagta
ggttctataaaattcagtgacgggttatgtgcatgctaacagatttcagtttgctggtgcttacgctggcct
ctccctcgacggctccgtcaccaccaaggaaccaactccagccctcatcgacgacgctcgacgagatc
gaattcatccgaggtcccgctcacttcaaaagtgggaaagaagcgcgctgagctcggccaccccaagcctt
tcagactctctcactgtgaagtcgaaacgaggactggctcgctggttatcccaactggctggaactctta
caaggagtaccgctcccatgttctcagggctatcaagaagctcaccgcatctcaccgtcatctcc
tctggtgcttctattgacccggttggaagaagatgctggtttcgatatcctgctcctggaatcggtg
actaccaccccttacccgagcctgatgttctgttgaggagtccaacctgtttgataacaataagtatgg
tcacatcattgggtgaggttgcttctaccaccccaacgggtggaactggctggagtggtaaccttatgct
taccctgggtgagctctgtgttggtggcgagccgctcgctctctcggttatgagcgcaacgcgcatcgta
ttcccggaacattctacgctcctatcctcaagaacgagaaccgttggaactgggctatcaccatgatcca
attcgccgcccactccgcatgaccacccgctccaccagctggatgtctggtcactcttcgagggccac
cccatgaccatactctccaccaccgcccgaacttcgacccctctactacgtcgctggttaagaacgagg
acaagggaactcttatctggaagggtgctgcgtataacaccaccaagggtgctgacgttcccggtctct
gtcttcaagggtgtcaagccggtgctcaagctgagcttactcttgaccacaagaagagaaggtatcct
tttgcttcaatgatcctcacaagggaacaatgttgttgatactaagaagactgttctcaaggccgatg
gaaagggtgctttcaacttcaagcttctaacctgagcgtcgctgttcttgagacctcaagaaggga
gccttactctagctag

도면23b

서열 번호 32

Fv51A의 단백질 서열

mvrffssilaaaacfvavesvnikvdskggnatsghqygflhedinnsgdggiiyaelirnrafqyskky
slsgwrpindaklslnridtplsdaipvsmnvkpgkgkakeigflnegywgmdvkkqkytgsfwvk
ghftaslrslntddvfgsvkvskankkqwehefvltpnknapsnntfaitydpkgadgaldfnlisl
fpptykgrkngrlrvdiaealeghpsllrfpggnmlegntnktwwdwdtlgplnrpgefegvwnyqgth
glgileylqwaedmnleiiivgvayglsldgsvtpkdqlplldaldeiefirgvpvtskwgkkr
kpfrlsyvevgnedwlagypgtwnsykeyrfpmfleaikkahpdlvissgasidpvgkddagfdip
igdyhpyrepdvveefnlfdnnkyghii**gevasthpnggtgwsnlpypwwisgvgeavalcgyerna**
dripgtftyapilknenrwwaitmiqfaadsamttrstswyvwslfaghpmthtlpttadfdplyy
vagk**nedkgtliwkgaa**ynttkgadvpslsfkgvkpqaqaeltltnkekdpfafndphkggnvvdtkk
tvlk**adgkga**fnfklpnlsvavletlkkgkpyss

도면24a

서열 번호 41

트리코테르마 리세이 유래의 GH10 자일라나제, Xyn3에 대한 뉴클레오티드 서열

atgaaagcaaacgtcatcttgtgcctcctggccccctggctgcgcctctccccaccgaaaccatccacc
tcgaccccgagctcgccgctctccgcgccaacctcaccgagcgaacagccgacctctgggaccgccaagc
ctctcaagcatcgaccagctcatcaagagaaaaggcaagctctactttggcaccgccaccgaccgaggc
ctctccaacgggaaaagaacgcggccatcatccaggcagacctcggccaggtagcgcgggagaacagca
tgaaatggcagtcgctcgagaacaaccaaggccagctgaactggggagacgccgactatctcgtaactt
tgcccagcaaaacggcaagtcgatacgcgcccaactctgatctggcaactcgagctgctgctgggtg
aacaatatcaacaacgcggatactctgcggcaagtcatccgcacccatgtctctactgtgttggcggt
acaaggcgcaagattcgtgcttgggtgagttttgaacaccacatgcccttttcttagtcgctctctctc
ctcttggaaacttctcacagttatagccgtatacaaacattcgacaggaaatttaggatgacaactactgac
tgacttgtgtgtgtgatggcgataggacgtgtcaatgaaatcttcaacgaggatggaacgctgcgctct
tcagttctttccaggctcctcgccgagaggtttgtctcgattgcctttcgtgctgctcgagatgctgacc
ctctcgccgtctttacatcaacgactacaatctcgaccgcaccaactatggcaaggtaacgggttgaa
gacttacgtctccaagtgagatctctcaaggagttccattgacggtattgggtgagccacgacccctaaat
gtccccattagagttctcttctagagccaaggcttgaagccattcagggaactgacacgagagccttctc
tacagggaagccagtcctcatctcagcggcgccgaggtctctggtacgctgggtgcgctccagcagctggca
acggtacccgtcaccgagctggccattaccgagctggacattcaggggcgaccgacgacgattacaccc
aagttgttcaagcatgctcgagctctccaagtgcgtcggtacacgctgtggggcatcagtgacaaggt
aagttgcttccctgtctgtgcttatcaactgtaagcagcaacaactgatgctgtctgtctttacctagg
actcgtggcgtgccagcaccacccctcttctgtttgacgcaaaacttcaaccccaagccggcatataacag
cattgttggcatcttacaatag

도면24b

서열 번호 42

Xyn3에 대한 단백질 서열

mkanvilcllplvlaalptetihldpelaalranltertadlwdrqasqsidqlikrkgklyfgtatdrg
llqreknaiiqadlqgvtpensmkwqlennqgqlnwgdadylvnfaqngksirghtliwhsqplpawv
nninnadtlrqvirthvstvvgrykgkirawdvvneifnedgtlrssvfrllgeefvsiafraardadp
sarlyindynldranygvnqlktyvskwisqgvpidgigsqshlsqggsgtlgalgqlatvpvtelai
teldiqgapttdtyqvvaqlsvskcvgitvvgisdkdswrastnpllfdfanfnpkpaynsivgilq

도면25a

서열 번호 43

트리코테르마 리세이 유래의 GH11 패밀리 자일라나제, Xyn2의 단백질 서열

mvsftsl1laasppsrascrpaaevesvavekrqtiqpgtgynngyfysywndghggvtytngpggqfsvn
wsnsgnfvggkgwpgptknkvinfsgsynpngnsylsvygsrnplicyiyivenfgtynpstgatklev
tsdgsvydiyrtrvngpsliigtatfyqywsrrnrhssgsvntanhfnawaqgltlgtmdyqivaveg
yfsngsasitvs

도면25b

서열 번호 162

트리코테르마 리세이 유래의 Xyn2의 뉴클레오티드 서열

ATGGTCTCCTTCACCTCCCTCCTCGCCGGCGTCGCGGCCATCTCGGGCGTCTTGGCCGCTCCCGCCGCGG
AGGTCGAATCCGTGGCTGTGGAGAAGCGCCAGACGATTCAGCCCGGCACGGCTACAACAACGGCTACTT
CTACTCGTACTGGAACGATGGCCAGCGCGCGTGACGTACACCAATGGTCCCGCGCGGCGAGTTCTCGGC
AACTGGTCCAACCTCGGCAACTTTGTGCGCGGCAAGGGATGGCAGCCCGGGACCAAGAACAAGTAAGACT
ACCTACTCTTACCCCTTTGACCAACACAGCACAAACAATACAACACATGTGACTACCAATCATGGAAT
CGGATCTAACAGCTGTGTTTTAAAAAAGGGTCATCAACTTCTCGGGAAGCTACAACCCCAACGGCAAC
AGCTACCTCTCCGTGTACGGCTGGTCCCGCAACCCCTGATCGAGTACTACATCGTCGAGAACTTTGGCA
CCTACAACCCGTCACGGGCGCCACCAAGCTGGCGGAGGTACCTCCGACGGCAGCGTCTACGACATTTA
CCGCACGCAGCGCGTCAACACGCCGTCCATCATCGGCACCGCCACCTTTTACCAGTACTGGTCCGTCCGC
CGCAACCAACCGCTCGAGCGCTCCGTCAACACGGCGAACCCTTCAACCGCTGGGCTCAGCAAGGCCTGA
CGCTCGGACGATGGATTACCAGATTGTTGCCGTGGAGGGTTACTTTAGCTCTGGCTCTGCTTCCATCAC
CGTCAGCTAA

도면26a

서열 번호 44

트리코테르마 리세이 유래의 GH3 패밀리 β -자일로시다제, Bx11의 단백질 서열

mvnnaallaalsailptalaqnnqnyanyasagqpdlypetlatitlsfpdcehgplknnlvcdssagyv
eraqalislfttleiilntqns~~sgpgvpriglpnyqv~~nealhgldranfatkggqfewatsfmpiltta
alnrtlihqadiistqarafsnsgrylgldvypnvngfrs~~plwgrgqetpgeda~~fflssaytyeyitgi
qggvdpehlkvaatvkhfagydlenwnnqsr~~lgfdaiitq~~dlseyytpqflaaaryaksrslmcaynsv
ngvpscansfflqtlilreswgfpewgyvssdc~~davynvfnphdy~~asnqssaasslragtdidogqtypw
hlnesfvagevsrgeiersvtrlyanlvrlgyfdkknqyrs~~lgwkdvvkt~~dawnisyeaavegivilknd
gtiplskkvr~~sialigp~~wanattqmqgnyygpapylispleaakkagyhvnfelgteiagnsttgfakai
aaakksdaiiylggidntiegegadrtdiawp~~gnql~~dlikqlsevgkplvvlqmgggqvds~~slksnk~~kv
nslvwggyppgsggvalfdilsgkrapagr~~lvttqypaeyvhqf~~pqndmnlrpdgksnpgqtiwytgkp
vyefgslfyttfketlashpkslkfntssilsaphgytyseqipvftfeaniknsgktespytamlfv
rtsnagpapypnkwlv~~gfdrladikpgh~~ssklisipipvsalarvdshgnrivypgkyela~~lntdes~~vkile
felvg~~eevtienwpleeq~~ikdatpda

도면26b

서열 번호 163

트리코테르마 리세이 유래의 Bx11의 뉴클레오타이드 서열

atgggtgaataacgcagctcttctcgcgcgcctgtcggtctctctgcccacggcctggcgacagaacaatc
aaacatacgcacaactactctgtcagggccagcctgatctetaccccgagacacttgcacgcctcacact
ctcggtccccgaactgcgaacatggccccctcaagaacaatctcgtctgtgactcatcgcccggtatgta
gagcgagccccagggcctcatctcgtctctccacctcgaggagctcattctcaacacgcgaactcgggcc
ccggcgtgcctcgcctgggtcttccgaactaccaagtctggaatgaggtctctgcacggcttggaaccgcgc
caactctgcgccacaagggcgccagcttccaatggcgacactcggtccccatgccatcctcactacggcg
gcctcaacccgcacattgatccaccagattgcgcacatcatctcgacccaagctcgagcattcagcaaca
gcgcccggttacggtctcgcagctctatgcgccaaacgtcaatggcttccgaagccccctctggggcggtgg
ccagggagacgcgcgcggaagacgccttttccctcagctcgcctatacttacgagtacatcacgggcatc
caggggtggcgctcgaccctgagcacctcaaggttgcgccacgggtgaagcactttgcgggatacgacctcg
agaactggaaacaaccagctcccgctctcggtttcgacgccatcataactcagcaggacctctccgaatacta
cactccccagttcctcgtctcgggcccggttatgcaaatgcacgcagcttgatgtgcgcatacaactccgtc
aacggcggtgccagctgtgccaacagctctctctcgcagacgcttttgcgcgagagctggggcttccccg
aatggggatacgtctcgtccgattgcgatgcgcgtctacaacgttttcaacctcatgactacgccagcaa
ccagctcgtcagccgcgcgcagctcactgcgagccggcaccgatatcgactcggtcagacttaccctggg
cacctcaacgagctccttgtggccggcggaagtctcccgccggcgagatcgagcggctccgtcacccgtctgt
acgccaaacctcgtccgtctcggatacttcgacaagaagaaccagttacgcgtcgtcgggtggaaggatgt
cgtcaagactgatgcctggaaacatctcgtacgaggtcgtctgtgagggtcagctcctcgtcagaagacgat
ggcactctccctctgtccaaagaaggtgcgcagcattgctctgatcggaacctggggccaatgccacaaccc
aaatgcaaggcaactactatggccctgccccatacctcatcagccctctggaagctgctaagaaggccgg
ctatcacgtcaactttgaactcggcacagagatcgccggcaacagcaccactggctttgccaaggccatt
gctgcccgaagaagtcggatgccatcatctacctcgggtggaattgacaacaccattgaacaggaggcg
ctgaccgcacggacattgcttggcccggtaatcagctggatctcatcaagcagctcagcgaggtcggcaa
accccttgtcgtcctgcgaatggcggtggtcaggtagactcatcctcgtcgaagagcaacaagaaggtc
aactccctcgtctggggcgatataccggccagtcgggagggcgttgcctcttcgacattctctctgga
agcgtgctcctgcgcgcgcagctggtcaccactcagtaaccggctgagtatgttcccaattccccagaa
tgacatgaacctccgaccgatggaaagtcaaacctggacagacttacatctggtacacccggcaaccc
gtctacgagatttggcagtggtctctctacaccacttcaaggagactctcgccagccaccccaagagcc
tcaagttcaacacctcatcgatcctctctgctcctcaccccggtatacacttacagcgagcagattccgt
cttcaccttcgaggccaacatcaagaactcgggcaagaaggagtcctccatatacggccatgctgtttgtt
cgacaagcaacgctggccagccccgtacccgaacaagtggctcgtcgtcggattcgaccgacttgccgaca
tcaagcctggctcactcttcaagctcagcatccccatccctgctcagtgctctcgcgcgtgttgattctca
cggaacccgattgtataccccggcaagtatgagctagccttgaacaccgacgagctctgtgaagcttgag
tttgagttggtgggagaagaaggttaacgattgagaactggcgttggaggagcaacagatcaaggatgcta
cacctgacgcataa

도면27c

서열 번호 47

포도스포라 안세리나 유래의 GH51 패밀리 효소, Pa51A에 대한 코돈 최적화된 cDNA

atgatccacctcaagcccgccctcgccgcccctcctcgccctcagcaccacatgctgcccacgcacctct
tcgtcaagagcagcgccggaacacagaccacccgacatcatgtacggcctcatgcaagagacatcaacaa
cagcgccgacggcgccatctacgcccagctgatcagcaaccgcccctccagggcagcgagaagtcccc
agcaacctcgacaactgggtcccccgctcgccggcgccacccctcaccctccagaagctcgccaagccctgt
cctotgcccctcccactcctcgtaacgtcgccaaccccaaggagggttaagggttaagggtcaaggacaccaa
gggcaagaaggtcgccctcgccaacgcccggcttttggggcatggacgtcaagcgccagaaatcacccggc
agcttccacgtcaccggcgaggtacaagggcgacttcgaggtcagcctccgcagcgccattaccggcgaga
ccttcggcaagaaggtcgtaagggcgccgagcaagaagggaagtggaacggagaaggagttcgagctggt
ccccctcaaggacgcccccaacagcaacaacacccctcgctcgctccagtgggacgcccagggcgccaaggac
ggcagcctcgacctcaacctcatcagcctcttcccggcccaccttcaaggggccgcaagaacggcctccgca
tcgacctcgcccagaccatggtcgagctgaagcccacccctcctccgctttcccggcgcaacatgctcga
gggcaacacccctcgacacctggtggaagtgttacgagaccatcgcccccctgaaggacccgcccgtgcatg
gcccggctctgggagttaccagcagacgctgggcccctcgccctgggtcgagttacatggagtgggcccagcaga
tgaacctcgagccccatcgctggcgctcttctgctggcctggcctggatggcagctttgtccccgagagcga
gatgggctgggtcatccagcaggtctctcgatgagatcgagttcctcacccggcgacgccaagaccaccaag
tggggcgccgctccgcccgaagctcgccacccctaaagccctggaagggtcaaatgggtcgagatcggaacg
aggactggctcgccggcgccgacctcgccgcttcgagagctacatcaactaccgcttcccatagtgatgaa
ggccttcaacgagaaataccccgacatcaagatcatgtccagcccctccatcttcgacaacatgaccatt
caagccggtgctgcccgtgacccaccacccctacctcacccccgaagaaattgtcgagcgcttcgccaagt
tcgacaacctcagcaaggacaaacgtcacccctcatggcgaggccgcccagcaccaccccccaacggcgcat
tgcttgggagggcgacctcatgcccctgcccctgggtggggcgccagcgtcgccgaggccatcttctctatc
agcaccgagcgaacggcgacaaagatcatcgccgcccacccacgcccctggcctccgatctctcgaccgct
ggcagtgagcatgacctgggtccagcagcccgccgacccctgcccctcaccaccccgcagcaccagctggta
cgtctggcgcatctcgccaccacatcatctcgagagacctcccgtcgacgcccccgccggcaagccc
aacttcgacccccctcttctacgtcgctggcaagtcggagagcgccgacccgcatcttcaaggccgcccgtct
acaacagcaccgagagcatccccgtcagcctcaagttcgacggcctcaacgagggcgccgtcgccaacct
cacccgtcctcacccggccccgaggaccctacggctacaacgaccccttcccccgcacacacgtcgtaag
gaaaagaccaccttcatcaaggccggcaaggcggaagttcacctttacccctccccggcctctctgtcg
ccgtcctcgagacggccgacgcccgtgaagggtggcaagggaagggaagggaagggttaagggttaacta
a

도면27d

서열 번호 48

지베렐라 제아에 유래의 GH43 패밀리 효소, Gz43A에 대한 뉴클레오타이드 서열

atgtatcggaaagtggccgtcatctcgccctcttggccacagctcgtgctaccaacgacgactgtctc
tcatcactagtagatggactcgccgacccctcggtcatgtctttaacgacacctgtggctctacccgtc
tcatgacatcgatgctggatttgagaatgatcctgatggaggccagtaacccatgagagattaccatgtc
tactctatcgacaagatctacggttccctgcccgtcgatcacggtacggccctgtcagtgaggatgtcc
cctgggcccctcgacagatgtgggctcctgacgctgcccacaagaacggcaataactacctatacttccc
tgccaaagacaaggatgatattcagaatcgccgttgctgtctaccaacccccggcgaccattcgtc
cccgacaagagttggatccctcacactttcagcatcgaccccgccagtttctgctgatgatgatgacagag
cctacttggcatgggtgggtatcatgggtggccagcttcaacgatggcaggataagaacaagttacaacga
atctggcactgagccaggaaacggcacccgctgcccgtgagccctcagattgccaagctgagcaaggacatg
cacactctggcagagaagcctcgccacatgctcattcttgaccccaagactggcaagccgctcctttctg
aggatgaagaccgacgcttcttcgaaggaccctggattcacaagcgcaacaagatttactacctcaccta
ctctactggcacacccactatcttctgtatgagacttcaaagacccccctatggtccttacacctaccag
ggcagaattctggagccagttgatggctggactactcactctagtatcgtcaagtaaccagggtcagtggt
ggctattttatcacgatgccaagacatctggcaaggactatcttcgcccaggtaaaggctaagaagatttg
gtacgatagcaaaaggaaagatcttgacaaagaagccttga

도면27e

서열 번호 49

푸사리움 옥시스포룸 유래의 GH43 패밀리 효소, Fo43A에 대한 뉴클레오타이드 서열

atgtatcgggaagtggcggtcatctcgcccttcttggccacagctcgtgctcaagacactaatgacattc
ctccctgatcaccgacctctggtcgcgagatccctcggctcatgttttcgaaggcaagctctgggttta
cccatctcagcatcgaagccaatgttgtaacggcacaggaggcgtcatacggcatgagggttac
catacctactccatgaagagcatctatggtaaagatcccggttgtagaccacggcgctcgtctctcagtcg
atgacgttccctggcggaagcagcaaatgtgggtccttgacgcagctcataagaacggcaaatattatct
gtacttccccgccaaaggacaaggatgagatcttcagaattggagttgctgtctccaacaagcccagcggt
cctttcaaggccgacaaagagctggatccctggcacgtacagtatcgatcctgctagctacgtcgacactg
ataacgaggcctacctcatctgggcggtatctgggcgggccagctccaagcctggcaggataaaaagaa
ctttaacgagtcgtggattggagacaaggctgctcctaacggcaccaatgccctatctcctcagatcgcc
aagctaagcaaggacatgcacaagatcacgaaacaccccgcatctcgtcattctcgcccccgagacag
gcaagcctcttcaggctgaggacacaagcgacgattcttcgaggcccttggatccacaagcgcgga
gctttactacctcatgtactccaccggtgataccacttcttctacgtacttccaagaacatctac
ggtccttatacctacccgggcaagattcttgatcctgttgatgggtggactactcatggaagtattgttg
agtataagggacagtggtggtcttcttctgtgatgcgcatacgtctggttaaggattacctcgacaggt
gaaggcgaggaagatctggtatgacaagaacggcaagatcttgcctcaccgtccttag

도면27f

서열 번호 50

페니실리움 푸니쿨로숨 유래의 GH51 패밀리 효소, Pf51A에 대한 뉴클레오타이드 서열

atgtaccgggaagctcgccgtgatcagcgcccttcttggcgactgctcgccgcatcaccatcaacgtcagcc
agagcgcgccgcaacaagaccagcccgctccagtaacggcctcatgttcgaggacatcaaccaaggcgccga
cggcgccctctacgcgcagctgggtcggaaacgggccttcagggcagcaccgtctacccggccaacctc
gacggctacgactcgtgaaacggcgcgatctctcgctccagaaacctcaaccaaccgctcagcccgagca
tgccctcgtcgtgaaacgtcgccaagggtcgaaacaacggcagcactcggttcgccaacgaggggtggtg
ggcatcgaggtcaagcgcgacggtaacgcggcagcttctacgtccaggcgactaccagggcgacttc
gacatcagcctccagagcaagctcaccagagggtcttcgcgacggcggaaggtccggtcgagcggaagc
acgaggactgggtccagtaacagtaacgagctggtcccgaagaaggcccgagcaaaccaacaacacct
caccatcaccttcgacagcaagggcctcaaggacggcagcctcaactcaacctcatcagcctcttcccg
ccgacctacaacaaccggccgaacggcctccggatcgacctcgtcgaggccatggcgagctggagggca
agttcctcgcgttccccggcggtcggacgtggaggcggtccaggcccgtaactggtaacagtggaaacga
gaccgtcgcgacactcaaggacgcctactcgcccgagcgctggacctacgaggagagcaacggcatc
ggcctcatcgagtacatgaactggtcgacgacatgggcctcgagccgactcctcgccgtctgggacggcc
actacctcagcaacgaggtcatcagcgagaaacgacctccagccgtacatcgacgacacctcaaccagct
cgagttcctcatggggcgccccggacactccctacgggtcttggagggttagcctcggtacccgaagccg
tggaacctcaactacgtcgagatcggaacgaggacaacctctacggcggtcctcgagacctacatcgct
accggttcaggccctactacgacgccatcacccgcaagtaaccgcacatgaccgtcatggagagcctcac
cgagatgccccggccccgctgcccggcgctcggaactaccaacgtaactcgacgcccgaaggcttcgtcagc
cagttcaactacttcgaccagatgcgggtcaccaaccgcaacgctgaacggcgagatcgccaccgtctacc
ccaacaaccggagcaactcgtggcggtggggcagcccggttcccgctctaccggtggtggatcggtccgt
ggctgaggcggtcttctcctcctcgccgagggagcggaacagccgaagatcctcgccgacgctacgcccc
atgttccgcaacattacaactggcagtgaggcccgacctgatcgcttcgacgcccagacagcagccgga
cgtcgctctacttcttggcagctcatcaagctcctcagcaaccaagatcaccagaaacctgcccac
gacgtggtctgggggggacatcgcccgctctactgggtcgccggccggaaacgacaacacggcgagcaac
atcttcaagggccgctctacaacagcaccagcgacgtcccggtcaccggtccagttcgccgggtgcaacg
ccaagagcgccaacctcaccatctctcgtcggaacgccccacggcagcaactacccggggggccccga
ggtcgtcaagacggagatccagagcgtcacggccaacggcgcccttcgagttcagcctcccgaaac
ctgtcggtgggtggtgtaagacggagtag

도면28a

서열 번호 51

트리코테르마 리세이 유래의 엔도글루카나제, Eg4의 핵산 서열

atgataccagaagcttttccaacctttcttctcaccgcactagcggtggcaaccgggtgtgttggaacaggac
acatcaacaacattgtcgtaacggagtggtactaccagggatatgatcctacatcggtcccatatgaatc
tgacccgcccatagtgggtgggtgacgggtgacgatcttgacaacgggttcgtctcaccgacgcataat
cagagcccggaacatcatctgccacaagaatgccaccaacgccaaggacacgcgtccgtcaaggccggag
aacctattccctccagtggtgacagttccttgccgcacccaggcccatcgctcgactacctggccaa
ctgcaacggcgactgagagaccgtggacaagacgtcccttgagttcttcaagattgacggcgctcggtctc
atcagcggcgagatccgggcaactgggcctcggaagtggtgattgccaaacaacacacctgggttgtca
agatcccgaggatctcgcccgggcaactacgtgcttcggcaacgagatcatcgcttgacagcgccgg
gcaggcggaacggcgctcagaaactacacctcagtgcttcaacctcgccgtccaggctccggatctctgcag
ccgagcggcgctcaagggaaccgcgtctaccactccgatgaccccggtgtcctcatcaacatctacacca
gccctcttgctacaccattcctggacacctccgtggtatcaggcctcccccagagtgctgccaggggcag
ctccgcccgcagcgccactgccagcgccactgttcttgccggtagcgggacgggaaacccgaccagtaag
actacgacgacggcgaggacgacacaggcctcctctagcaggggcagctctactcctcctgctactacgt
cggcacctgggtggaggcccaacccagactttgtacggccagtggtggcagcggtacagtggtcctac
tcgatgcgcgcgcggccacttgctctacctgaaccatactacgcccagtgcttaactag

도면28b

서열 번호 52

트리코테르마 리세이 유래의 엔도글루카나제, Eg4의 단백질 서열

MIQKLSNLLVTALAVATGVVGHghhndivingvwvyqaydpttffpyesnppivvgwtaadldngfvspday
qnpdiichknaatnakghasvkagdtlffqwpvpwphpgpivdylancngdcetvdkttleffkidgvg
lsggdpgtwasdvlnnntwvkipdnlapgnyvlrheialhsaggqangaqnyppqcfniavsqsgslq
psgvlgtldlyhatdpgvliniytsplnyiiipgptvvsglptsvaaggssaataatasatvpggsgptsrtt
ttarttqassrpsstppattsapaggptqtlygqcgsgysgptrcappatcstlnpyyaqcln

도면29a

서열 번호 53

포도스포라 안세리나 유래의 GH3 패밀리 β -글루코시다제, Pa3D의 핵산 서열

atggctcttcaaaccttcttctgctggcggcagccatgctggccaaacgagagacaaacaggcgaaaagg
tctctcggaagcaccgctctggcgctcaagcatgggcccgcggccactcccaggtgcccgcactctggc
cagaatgtcacagcaagacaagatcaacatggtcacgggcatggctgggacagagggccttgctggga
aacacagctgccaatcagctccatcaactatcctcaaatctgtcttcaggatggaccattgggcattcgt
tcggcactggtaccacgccttcacacctggcgctccaagctgcttcgacatgggacgttgatctgatccg
gcagcgcggtgcttaacctggcgccgaagccaaggctgcccgcattcacatccttttggggcccgttgcc
ggtgcccctgggaagattcccacggcggtcgcaactgggagggatttggcgccgaccccctacottgccc
gtattgccatgaaggagaccatcgagggatttcagtcagcagggctccaggccaacgccaagcactacat
tgcaaacgaacaagagctcaaccgcgagaccatgagcagcaatgtggatgaccgcactcagcacgagctc
tacctctggccctttgcgacgcggtgcacgccaactgcgcagcgtcatgtgcagttacaacaagctca
atggcacgtgggcttgcgagaatgacaaggctctgaatcagatcctgaagaaggagctcgattccaggg
ctacgttctcagcgactggaatgctcagcacagcactgctctgtctgctaacagtggtctggacatgact
atgcccgttaccgatttcaacggccgcaatgtctactggggccctcaactgaacaacgctgtcaacgccg
gccaggttcagagatccagatcagacagcatgtgcaagagaatcttggctggctgggtacttgctcggtca
gaaccagggtatcccgcctcaacatcaggggccaacgttcaggggcaaccataaggagaacgtacgtgct
gttgccagagacggcatcgtcttctgctgaagaacgatggaattctgcgcgtttccaagccgagaaagattg
ctgtcgtgggctcccactccgtcaacaatcccagggaatcaacgcctgtgttgacaagggtgcaatgt
tggaaccccttggtcatgggtgggttcaggcagcgtcaactacccctatctcgtgtcccgcgtacgatgct
ctccggactcgtgctcaggccgatggcacaacaatcagcctccacaacactgacagcaccacgggtgtgt
caaacgttgtgtctgacgctgatgctgttgggtgtgcatcactgcgcgttctggtgaagggtacatcac
tgtcgaggggccacgctggcgacccgagccaccttgacccgtggcacaatggcaaccaactgttcaggct
gcgcggctgccaacaagaacgtcatcgttgggtgtgacagtggttgccagatcacctgggagactatcc
tcaacaccaatggagtcgcgcgattgtgtgggtgtgtctccggggccaagagaatggcaacgctcttgt
tgatgttctctacggcttggtttcgccatctggaaagcttccctacaccattggcaagagggagtcggac
tatggcacagccgttggtcgtgggatgataaacttcaggaggggccttttgggtgactaccgtcactttg
acaatgccaggatcgagccgcgtatgagtttggcttggctcttgaagtccagcggcgagttgggt
ttgatttcaagctttcctaacctgataaaacagcttacaccaatttcaccttctccgacatcaagattac
ttccaatgtcaagccggggcccgctactggccagaccattcccggcggaactgcgcgactgtgggagac
gttgcgacagtcactgcaaccatcaccaactcgggtgctgctcgagggcgctgaggttgcccagctttaca
tcggcctgcgctcctcggtcctgcctctcccgcgaagcagctgcgtggattttccaagctgaagctggc
ccgggtgcccagcgccactgcacattcaacctcagacgcagagatctcagctattgggatacccgctc
cagaactgggtcgtgcccagcggaacttgtcgtcagcgtcggcgccagctcgagagatatccgcttga
cgggcaccatcacggcgtag

도면29b

서열 번호 54

포도스포라 안세리나 유래의 GH3 패밀리 β -글루코시다제, Pa3D의 단백질 서열

malqtffilaamlanaettgkvsrqapsgaqawaaahsqaaatlarmsgqdkinmvtgigwdrGPCVG
ntaaissinypqicldgplgirfgtgtaftpgvqaastwdvdlirqrGaylgaEakgcGihillgpva
gaigkiphgrnwegfgadpylagiamketiegiqsagvqanakhyanegelnretmssnvddrtqhel
ylwPfadavhanvasvmcsynkngtwacendkalnqilkkelgfqgyvlsdwnaqhstalsansgldmt
mpgtdfngnrnywgpqlnnavnagqvqrsrlldmckrilagwylggnggypainiranvqgnhkenvra
vardgivilkndgilplskprkiaavvghshsvnpqginacvdkgcnvgtlgmwgsGsvnypylvspyda
lrtraqadgtqislhntdstngvsnvvsdadavvvvitadsgegyitveghagdrshldpwhngnqlvqa
aaaanknvivvvhsvgqitletilntngvraivwaglpqgengnalvdvlyglvpsgklpytigkresd
ygtavvrGddnfreglfvdyrhfdnariepryefgflsytnftfsdikitsnvkpgpatggtipggpad
lwedvatvtatitnsgavegaevaqlyiglpssapasppkqlrgfsklklapgasgtatfnlrrrdlsyw
dtrlqnwvvpvsgnfvsVgassrdirltgtita

도면30a

서열 번호 55

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β -글루코시다제, Fv3G의 뉴클레오티드 서열

atgttttcttcttccatattctgtttggcggccctgagctctgatgagccagggtctactagctcagagcc
aacccggaatgtcatcaccgatgatacctacttctacggccaatcgccaccagtgtatcctacacgtaa
gcactctctctgatttcccaacgaaagcaatactgatctcttgaccagcggaacaggtagacaccggctc
atgggctgcgctgtagccaaagccaagaacttgggtgtccagttgactcttgaagagaaagtcaacttg
actacaggaggccagacgaccaccggctgctctggcttccctggcattcccggtgtaggctttccag
gactgtgttttagcagacgctggcaacgggtgtccgcaacacagattatgtgagctcgtttccctccgggat
tcatgtcgggtgcaagctggaatccggagttgacctacagccggagctactacatgggtgctgaggccaaa
gccaaggcggttaacatccttctcgttccagtatattggacctttggccgagtagttgaagggtggacgca
actgggaagggttttccaatgatccctacctggcgggtaaataggggcatgaagctgtcggcggtatcca
agacgcccggagttgttgcatgcccgaatacttcttctgctcaagagcaggagaccatagacttggcgcg
tctgtcactggggctgatgcaatctcatcaaatctcgtatgacaagacactccatgaattatatctcgtgt
aagcacatcatatcttggctgagtagatgaacctactaacaccggaaactgggcttttcgctgatgcagt
ccacgcccggacttggcagtgatgtgacgtacaaacagagcaaaacaattcacacgctgcccactcgg
aagctttctcaatggccttctcaaggcgagttaggattccagggttttgcgtctcggactggggcgac
agcaatctgggtatggcttcagcattggctggcctggatgtgtcatgcccagctcgatcttgtggggtgc
caaccttacccttgggtgtgaacaacgggaactattcccgagtcacagggttgacaatatgggttacacggtac
gcgaagctctcagccttacttctcaattcttttgaactgacaatcgtgtaggctccttgcacttgggtatc
agttgaaccaggaccaagacacccaagcccgagtcacggactcgtcgcaagctttgggagcctcacc
agtagtcgacgctcgcaacgcaagctccaagcctactatctgggacgggtgcagtcgaggggccatgttctt
gttaagaacaccaacaacgcaactgccattcaagcccaacatgaaactcgtttcttgttcggatactctc
acaaagctcctgataagaacatccagaccccgcccaaggcatgttctcgccttgggtctatcggtgccc
atccgccaacatcactgagctgaacctcggctttctcggaaatttgagctctcacatactccgccatcggc
cccaacggaacatcatctcgggtggaggtcgggtgccagcgttggactctgttcagctcacccttcg
atgcattcgtttctcgggccaagaaagagggtactgcgcttttctgggattttgagagctgggacctta
tgtgaacctacatctgaagctttgcatcgttctgtgtaatgcatgggctagcgaaggctgggtagacct
gcaacctatgatgcctatactgatgagctcatcaataacgctcgtgacaagtgcgctaactactattgtg
ttcttcacaatgctggaacacgacttgtggatggcttcttgggtcaccacaacgctcaccgctattatcta
cgctcatctccaggtcaggatagtgagatgctctggtatcttctgctctatggcgatgagaacctatct
ggtcgctccttaccacggttgcgccgaacgagacggattatgggtcactgctgaagccagacttgactc
tcgcccccacaccagtaaccaacactttcccgagtcgcgacttctccgaggggtatttctcatgactaccgaca
tttcgatgctaagaacatcacgcctcgtctcgagtttgggttcggcttgagctacacaacctttgagtac
gctagtctccagatctcaaagtcaccaggccagacacccgaataccagctgggtgctcttaccgagggag
gcggttcagatttctgggacgctgcttgcactgtcacagcagcgtcaggaaactgggtctgtcgcacgg
caaggaggttgacacagctatacgttgggttccagggtggtcctatgagacagctacgtggctttacgaaa
ccagctattaaggctggagagacggctacagtgacctttgagcttactcgcgcgacttgagtgtctggg
atgttaatgcgcagagtggaacttcagcaaggcaactatgctatctacgttggccgaagtagtcgaga
tttgcctctgcaagtaccttgagcatctag

도면30b

서열 번호 56

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β -글루코시다제, Fv3G의 단백질 서열

mfpssisciaalslmsqqlaqsqpenvitddtyfygqspvpythtgsaaavakaknlvsqtleekv
nlttgqgttttgcsgfipgiprvfgpglcladagnvrntdyvssfpgsihvgaswnpeltysrsyymgae
akakgvnillgpvfgplgrvvegggrnwefsndpylagklgheavagiqdagvvacgkhflaqeqethrl
aasvtgadaissnlddktlhelylcvmsynrannshacqnsklngllkgelgfggfvsdwaqqsgm
asalagldvmpssilwganltlgvnngtipesqvdnmvtrllatwyqlnqdgqteapghlaaklweph
pvvdarnasskptiwdgaveghvlvknntnalpfpknmklvslfgyshkapdknipdaqgmfsawsiga
qsanitelnlglgnlsityaiapngtiisggsgasawtlfsspfdafvsrakkegtalfwdfeswdp
yvntseacivagnawasegwdrpatydaytdelinnvadkcantivvlhnagtrlvdgffghpnvtaii
yahlpqgdsgdalvsllygdenpsgrlpytvarnetdyghllkpdltlapnqyqhfpqsdffsegifidyr
hfdaknitprfefgfglsyttfeyaslsqsksqagtpyepagaltegrsdlwdvvtasvrntgsvd
gkevaqlyvgvpggpmrqlrgftkpaikagetatvtfeltrrdlsvwdvnaqewqlqqgnyaiyvgrssr
dlplqstlsi

도면31a

서열 번호 57

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β -글루코시다제, Fv3D의 뉴클레오타이드 서열

atggctagcattcgatctgtgttgggtctcgggtcttttggccgcggtgtcaatgcccagcctacgatg
cgagtgcgcgcgtgaagatgctttcagctgggtccagcccaagaacaccactattcttggacagtaagg
ccattcgccctcattaccctgccagatgtttcaccactacccaagtgcactgaggtgtactgacatt
ctagacaatgtactggaagggtcgggaagatgccttcgccaagggtcaaaactttgtctcccaactaa
ccctcgaggaaaaggccgaocatggtcacaggaaactccaggtccttgcgtcggcaacatcgtcgccattcc
ccgtctcaacttcaacgggtctctgtcttcacgacggccccctcgccatccgagtagcagactacgccagt
gttttccccgctggtgtatcagcgcgttcacgtgggacaaggacctcctctaccagcgggtctcgcca
tgggtcaagaggttcaaggccaagggtgtctcacatcctcctcgccccgtcgccggtcctcttggccgctc
ggcatactctggtcgttaactgggagggtttctcgccggaaccttacctcactggtattgcgatggaggag
actatcatgggacatcaagatgctggtgttcaggctactgcgaagcactttatcggtaatgagcaggagg
tcattgcgaaacctacttttgtcaaggatgggtatatattggtgaggttgacaaggaggctctttcgtctaa
catggatgatcgaaacctgacagagctttacctctggcccttttgccaatgctgttcatgccagggtctcc
agcatgatgtgctcgtaccagcgtctcaacgggtcctacgcctgccagaactcaaaggctcctcaacggaa
tctgtcgtgatgagcttgggttccagggctacgtcatgtcagattgggtgcccaccacgcgggtgttgc
tgccatcaacagcgggtctcgacatggacatgccccgtggtatcggtgcctacggaacatactttaccaag
tccttcttcggcggaacctcaccgcgcgtcaccacggcaccctcgacgagaccgcgtcaacgaca
tgatcacccgcacatcactcctactcttggctcgccaggacaaggactatcctcgcgtcgaccctc
cagcgggtgatctcaacaccttcagcccccaagagctcctggttcgcgcaggttcaacctcaccggcgagcgc
agccgtgacgtccgcggttaaccaaggcagcttgatccgcaagcacggcgccgaggtctaccgtcctctca
agaaagagaagaacgccttccctcaagaagcccaagtcacgtgctctttggcaacgatgctggtga
tatcactgaggggtttctacaaccagaatgactacgaatttggcactcttgttgcgtggtggtggtctgga
actggtcgtttgacataccttgttctgcctctagccgccatcaatgctcgtgctaagcaggacggtaactc
ttgttcacgagtggtgaacaacactcttattgctaccaccaacgtcactgatctctggtacccctgctac
tcccgatgtctgcctcgttttctgaagacttgggtgagggaggtgctgatcgtgagcacctctccgtt
gactgggagcgttaatgatgttgggtgagctctgttgcgaagtactgcaataacactgtcgtcgtcactcact
cttctggtatcaacactcttccctgggtgacaccccccaacgtcaccgctattctcgtcgcacacttccc
cggtcaggagctctggcaactcctcgttgacctcctctacggcgatgtcaacccctctggtcgtctccc
tacaccatcgcttcaacggcaaccgaactacaacgctccccccaccactgcgcgtcaacaccacggcaagg
aggactggcagctcttgggtcgacgagaagctcgagattgactaccgctacttcgacgcgcacacaactctc
cgtccgctacgaattcggcttcggtctctcctactcacccttcgaaatctccgacatctccgctgagcca
ctcgcatccgacattacctcccaagccgaggtctcctccgtgcagcccgcggaacccccgcctctggg
agaccgtctacaacgtgacccgtctccgtctccaaacgggcaagggtcgacggcgccactgtccccagct
atcgtgacattccccgacagcgcgcctgcccgtacaccacccaagcagctccgtgggttcgacaagggtc
ttccttgaggctggcgagagcaagaggtgcagctttgagctgatgcgcgtgatctgagctactgggata
tcatttctcagaagtgggtcactcctgagggagaggttactattcgtgttggaattcagcagtcgggactt
gaaggaggagacaaagggttactgttgggtgaggcgtaa

도면31b

서열 번호 58

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β -글루코시다제, Fv3D의 단백질 서열

masirsvlvsqllaagvnaqaydasdraedafswvpknntilgqyghsphypannatgkgwedafaka
qnfvsqitleekadmvtgtgpgpcvgnivaiprlnfnglclhdgplairvadyasvfpagvsaaasswdkd
llyqrglamgqefkakgahillgpvagplgrsaysgnrwegfspdpyltgiameetimghqdagvqata
khfigneqevmrnptfvkdgyigevdkealssnmddrtmhelylwpfanavhakassmmcsyqrlngsy
acqnskvlnqilrdeifggyvmsdwgathagvaainsgldmmpggigaygtyftksffggnltravt
ngtldetrvndmitrimtpyfwlgqdkdypsvdpssgdlnfsspksswfrelnltgersrdvrgnhgdl
irkhgaestvllkneknalpkkpksiavfgndagditegfynqndyefgtlvagggsgtgrltylvsp
laainarakqdgltlvqgwmntliattntvtdlwipatpdvclvflktwaeeaaadrehlsvdwdgndvve
svakycnntvvthssgintlwpadhpntailaahfpgqesgnsldllygdvnpssgrlpytiafngt
dynappttavnttgkedwqswfdekleyryfdahnisvryefgflsystfeisdisaeplasdits
qpeditpvgpggpawetvynvtvsvsntgkvdgatvpqlyvtfpdsapagtppkqlrgfdkvfleage
sksvsfelmrdrldisqkwllipegeftirvgfssrdlkeetkvtvvea

도면32a

서열 번호 59

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β -글루코시다제, Fv3C의 뉴클레오타이드 서열

atgaagctgaattgggtgcgcgcagccctgtctataggtgctgctggcactgacagcgcagttgctcttg
cttctgcagttccagacactttggctgggtgtaaggctcagtttttttccaccatttctcgtctaatctc
agccttgtttgccatatagccttggctcgtcggacgcacgcacagatcgcgatcatttctcctccttgc
agccttgggttctcttctacgatcttccctccgcaattatcagcgccttagtctacacaaaaaccccgag
acagctctttcattgagtttgtcgacatcaagttgcttctcaactgtgcatttggctgggtgtctacttct
gcctctagacaaaccaaatctggggcgcaattgaccgctcaaaccttggttcaataaaccttttttattcgag
acgcacatttataaatatgcgcctttcaataataccgactttatgcgcggcggtgctgtggcggtgat
cagaaagctgacgctcaaaagggttgtcagcagagatacactcgcatactcgcgcctcattatccttcac
catggatggacccctaattgctgttggctgggaggaagcttacgcacaaagccaagagctttgtgtcccaact
cactctcatggaaaagggtcaacttgaccactgggtgttgggtgaagcagctccttgcaaacagggtatctca
atccctcagctaaacaacttctcagatggcgaaggcgaaacgctgtgtaggaacgtgggatcaattcctcg
tctcggatgacgaggtctctgtctccaggatggctccttggaaattcgtctgtccgactacaacagcgt
tttcccgctggcaccacagctgggtgttcttggagcaagctctctgtgtatgagagaggtctcctgatgg
gcactgagttcaaggagaaagggtatcgatatcgtcttggctcgtactggacctcttggctcgcactgc
tgctgggtggacgaaactgggaaggcttccacgcttgatccttatatggctggccacgccatggccgagggc
gtcaagggtatttcaagacgcaggtgtcattgttgtgttaagcattacatcgcaaacgagcagggtaagc
cacttggacgattttaggaattgacagagaactgacccctctgttagagcacttccgacagagtgccgaggg
tcagttcccgcaagtaacaactctccgaggtctctctcctccaaactggatgacaagactatgcacgagct
ctacgcctggcccttctcgtgacgcgcgtccgcgcgcgcgtcgggtccgctcgtgtctgtacaaaccagatc
aacaactcgtacggttggcagaactccaagctcctcaacgggtatcctcaaggacgagatgggttccagg
gtttcgtcattgagcgaattggggcgcccaagcataaccggtgcgcgttctgcgcgtcgtgtgtcgtatgag
catgctgtgtgacactgccttcgacagcggatacagcttctggggcggaacttgactctggctgtcctc
aacggaaactgttccgcgcgtggcaggttatgacatggctctcgcgaatcatgtctgccttcttcaagggtg
gaagacgatagaggtatcttccgacatcaacttctcctcctggaccgcgcacaccttccggttctgtgca
taacttgtctcaagagaaccgcgagcaggtcaacttggagtcacagtcacgacacacacagagccac
atccgtgagggcgcgtgcgaagggaagcgtcgtgctcaagaacacccgggtcccttccctcaagaacccaa
agtctcctcgtgtcattgggtgaggacgcgcgtcccaacccctgctggacccaatgggtgtgtgtgacgcgtg
ttgcgataaatggtacccctggctatgggttggggctcgggaacttcccaattcccttacttgatcaccctc
gatcaagggtctctaatcgagctactcaagacggaaactcgatatgagagcatcttgaccaacaacgaat
gggttccagtacaagctcttgtcagccagcctaactgacgcgtatcgttttgcgaatgcccactctgg
tgagggatacattgaagtcgacggaaacttgggtgatcgaagaacctcaccctctggcagcagggagac
gagctcatcaagaacgtgtcgtccatatgcccacaacacattgtagtcttgcacacgcgtcggccctgtcc
tactcgcgcgactacgagaagaaccccaacatcactgcatcgtctgggtgtgttcccgcccaagagtc
aggcaatgccatcgtgatctcctctacggcaagggtcagccctggccgatctccttcaacttggggccgc
accgcgagagctacgggtactgaggttctttatgaggcgaacaacggccgtggcgctcctcaggatgact
tctctgaggggtgtcttcatcgactaccgtcacttccgaccgacgatctccaagcaccgatggaaaagagctc
tcccaacaacacccgctgctcctctctacgagttcgggtcaggtctatcttgggtccacctttagtactct
gacctcaacatccagaagaacgtcgagaacccctactctcctcccgctggccagacatccccgccccaa
cctttggcaacttcagcaagaacctcaacgactacgtgttcccaaggcggtccgatacatctacaagtt
catctaccccttctcacaacccctcctcatccgcagcgaggtatccaacgatgggtggccagtttggtaag
actgcgaagaggttctcctcctcccaacgcctcaacgggtcagcccagcctcgtcttccgcctctgggtg
cccaggtggtaacctcaattgtgggacatcttgtacacgcgtcacagccacaatcaccaacacagcga
cgccacctccgacgagattccccagctgtatgtcagcctcgggtggcgagaacgagcccatcgtgttctc
cgcggttccgacggtatcgagaacattgtctccggccagagcgccatcttcaacgctcaattgacccgtc
gcgatctgagtaactgggatacaaatgcccagaactgggtcatcactgaccatccaagactgtctgggt
tggaagcagctctcgaagctgcctctcagcgccaagttggagtaagaaagccaacaagggttgttttt
tggactgcaattttttgggaggacatagtagcgcgcgcagttacgtc

도면32b

서열 번호 60

푸사리움 베르티실리오이테스 유래의 GH3 패밀리 β-글루코시다제, Fv3C의 단백질 서열

mklnwvaalsigaagtdsavalasavpdtlagvkkadaqvvttrdtlayspphyppswmdpnavgweea
yakaksfvsqiltlmekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirisdynsafpagtttag
aswkslwyergllmgtefkekgidialgpattgplgrtaaggrnwegftvdpymaghamaeavkqi dag
viacakhyaneghefrqsgvevqsrkyniseslssnlddktmhelaywpfadavragvgsvmcsynqinn
sygcqnsklngilkdemgfggfvmssdwaaghtgaasavagldmsmpgdtafdsqysfwggnltlaving
tvpawrvddmalrimsaffkvgtiedlpdinfsswtrdtfgfvhtfagenreqvnfgvvnqhdhkshir
eaaakgsvvlkntgslplknkplflavigedagpnpagpnpgcgrgndngtlamawgsqtsqfpyltpdq
glsnratqdgtryesiltnewasvqalvsqpnvtaivfanadsgegyievdngfgrknltlwqqgdel
iknvssicpntivvlhtvgpvlladyeknpnitaivwaglpqgesgnaiadllygkvspgrspftwgrtr
esygtevlyeanngrgapqdddfsegvfidyrhfdrrspstdgksspnntaaplyefghglswstfeysdl
niqknvenpysppaggtipaptfgnfsknlndyvfpgkvryiykfiypflntssaseasndggqfgkta
eeflppnalngsaqprlpasgapgnpqlwdilyvtatitntgnatsdeipqlyvslggenepirvlrg
fdrieniapggsaifnaqltrrdlsnwdtnaqnvwitdhpktvwvgsssrklplsakle

도면33a

서열 번호 61

트리코테르마 리세이 유래의 GH3 패밀리 β-글루코시다제, Tr3A의 뉴클레오티드 서열

atgcgttacccgaacagcagctgcgctggcacttgccactgggcccctttgctagggcgagacagtcagtata
gctgggtcccatactgggatgtgatattgtatccttgagacacccatgctgactcttgaatcaaggtagctca
acatcgggggcctcgctgagggcagttgtacctcctgcagggaactccatggggaaccgcgtacgacaagg
cgaagcgccgcatctggcaagctcaatctccaagataaggtcggcatcgtgagcgggtgctgggtggaacgg
cggtccttgctgtggaacacatctcggcctccaagatcagctatccatcgctatgccttcaagacgga
ccctcgtgtgtcgatactcgacagggcagcacagccctttacgcccggcggttcaagcggcctcgacgtggg
atgtcaatttgatccgcgaacgtggacagttcatcggtgaggaggtgaaggcctcggggattcatgtcat
acttggtcctgtgggtgggcccgtgggaaagactccgagggcggtcgcaactgggagggcttcggtgtc
gtcccatatctcacgggcattggccatgggtcaaaccatcaacggcatccagtcggtaggcgtgcaggcga
cagcgaagcactatctcctcaacgagcaggagctcaatcgagaaaccatttcgagcaaccagatgacgg
aactctccatgagctgtatacttgccatttgccgacgggttcaggccaatgtcgttctgtcatgtgc
tcgtacaacaagggtcaataccacctgggctgaggatcagtaacgctgcagactgtgctgaaagacc
agctgggggttcccaggctatgtcatgacggactggaaacgcacagcacagactgtccaaagcggaattc
tggtgtgacatgtcaatgcctggcacagacttcaacggtaacaatcggtcgtgggttcagctctcacc
aatggcggtaaaatagcaatcaggtccccacgagcagagtcgacgataatgggtgactcgtatcctcggccat
ggtacttgacagggccaggaccaggcaggtatccgtcgttcaacatcagcagaaatgttcaaggaaacca
caagaccaatgtcagggaattgcccaggacggcatcgttctgctcaagaatgacgccaacatcctgccc
ctcaagaagcccgtatgacattgcgctcgttggtatcgcgcgaatcattggtaaccacgccagaaactcgc
cctcgtgcaacgacaaaaggctgcgacgacggggccttgggcatgggttggggttcgggcgcgtcaacta
tcogtacttctgcgcgccatcagatgccatcaataccagagcgttctcgacgggcacccagggttaccttg
agcaacacgcgacaacagctcctcaggcgcatctgcagcaagaggaaaggacgtcgccatcgtcttcatca
ccgcccactcgggtgaaggctacatccgctggagggcaacggggcgatcgcaacaacctggatccgtg
gcacaacggcaatgccctggtccaggcggtggccggtgccaaacagcaacgtcattgttgttgcactcc
gttggcggccatcattctggagcagattcttgccttccgcaggtcaaggccggtgtctggggcggtcttc
cttctcaggagagcggaatgcgctcgtcgacgtgctgtggggagatgtcagcccttctggcaagctgggt
gtacaccattgcgaagagccccaatgactataaactcgcacgtcgttccggcgggcagtgacagcttcagc
gagggactgttcatcgactataagcacttcgacgacgccaatatcacgcgcgggtacgagttcggtatg
gactgtgtaagtgttctaacctgaacaatctattagacaggttgactgacgagatgactgtggaatgatag
cttacaccaagtccaactactcacgcctctcctgcttgcgacggccaagtctggctcctgcgactggggc
cgttgtgcccgggagggcccagtgatctgttccagaatgtcgcgacagtcaccgttgacatcgcaactct
ggccaaagtgaactggtccgaggttagccagctgtacatcacctacccatcttcagcaccaggaacccctc
cgaagcagctgcgaggctttgccaaagctgaacctcacgcctgggtcagagcggaacagcaacgttcaacat
ccgacgacgagatctcagctactgggacacggcttcgcagaaatgggtggtgcgctcggggctgtttggc
atcagcgtgggagcagcagcgggatatcaggctgacgagcactctgtcgttagcgtag

도면33b

서열 번호 62

트리코테르마 리세이 유래의 GH3 패밀리 β -글루코시다제, Tr3A의 단백질 서열

MRVRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKNLQDKVGIVSGVGWNG
GPCVGNTPSPASKISYPSLCLQDGLGVRYSTGSTAFTPGVQAASWDVNLIRERQFIGEEVKASGIHVI
LGPVAGPLGKTPQGGRNWEGFVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDR
TLHELYTWPFDADVQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPYVMTDWNQHTTVQSANS
GLDMSMPGTDFTNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNH
KTNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAIIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNY
PYFVAPYDAINTRASSQGTQVTLSDNTSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPW
HNGNALVQAVAGANSNVIVVHSGAIIILEQILALPQKAVVWAGLPSQESGNALVDVLWGDVSPSGKLIV
YTIKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDANITPRYFEGYLSYTKFNYSRLSVLSTAKSGPAT
GAVVPGGSDILFQNVATVTVDIANSQGVGTGAEEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATF
NIRRRDLSYWDITASQKVVVPSGSGISVGASSRDIRLTSTLSVA

도면34a

서열 번호 63

트리코테르마 리세이 유래의 GH3 패밀리 β -글루코시다제, Tr3B의 뉴클레오타이드 서열

atgaagacgttgtcagtggttgcgtgcgcgccttttgccggccgtagctgagggccaatccctaccgcctc
ctcactccaaccaggcgtagctgcgcctcctttctacccttcgccaatggatggacccagtgctccaggctg
ggagcaagcctatgcccaagctaaggagtgctgcggccttgactctcttgagaaaggtcaacctcacc
accggtgttgcttgatgggtgagaaagtgcttggaacggttggtacccgtgcctgccttgggcatgcaa
gtctttgcatgcaggacggcccccctgggtctccgattcaacacgtacaaacagcgctttcagcggttgctt
gacggccgcgcgcagctggagccgacacctttgggttgaccgcggtaccgctctgggctccgagggcaag
ggcaagggtgtcgatgtctctcgcgaccgctggctggccctctcgctcgcaaccccaacggagggcgt
acgtcgaggggttcggctcgatccctatctgcccgggtttggctctggccgataccggtgacgggaatcca
gaacgcggggcaccatgcctgtgccaagcacttccctcctcaacgagcaggagcatttcgcgcaggtcggc
gaagctaaccgttacgggatacccccacacggaggtctgtcttccaacggttgatgacaagacgattcacg
aggtgtacggctggcccttccaggatgctgtcaaggctgggtgctgggtccttcattgtgctcgtaacca
ggtcaacaactcgtagccttgccaaaactccaagctcatcaacggcttgctcaaggaggagtagcgtttc
caaggctttgtcatgagcgactggcaggccagcacagggtgtcgctctgctgttgccggtctcgata
tgaccatgcctggtagacacgccttcaacacggcgcatcctactttggaagcaacctgacgcttgctgt
tctcaacggcaccggtcccgagtgggcgacttgacgacatgggtgatcgatcatggtcctcttctcaag
gtgggcaagacggttgacagcctcatgacaccaactttgattcttggaaccaatggcgagtagcgtacg
ttcaggccgcctgcaatgagaactgggagaaggtaactacggcgctcgatgtccgcgcaaacatgagaa
ccacatccgcgaggttgccgccaagggaactgtcatcttcaagaacaacggcatcctgcccttaagaag
cccaagttcctgacggtcattgggtaggatgctggcggaacccctgcccggccccaacggctgcggtgacc
gcggctgtgacgacggcactcttgccatggagtggggatctggtagtactaccaacttccctaccctcgtaac
cccgcagcgccgctcgagagccaggtctccaggacggcaccgctacgagagcatcctgtccaactac
gccatctcgagacccaggcgctcgtagccagcccgatgccattgccattgtctttgccaaactcggata
gcggcgagggctacatcaacgtcgatggcaacgagggcgaccgcaagaacctgacgctgtggaagaacgg
cgacgatctgatcaagactgttgctgtgtcaaccccaagacgattgtcgatccactcgacggccccc
gtgattctcaaggactacgccaacaccccaacatctctgccattctgtgggcgggtgctcctggccagg
agtctggcaactcgctgggtcgacattctgtacggcaagcagacccgggcccgcactcccttcactgggg
cccgctcgctggagagctacggagttagtggtatgaccacgcccacaacggcaacggcgctcccaggat
aacttcaacgagggcgcccttcacgactaccgctactttgacaagggtggctcccggcaagcctcgagct
cggaacaaggctcccacgtacgagtttgcttcggactgtcggtcgacgttcaagttctccaacctcca
catccagaagaacaatgtcgcccccatagagcccccacggcaagacgattggcgctccctctctgggc
agcttcagcaagaaccttaaggactatggcttccccaaagcgttcgcccacatcaaggagtttatctacc
cctacctgagcaccactacctctggcaaggagcgctcggtgacgctcactacggccagactcggaagga
gttccctcccgcgggtgccttgagcggcagccctcagcctcgctctcgggcctctggcgaaacccggcggc
aacgcgagctgtacgacattctctacacggtagcggccaccattaccaacacgggctcggtcatggacg
acggcgttcccagctgtacctgagccacggcggtcccaacgagccgcccagggtgctcggtggcttcca
ccgcatcgagcgcatgtctccggccagagcgtagcgttcaaggcagacgtgacgcgcgtgacgtgtcc
aactgggacacgaagaagcagcgagtggtcattaccgactacccaagactgtgtacgtggcgagctcct
cgcgcgacctgcgcgtgagcgcccgctgccatga

도면34b

서열 번호 64

트리코테르마 리세이 유래의 GH3 패밀리 β -글루코시다제, Tr3B의 단백질 서열

mkltlsvfaaallaavaeanpyppphsnqaysppfypspwmdpsapgwqayaqakefvsagltllekvnlttgvvgwmgekcvgnvgtpvrlgmrsicmqdgpplglrftynsafsvgltaaswsrhlwvdrgtalgseakgkgvdlvllgpvagplgrnpggrnvefgsgdpylaglaladtvtgignagtiacakhflneqehfrqvgeangygyppitealssnvddktihevgywpfqdavgkagvgsfmcsvngvnnsyacqnsklingllkeeygfggfvmsdwqaghtgvasavagldmtmpgdtafntgasyfgsnltlavingtvpewridmvmrimapffkvgktvdsldtnfdswtnggygyvqaavnenwekvnygvdranhanhirevgakgtvifknngilplkkpkfltvigedaggnpagngcgdrgcdgtlamewsgsttnfpylvtpdaalqsgalqdgtryesilsnyaisqtqalvsgpdaiaivfansdsggyinvdgnegdrknltlwknngddliktvaavnpkti vvihtgppvilkdyanhnpsailwagapqgesgnsldilygkqspgrtpftwgspslesygvsvmttppnngngapqdnfnegafidryfdkvapgkprsscdkaptiefgflswstfkfslhiqknnvgpmsppngktaapslgsfksknldkygfpknvrrikefiypylstttsgkeasgdahyggatkeflpagaldgspqprsaasgepggnrqlydilytvtatitntgsvmdavpqllylshggpnepkvlrgfdrieriapggsvtfkadltrrdlsnwdtkkqwwitdypktyvgsssrldplsarl

도면35a

서열 번호 65

트리코테르마 리세이에서의 발현을 위해 코돈 최적화된, 탈라로마이세스 에메르소니 유래의 GH3 패밀리 β -글루코시다제, Te3A의 뉴클레오타이드 서열

atgcgcgaacggcctcctcaaggtcgccgcttagccgctgccagcgccgtcaacggcgagaacctcgccacagcccccccttctaccaccagccctggggcaacggccagggcgactggggcgaggcctaccagaaggccgctccaggttcgtcagccagctcaccctcgccgagaaggtcaacctcaccaccggcaccggctgggagcaggaacgctcgctcgccaggtcgccagctcgccagcatccccgcttaggcttcccgccctctgcacgcagacagccctcgccgctcgccgacacagactacaacagcccttccctgcccggcttaacgtcgccgccacctgggacgcaacttagcctacgcagagggctcgccatggggcgaggaacaccggcgcaaggggctcgacgtccagtttaggccccgtcgccggcccccttagggcgctctcctgatgcggccgcaactgggagggcttcgccccgacccccgtcctcaccggcaacatgatggccagcaccatccagggtaccaggatgtggcgctcattgctcgcccaagcaacttcctctctacgagcaggaacacttcggccagggcgcccaggacggctacgacatcagcgacagcatcagcgcccaacggcgacagcaagacatgcacgagttatacctctggcccttcgccgatgccttcggcgccggtgtcgccagcgtcatgtgcagctacaaccaggtcaacaacagctacgctgcagcaacagctacacatgaacaagctcctcaagagcgagttaggcttccagggttcgtcatgaccgactggggcgccaccacagcgccgctcgccctcgccggctcgacatgagcatgcccggcgacattgctctcgacacgggcaacgtctttctggggcaccacactcaccgttcgctcctcaacggctccatccccgagtgccgctcgacgacatggccgctcgccgcatgagcgctactacaaggtcgcccgagaccgctacagcgctcccccaacttcgacagctggaccctcgacacactcggccccgagcactacgctcgcccgaggccagaccaagatcaacgagcagctcgccggcaaccacgcggagatcatccacgagatcgccggccgctccgccgtcctcctcaagaacaaggcgccctccccctcactggcaccgagcgttcgctcggtgtctttgcaaggatgctggcgacaaacccctggggcgctcaacggctgcagcgaccggctgcgacaacggcacccctcgccatgggctggggcgagcgccaccccaacttccctacctcgtaacccccgagcagccatccagcgcgaggtcctcagccgcaacggcgaccttcacggcatcaccgacaacggcgcttagccgagatggccgctgcccctctcaggccgacacctgctctgtctttgccaacggcgactccggcgagggctacatcaccgtcgatggcaacgaggcgacccgcaagaacctcaccctctggcgagggcgccgaccaggtcatccacaacgtcagcgccaactgcaacaacaccgtcgtcgtcttaacacacgtcgccccgctcctcatcgacgactggtaacgacccccaacgtcacccgcatcctctggggcggttaacccggtcaggaaagcgccaacagcctcgtcgacgtcctctacggcgctcaaccccggaagaccccccttaccctggggcgagacggcgacgactatggcgccccctctcatcgtaagccttaacaacggcgagggcgcccccgagcaggttcaaccgagggcatcttcacgactacggcgcttcgacaaagtaacaacatcacccccctctacgagttcggttcggcttcggcctcagctacaccaacttcgagttcagccaagttaaagctccagcccatcaacgcctcctacacccccgcagcggtttaacgaaggccgcccagagcttcggccagccctccaatgccaagcgacaacctctacccctagcgacatcgagcggtccccctctacatctacccctggctcaacagcaccgacctcaaggccagcgcccaacgacccccgactacggcctccccacggagaagtacgtcccccccaacggcccaacggcgacccccagccattgacctgcccggggtgccccggggcgaacccccagcctctacgagccgctcgccccgctcaccacatcatcaccacacgggcaaggtcaccggcgacgaggtccccagctctatgtcagcttagggggccctgacgacgcccccaaggtcctccgcggtcttcgacccgacatccccctcgccccgtggccagcagtaacctggaccaaccacctcactcgcccgacatcagcaactgggacccccgtcaccagaaactgggtcgtcaccacactacaccaagaccatctacgtcggaacagcagccgcaacctccccctccaggccccctcaagccctacccccggcatctgatga

도면35b

서열 번호 66

탈라로마이세스 에메르소니이 유래의 GH3 패밀리 β -글루코시다제, Te3A의 단백질 서열

mrnqlkvaalaasavngenlaysppfyppswangqgdwaeayqkavqfvsqtltaekvnlttgtgweq
drcvqgvgsiprlgfpplcmqdsplgvrdtdynsafpagvnvaatwdrnlayrrgvamgeehrgkgvdvq
lgpvagplgrspdagrnwegfapdpvltgmmastiqgiqdagviacakhfilyeqehfrggaqdydis
dsisanaddktmhelylwpfadavragvgsvmcsynqvnnssyacsnsytmnklikseigfggvmtdwgg
hhsqvgasalagldmsmpgdiafdsgtsfwgtnltvavlngsipewrvddmavrimsayykvgrdrysvpi
nfdswtldtygpehyavggqgtkinehvdvrgnhaeiheigaasavllknkggplltgterfvqvfqkd
agsnpwgvngcsdrgcdngtlamgwsgtanfpylvtpaqaiqrevlsrngtftgitdngalaemaaaas
qadtlclvanadsgegyitvdgnegdrknltlwqgadqvihnvsancnntvvvlhtvgpvliddwydhpn
vtailwaglpqgesgnsldvlygrvnpktpftwgrarddygaplivkpnnkggapqddftegifidyr
rfdkynitpiyefgfglsyttfefsqlnvqipinappypasgftkaagsfgqpsnasdnlypsdiervpl
yiyplwnstdlkasandpdyglptekeyvppnatngdpqpidaggapgnpslyepvarvttiitntgkv
tgdevpqiyvslgppddapkvrlrgfdritlapggqylwtttlrrdisnwdpvtqnwvvtntyktiyygn
ssrnplqaplkpypgi

도면36a

서열 번호 67

아스페르길루스 니케르 유래의 GH3 패밀리 β -글루코시다제, An3A의 뉴클레오타이드 서열

atgcgcttcaccagcatcgagggcgtcgccctcaacggcgtcagcctcgccagcgccgacgagttagcct
acagccccctactacccagccctgggccaacggccagggcgactgggcccagggcctaccagcgccg
cgctcgacatcgctcagccagatgacccctcgccgagaaggtcaacctcaccacggcaccggctgggagtta
gagttatgcgtcgccagactgggtggcgtcccccgccctcgccatccccggcatgtgcccaggaacagcc
cctcgcgctcgccgacagcgactacaacagcgcccttcctcgccggtcaacgtcgccgcccactggga
caagaacctcgctacctcgccggccagggcctggggcaggaattcagcgacaagggcgccgacatccag
ttaggccccgtcgccggccctttaggccgctctcccgacggcgccagaaactgggaggggttcagccccg
accccgctctcagcgcgctcctctcgccgagactatcaaggccatccaggatgctggcgctcgccgac
cgccaagcactacattgcctacgagcaggaacacttcggccaggcccccgaggcccgagggctacggcttc
aacatcaccgagagcgccgacgccaacactcgacgacaagaccatgcacgagttatacctctggcccttcg
cgacgcccattagagctggcgctgggtgctgcatgtgcagctacaaccagatcaacaacagctacggctg
ccagaacagctacacccctcaacaagctcctcaaggccgagttaggcttcagggttcgctatgtccgac
tgggccgcccaccacgcccgtcagcgccgcttagccggcctcgacatgagcatgcccggcgacgtcg
actacgacagcgccacagctactggggcaccacactcaccatcagcgctcctcaacggcaccgtccccca
gtggcgctcgacgacatggcgctcgccatcagggccctactacaaggtcgcccgcgacccgctctgg
accccccccaacttcagcagctggaccccgacgagtagcggttcaagtaactactacgtcagcgagggcc
cctatgagaaggtcaaccagttcgtaacgtccagcgcaaccacacgaggttaatccgcccgcacggcg
cgacagcaccgctcctcctcaagaacgacggcgccctccccctcaccggcaaggaacgctcgtcgccctc
atcgcgaggaacggcgacgcaacccctacggcgccaaacggctgacgcaacggcgctgcgcaacggca
cctcgcctatgggctggggcagcgccaccgccaacttccttacctcgctacccccgagcaggccatcag
caacgaggtcctcaagaacaagaacggcgctctttaccgcccagcaactgggcccacgacagatcgag
gcttagccaaagaccgctctgtcagcctcgcttttgtcaacggcgacagcgggcgagggctacatcaacg
tcgacggcaacactcgccgacggcgcaacactcaccctctggcgcaacggcgcaacgtcatcaaggccgc
cgccagcaactgcaacaacaccatcgctcatcaccacagcgctcgcccccgtcctcgtaacgagtggtac
gacaacccccacgtcaccgcatcctctggggcggttaccggccaggaagcgggcaacagcctcgccg
acgtcctctacggccgctcaacccctggcgccaaagagcccccttcactggggcaagaccgagggccta
tcaggactacctctacacggagcccaacacggcaacggcgccccccaggaagatttcgtcgagggcgct
tttatcgactaccggcgtttgacaagcgcaacgagactccatctacgagttcggctacggcctcagct
acaccaccttcaactacagcaacctccaggtcgaggtcctcagcgccctgcctacgagcccgccagcg
cgagactgagggcgccccacacttcggcgaggtcggaacggccagcgactactataccccgacggcctc
cagcgcatcaccagttcatctacccctggctcaacagcaccgacactcgaggccagcagcgcgacgct
cttaaggccaggacgctcctcagctacctccccgagggtgccaccgacggcagcgctcagcccattacc
tgccgtggcggtgctggcgcccaacccccagactctacgacgagctgatccgctcagcgtaacacatcaag
aacaccggcaaggtcgctggtagcagaggtccccagctctacgtcagcttagggcgccctaacgagccca
agatcgctcctcgccagttcgagcgcatcaccctccagcccagcaaggaaactcagtgagcaccaccct
cactcgccgacactcgccactggaacgtcgagactcaggactgggagatcaccagctacccaagatg
gtctttgcccggcagcagcagcgcaagctccccctcgccgacgctccccaccgtccactgatga

도면36b

서열 번호 68

아스페르길루스 니케르 유래의 GH3 패밀리 β -글루코시다제, An3A의 단백질 서열

mrftsieavaltavslasadelayspppyppswangggdwaeyqravdivsqmtlaekvnlttgtgwel
elcvqgtgvgprlgipgmcaqdsplgvrdsdynsafpagvnaatwdknlaylrgqamqgefstdkgadiq
lgpaagplgrspdggrnwefspdpalsgvlfaetkigiqdagvvatakhyaieqehfrqapeaqgygf
nitesgsanlddktmhelylwpfadairagagavmcsynqinnsygcqnsytlnklkaelgfggfmsd
waahhavgsgalagldmsmpgdvdydsqtsywgtntltisvlngtvpqwrddmavrimaayykvgrdlw
tpnfnsswtrdeygfkyyyvsegpyekvnqfwnvqrnhselirrigadstvlkndgalpltgkerlval
igedagsnpygangcsdrgcdngtlamgwsggtanfpylvteqaisnevlknkngvftatdnwaidgie
alaktasvslfvnadsgegyinvdgnlgdrnltlwrngdnvikaaasncnntiviihsvgpvlvnewy
dnpnvtailwggpggesgnsladvlygrvnpgakspftwgtkreayqdylytepnngngapqedfveg
fidyrqgfdkrnetpiyefgyglsytnfnysnlqvevlsapayepasgeteaaptfgevgnasdylypdgl
qritkfiypwlnstdleassgdasygqdasdylpegatdgsaqpilpagggaggnprlydelirvsvtik
ntgkvagdevpqiylvslggpnepkivlrqferitlqpsketgwsttltlrdranwnvetqdwetsypkm
vfagsssrklplrslptvh

도면37a

서열 번호 69

푸사리움 옥시스포름 유래의 GH3 패밀리 β -글루코시다제, Fo3A의 뉴클레오타이드 서열

atgaagctgaactgggtgcgcgcagccctctctataggtgctgctggcactgatgggtgcagttgctcttg
cttctgaagttccaggcactttggctgggtgtaaggctcggttttttaccatttctctacctaactctcag
ccttggttgcacatagcccttattcgctcggaagctacgcaccaaatacgcgatcatttctcccttgacag
ccttggtttcttttttgcagcttctccctccgcgaatcgccagcacccttagcctacacaaaaaccccgaga
cagttctcattgagtttgcgacatcaagttgcttctcaagtggtgcatttgcgtggctgtctacttctgcc
tctagaccaccaaactcggcgcaattgatcgctcaaaccttgctgaataagccttttattcgagagct
ccaatttttacagagaatgtacctttcaataataccgacgcttatgcgcggcggtggctgctgtgatgggt
gttgatcagaatactgacgctcaaaaggttgctacgagagatacactcgacactcactcctcactatc
cttccacatggatggatcctaattgccattggctgggaggaagcttacgcacaaagcaaaagaaactttgtgtc
ccagctcactctcctcgaaaaggctcaacttgaccactggtgttggttaagtagctccttgcaacagtg
atctcggtctccttgactaacgactctctcaggtggcgaaggcgaacgctgtgtaggaaacgtgggacaa
ttcctcgctcttggtatgagaggtctttgtcttcaggtatggtcctcttggaattcgtctgtccgattaca
cagtgcttttcccgctggcaccacagctgggtgcttcttggaagcaagctctctggtatgagaggggtctt
ctgatgggaactgagttcaaggggaaggttatcgatatcgctcttgccctgctactggctctcttgcc
gcaactgctgctgggtgagcaaaactgggagggctttacggtgatccttatatggctggccatgccatggc
cgaggccgtcaagggcatccaagacgcaggtgctcattgcttggctaaagcattacatcgcaaacgagcaa
ggtaagccaattggaaggtttgggaatcgacagagaaactgaccccttgtagagcacttccgacagagt
ggcgaggtcccaacacaggtacacatctccgagttctctcctcccaactggagcagacaagactttgc
acgagctctacgctggccctttgctgatgccgtccgcgctggcgtcggttcagtcagtgctcttaca
tcagatcaacaactcgtaggttgccagaactccaagctcctcaacggtatcctcaaggacagatgggt
ttccagggcttctgcatgagcgattgggcccgcagcacacgggtgctgcttctgcgctcgctgggtcttg
atatgagcatgctgggtgacccgcttcgacagtgatagcttctgggtggaaacctgactcttg
tgtcatcaacggaactgttcccgctggcgagttgatgacatggctctgcgaatcatgtcgccctcttc
aaggttggaagacggttagaggacctcccgacatcaacttctcctcctggacccgcgacaccttcggct
tcgtccaaacatttgctcaagagaaacccggaacagtcactttggagtttaacgtccagcacgaccaca
gaaccacatccgtgagttctgcgcgaagggaagcgtcatcctcaagaacacccgctccttccctcaac
aatcccaagttcctcgctgctcattgggtgaggaagcgggtcccaacctgctggacccaatggttgccg
accgtggttgcgacaattggtacctggctatggcttggggtcggaacttctcaattcccttacttgat
cacaccgaccaaggtctcagaacagagctgcccagacggaactcgatagagagcatcttgaccaac
aacgaatgggcccagacacaggtcttctcagccaaacccagtgacgctatcgttttgcaacgccc
actctggtgaggggttacattgaagtgcacggaacttcggtgatcgaaagacctcaccctctggcaaca
gggagacgagctcatcaagaacgtctcgtccatctgcccacacacattgtcgttctgcatacagctggc
cctgctcgtctcgccgaactacgagaagaaccccaacatcacccgcatcgtctgggtggtcttcccgcc
aagagttctggcaatgcatcgtgatctcctctacgcaaggtaagccctggccgatctccttcaacttg
ggcgccgacccgtgagagctacggtaccgaggttcttattgagggcaaacacgcccgtggcgctcctcag
gatgacttctcgagggtgtcttcattgactacgctcactttgatcgacgatctcccagcaccgatggca
agagcgtcccaacacacagctcctcctctacgagttcggtcatggtgtgtcttggactacacttga
gtattcagaactcaacatccagaagaacggttaactccactactcctcctcgtggtcagaccattcct
gcccacacctttggaacttcagaagaacctcaacgactacgtgttccctaagggtgtccgatacatct
acaagttcatctacccttctgaacacttctcctcctcgcgcagggacatctaacgacggcgccaggt
tggtaagactgcccgaaggttctcactccaaacgcccctcaacggctcagcccagcctcgtcttccctct
tctggtgcccagaggttaacccctcaatttggtggtatcctgtacacgctcagagcccaatcaccaca
caggcaacgccacctccgacgagattcccagctgtatgtcagcctcggtggcgagaaacgaacccgttcg
tgtcctccggtttcgacggtatcgagaacattgtcctccgcccagagcgcctcttcaacgctcaattg
accgctcgcatctgagcaactgggatgtggtatgccagaactgggttatcccgaccatccaagacgg
tgtgggttggaagtgtctcgcgaagctgctctcagcgcgaagttggaataa

도면37b

서열 번호 70

푸사리움 옥시스트롬 유래의 GH3 패밀리 β -글루코시다제, Fo3A의 단백질 서열

mklnwvaalsisgaatdgalavasevpgtlagvkntdaqkvvtrdtlahspphyppwmdpnaigweea
yakaknfvsqtllekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirlsdynsafpagttag
aswskslwyergllmgtfkgkgidialgpatgplgrtaaggrnwegftvdpythagamaeavkgiqdag
viacakhyaneghefrqsgevsrkyniseslssnlddktlhelaywpfadavragvsgvmcsynqinn
sygcqnsklngilkdemgfggfmsdwaaghtgaasavagldmsmpgdtafdsygsfwggnltlaving
tvpawrvddmalrimsaffkvgtvedlpdinffsswtrdtfgfvqtfaqenreqvngvngvqhdknhr
esaakgsvilkntgsiplnnpkflavigedagpnpagpncgdrgcdngtlamawsgtsqfpylitpdq
glqnraaqdgtryesiltnewaqtqalvsqpnvtaivfanadsggyievdgnfgdrknltlwqqgdel
iknvssicpntivvlhtvgpvladyeknpnitaivwaglpqgesgnaiadllygkvspgrspftwgrtr
esygtevlyeanngrgapqddffsegvfidyrrhfdrrspstdgksapnntaaplyefghlswttfeysd
niqkvnstysppaggtipaptfgnfsklnldyvfpgkvryiykfiypflntssasaseandggqfgkta
eeflppnalngsaqprlpssgapgnpqlwdilytvtatitntgnatsdeipqlyvslggenepvrvlrg
fdrieniapggsaifnaqlttrdlsnwdvdaqnwvitdhpkvwwgsssrklplsakle

도면38a

서열 번호 71

지배텔라 제아에 유래의 GH3 패밀리 β -글루코시다제, Gz3A의 뉴클레오타이드 서열

ATGAAGGCCAATTGGCTTGCCGCGGCCGCTTTATTTGGCTGCTGGCACCGATGCTGCAGTCCCTGACACTT
TGGCAGGAGTCAATGTAAGCTACTCTTCAATTTCACTCATCTCAACTTTGCCAGGCCACAACAATTTT
CTTCACCTACCGATCTTTTCCACCAATAACGCAACAGTTTCACAAAAATAAAGCCCAATCATGCTCTGA
TCGTTGAATCGCCATCTTCGTTTACATCGCGGTGTCTTTTCTTCTGTACTTCTCATTCGTTGTGT
TCTCTACATTTTCGACTGGCTGTTTAGCCTTGAGATTCTTCTCACTCCCGTGATGCCTAGATCACTCTC
TGAGGCGTTTAACTACTTGTAGAGATGCGCCTCTCATTTGTGTGTGCTAGTCGCGATAGTTGCTGGA
ATTGCAGTCCTTGATCTTCTACTGACACTCAAAAGCTCGTTGCGCGGGACACACTCGCTCACTCTCCTC
CTCACTATCCCTCGCATGGATGGAGCCCTAACGCTGTGCGCTGGGAGGACGCTACGCCAAGGCCAAGGA
CTTTGTCTCCAGATGACTCTCTAGAAAAGGTCAACTTGACCACTGGTGTGGTAAGTAACGAGCGAC
AAGACGCTACAATCCACTAACACGATCTCTAGATGGCAGGCGAACGTTGTGTGGAAACGTGGATCT
ATCCCTCGTCTCGGTATGCGAGGCTCTGTCTCCAGGATGGTCTCTCGGAATTCGCTTCTCCGACTACA
ACAGCGCTTTCCCTACTGGTGTACCGCTGGTGTCTTCTGGAGTAAGGCCCTTTGGTACGAGCGAGGACG
ATTGATGGGTACCGAGTTTAAAGGAGAAGGGTATCGATATTGCTCTCGGCCCTGCAACTGGTCTCTCGGT
CGCCACGCTGCTGGTGGACGAACTGGGAAGGCTTCACTGTGCGACCCCTACGCCGCTGGCCATGCTATGG
CTGAGACTGTCAAGGGTATCAAGATTCTGGAGTCATTGCTTGTGCTAAGCATTACATCGCAACGAGCA
AGGTATGTACAGGCCATTCAATGGCTTCAGGAACGAAACTTAACCTTAATAGAACACTTCCGTCAACG
AGCGATGTATGTCTCAAAAGTTCAACATTTCCGAGTCTCTGTCTTCCAACCTTGACGATAAGACTATG
CAGGAGCTTACAACCTGGCCTTTCCGCGACGCGCTCGCGCGCGGTGTTGGCTCCATTATGTCTCTTACA
ACCAGGTCAACAACCTCATATGCTTGGCAGAACTCCAAGCTCTCAACGGCATCTTCAAGGACGAGATGGG
TTTCCAGGGTTTCGTCATGAGCGATTGGCAGGCTCAGCACACCGGTGCGGCTCCGCTGTTGCGGCTCTT
GACATGACCATGCCTGGTGACACCGAGTTCAACACTGGCTTCAGCTTCTGGGGTGGAAACCTGACCCCTG
CTGTTATCAACGGTACTGTTCCCGCTGGAGAATCGACGACATGGCTACCCGAATATGGCTGCTTTCTT
CAAGGTTGGCCGATCTGTTGAGGAGGAACCCGACATCAACTTCTCAGCTTGGACTCGTGATGAGTATGGC
TTGCTCCAGACCTACGCCCAAGAGAACCGAGAAAGGTCAACTTGTCTGTTAATGTCCAGCAGCAGCACA
AGCGCCACATTCGCGAGGCTGGCGCAAAGGGATCCGTCGTCCTCAAGAACCTGGCTCACTTCTCTTAA
GAAGCCCCAGTTCTCTGCTGTCTTGGAGAGGACGCTGGTTCCAACCTGCGGACCCAAACGGTTGCGCT
GACCGTGGATGCGACACCGTACTTGGCATGGCATGGGTTCCGGAACCTCTCAATTCCCTTACCTTG
TCACCCCGACCAAGGCATCTCGCTCCAGGCTATTAGGACGGTACTCGTTATGAGAGCATCTCAACAA
CAACAGTGGCCCCAGACACAAGCTCTTGTGAGCCAGCCCAACGTCACCGCCATTGTCTTTGCCAATGCC
GATTCTGGTGAGGGCTACATCGAGGTTGACGGCAACTACGGCGACCGCAAGAACCTCACTCTGTGGAAGC
AAGGCGATGAGCTCATCAAGAAGCTCTGCTATCTGCCCCAACACCATTTGTTGCTCTTACACCGTTGG
CCCCGTCCTTCAACCGAGTGGCACAACAACCCCAACATCACCGCATTTGTTGGGCTGGTGTGCTGGA
CAGGAGTCCGGTAACGCCATCGCGACATCTCTACGGCAAGACCGCTGGACGTTCTCCCTTACCT
GGGTCGCACTTATGACAGCTATGGCACAAGGTTCTCTACAAGGCCAACAATGGAGAGGGTGGCCCTCA
AGAGACTTTTGGCAGGGCACTTTCATGACTACCGCCACTTTGACCGACAATCCCCAGCAGCAACGGA
AAGAGTGCCACCAACGACTCTTCTGCTCTCTACGAGTTCGGTTTCGGTCTGTCTTGGACTACCTTTG
AGTACTCTGATCTCAAGTCGAGTCTGTGAGCAACGCTCTTACAGCCCCCTGTGCGGAACACCATTC
TGCCCCCTACCTACGGCAACTTCAGCAAGAACCTGGACGATTACACATTTCCCTCAGGTGTCCGATACCTC
TACAAGTTCACTACCCCTACCTCAACACCTCTTCTCCGCTGAGAAGGCTTCCGGCGATGTCAAGGGCA
GATTTGGTGAGACCCGCGACGAGTTCTCCCTCCCAACGCTCTCAACGGTTTCATCGCAGCCTCTCTTCC
TTCCAGTGGTGTCTCCGCGGTAACCTCAGCTCTGGGACATTATGTACACCGTCACTGCCACCATCACC
AACACTGGTGACGCTACCTCGGATGAGGTTCCCGAGCTGTACGTACGCTCGGTGGTGGGGGAGGCTG
TCCGTGTCTCCGTGGCTTCGAGCGTCTTGAAAACATTTGCTCTGGTGAGAGTGCACATTCACCGCTCA
GCTTACTCGCGTGACCTGAGCAACTGGGACGTCAACGTCCAGAAGTGGGTCATCACCGATCACGCCAAG
AAGATCTGGGTGCGGACGAGCTCTCGCAATCTGCCCTCAGCGCCGACCTGTAG

도면38b

서열 번호 72

지베렐라 제아에 유래의 GH3 패밀리 β -글루코시다제, Gz3A의 단백질 서열

mkanwlaaavylaagtdaavpdtlagvnlvardtlaahspphyppspwmdpnavgwedayakakdfvsqmtl
lekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirfsdynsafptgvtagaswskalwyergr
lmgtfkekgidialgpatgplgrhaaggrnwegftvdpyaaghamaetvkgiqdsgviacakhyaneg
ehfrqrgdvmsqkfniseslssnlddkthelynwfpadavragvgsimcsynqvnnsyacqnsklngi
lkdemgfqgfvmssdwqaghtgaasavagldmtmpgdtefntgfsfwggnlflavingtvpawridmatr
imaaffkvgrsvveepdinfawtrdeygvqtyaqenrekvnfavnvqhdkhrhireagakgsvvlknt
gsiplkpkqflavigedagsnpagpngcadrgcdngtlamawsgtsqfpylvtpdqgislaiaqdgttry
esilnnnqwpqtgalvsqpntaivfanadsgegyievdgnygdrknltlwkqgdeliknvsaicpnitiv
vlhtvgpvlletewhnpnitaiwagvpggesgnaiadilygktsprspftwgrtydsygtkvlykann
gegagqedfvegnfidyrhfdrgspstngksatndssaplyefgfglswtffeysdlkvesvsnasysps
vgntipartygnfsknllddytfspsvrylykfiypylntsssaekasgdvkgfrfgetgdeflppnalngs
sqprlpssgagpgnpqlwdimytvtatitntgdatsdevpqlvyslggegepvrvlrgferleniapges
atftaqltrrdlsnwdvvnvqnwvitdhakkiwvgsssrnlplsadl

도면39a

서열 번호 73

넥트리아 헤마토코카 유래의 GH3 패밀리 β -글루코시다제, Nh3A의 뉴클레오티드 서열

atgcgggttcacgcgtctctctgcggcatttttcggggcttgccccatggttggttcgcaagctgaccaga
aaccactacagctcgggtgtgaacaataaactctggcgcatcactcctcactataccttcgcatggat
ggatcctgctgctcctggctgggaggaagcctatctcaaggcgaaagattttgtttcacagcttacccct
cttgaaaaggtcaacttgaccactggtgttggtgagtcacttggtttctctctcctgacgtgacactt
tgctttggcctgcttctctatctgctctactagcattgctaactcagagcagatggatggcggaacgtt
cgctcggaacagctgggttcactcctcctggttttggaatgcgtggtctctgcatgcagatggccccctcg
catccgcttgctgactataaactctgcctttctactggtattacagctggtgctcttgagccgtgccc
ctttggtaccaacgtggcctcctgatgggcaccgagcatcgtgaaaaagcagatcgacgttgacttgggc
ctgctactggtcctcttggtcgtactcctactggcgccgcaactgggaggggttctcggttgatcccta
cgttgctggcgttgccatggccgagactgttagcggcattcaagatgggtggtactatcgccctgtgctaag
cactacatcggaacgaacaaggtatgctctctcacttctcctcgtgataaactcgtccacaacaacct
agagcaccatcgccaaagcccccgaatccattggccgcggtacaacatcaccgagtcctcgtgctgaac
gttgatgacaagacccctccagagctctatctctggccggttcgcagatgcggtcaaggctggtgttggtg
ctatcatgtgttctaccagcagctgaacaactcttaacggttgccaaaactctaagcttctcaacggaat
tctcaaggacagctaggtattccagggtctcgtcatgagtgactggcaagcccaacatgctggagctgct
accgctgttgacaggccttgacatgaccatgcccgggtgacactttgttcaaacccggatacagcttctggg
gtggtaacctgaccctcgtgtgtagtcaatggcactgttccgactggcgtattgacgacatggotatgag
aatcatggcagcttctctcaagggttggaagactgttgaggacctcctgacatcaactttctcttggtg
tctcgagacacttttggtcactgttcaagccgctgcccagaagaactgggaacagatcaacttcggagttg
atgttctgacagaccagcgaacacattcgaactctcggccgccaagggcaccgctcctccttaagaactc
tggtcatttgctctgaagaagcccaagttccttgccgtcgttgccgagagcgcggccccaacccctgct
ggccccaacggctgttaacgaccggatgttaacaacggcactctggccatgtcctggggtcaggaacag
cccagttcccttacctcgttactccgactcagcgtacagaaccaggctgtcctcgacggcactcgtcga
cgagagtgtcttgccgaacaaccagtggaacagacacgcagctctcattagccaacctaacgtgacggct
attgtgtttgccaatgccaatccggagagggatatacgtatgttgacggcaacgaaggcgatccggaaga
atttgacctgttggaacgaggggtgatgacctaatgaagaacgtctcctcaatctgccccaacaccattgt
tgttctgcacactgttgccctgtcactcctgacggaatggtatgacaaccgaacattaccgcatagtgtg
tggtgctggtgtacctggacaggagtcgggcaatgctcttggtgacatcctttatggcaaaaacaagccctg
gtcgtctcctctacatggggctgcacccgaagaggttacggcactgagtcctatacagagcccaacaa
tggtcaggggtgctcctcaagatgatttcaaggaggaggtctttatcgactatcgtcattttgaccagggt
tctcctagcaccagcggcagcaagtctaatgatgagtcagtcctcctacaggtttggccatggctctgt
cctggaccacgtttgagtactctgaactcaacattcaagctcacaacaagattccttcgatcctcctat
tgccgagacgattgcccgtccggctccttggaactacgtacgcaccttgccgattacaggttccccgat
ggaattcgtacatctaccagttcactcctcctggttgaaacttctcctccggaagagaggtctctg
gcgatcccgactacggaagacggccgaagaggttctgccccccggagctctcgacgggtcagctcagcc
gcgacctccatcctctggtgctccagggtggaacccctcatctttgggatgtgtgtgacactgttagtgt
atcatcaccacactggcaacggccactcggacgagatcccgagctctacgttagtctcgttggtggcaga
acgagcccgctccgctccttcgggggttcgacgaattgagaacattgcccctggccagaggtgtcagatt
cacaactgacatcactcggcgacacctgagcaactgggacgtcgtctctcagaactgggtcattacagac
tacgagaagaccgtatattgctgggagcagctcccgcaacctgcctctcaaggcaacctgaagtaa

도면39b

서열 번호 74

넥트리아 해마토코카 유래의 GH3 패밀리 β -글루코시다제, Nh3A의 단백질 서열

mrftvllaaafsglvpmvgsqadqkplqlgvnnntlahspphyppswmdpaapgweeaylkakdfvsqtl1
lekvnlttgvgwmgercvgnvgsiprfmrglcmqdgplgirlsdynsafptgitagaswsralwyqrgl
lmgtehrekgidvalgpapglgrtptggrnwegfsvdpyvavvamaetvsgiqdgggiacakhyneg
ehhrqapesigrniteslssnvddkthelylwpfadavkagvgaimcsyqqlnnsygcqnsklngi
lkdelgfgqfvmstdwqahagaatavagldmtmpgdtlfnfgysfwggnitlavvngtvpdwriddmamr
imaaffkvgtvedlpdinfsswsrdtfgyvqaaqenweqinfvdrhdhsehirlsaakgtvilkns
gslplkpkflavvgedagpnpagpncndrgcnngtlamswsgtaqfpylvtppsalsalngavidgtry
esvlrnnqweqtrslisqpnvtaivfanansgegyidvdgnegdrknltlwnegddliknvssicpntiv
vlhtvgpviltewydnpnitaiwagvpggesgnalvdilygktsprgrspftwgrtrksygtdivlyepnn
gggapqddftegvfidyrhfdqvspstdgskndesspiyefghglswttfeyselniqahnkipfdppi
getiaapvlgnystldadytfdpdiriyqfiypwlnntsssgreasgdpdyktaeeflppgaldgsaqp
rppssgapggnphlwdvlytsaiitntgnatsdeipqlyvslggenepvrvlrgfdrieniapqgsrvf
ttditrrdlsnwdvvsqnwvitdyektvyvgsssrnlplk

도면40a

서열 번호 75

버티실리움 달리아에 유래의 GH3 패밀리 β -글루코시다제, Vd3A의 뉴클레오타이드 서열

ATGAAGCTGACCCCTCGCTACTGCCCTTACTGGCAGCCAGCGGGTGTGTCTCTGCGGACAACCCAAGCTCA
AGGTACGTACTTGGCTCTTTTTTCAACAAGGAAACCAACCCGACCCATAATGGTGATTGAGCAGTCTGTGCT
TTCTTCAACCCGAATCAAACCCATGCCGTGTTCGCGCATGCCCTTTCGATCGTCTGTTGTGTGAACCC
ACGCTCTTCAAGCATCGCACATAGCACCACTCCATCTTCAATTTTCGAGCAATTCGGGCCGCAGAGAGCG
GTCTTTCACTTCAACCAATCGTTTCATGCCCTCGTCCCCACTGCCATGTTTCTTCCCAGTATTCTACTTC
TGAGAGCCTTGACCACCGTGTGTCGACATCTCGTCGCCAAGGCTCGTTGACACGGAATCTGTTCCCTTGG
AATTAATATTGAAACAATGCTGACCAGCATCTCAGCGCCAGACTAACAGCTCTAGCGAGCTCGCCTTT
TCCCTCCGCACTACCTTCTCCATGGATGAACCCCAAGCGACTGGGTGGGAGGACGCTACGCCCGTG
CCAGAGAGGTGGTAGAGCAGATGACTCTGCTCGAAAAGGTCAACCTGACGACAGGTGTGCGGTAAGCTTC
ACAGACCCCGTCTTGCCATCCAAAGTCATCTGACAGAATCCTAGCTGGAGCGGTGATCTCTGCGTCGGAA
ACGTGCGCTCGATCCCCGAATCGGCTGGAGGGGGCTTTGTTTGAGGATGGCCACAGGGTATCCGTTT
CGCGGACTACGTCTCTGATCTTCACTTTCGAGCCAGACAGCCGCGCTACCTGGGACCGAGGGCTTCTGTAC
CAGCGCGCTCACGCCATTGGCGCGAAGGAGTAGCCAAGGGCGTCGACGTCGCTCTCGGGCCCGCATTG
GCCCTCTAGGTGCGCTTCCCGCCGAGGTGCTAACTGGGAGGGTTTCGCGGTGGACCTTACCTCAGTGG
CGTTGCTGTGCGCGAATCCGTGAGGGGCATCCAGGATGCTGGTGTCTATTGCCAACGTCAAGCACTACATC
GTCAATGAGCAGGAACATTTCCCGCAGGCTGGCGAGGCTCAAGGTTACGGCTACGATGTGACGAGGCAT
TATCGTCGAACGTTGACGACAAGACCATGCGATGAGCTTTACCTTTGGCCATTTGACAGCGCTGTCCGTGC
TGGAGCCGCGAGTGTCTGTTCTTATCAACAGGTGGGGCAATACCATTTCTCTCTCTTTCTCTGACG
ACAGTGCACTGACCGACCTTTTTCGCCAAGATCAACAACAGTTACGGCTGTCAAAACTCACATCTCTG
AATGGGCTCTCAAGGACGAACCTCGGCTTTCAGGGGTTCGTCCTCAGCGATTGGCAAGCGCAGCATGCTG
GTGCTGCCACTGCCGTGTGTCGACTTGCATGCGCATGCCCGGTGACACTCGCTTCAACACCGGAGTCGC
CTTCTGGGGCGCTAACCTTACCAATGCCATTTTGAACGGCACCGTTCCCGAATATCGGCTCGATGACATG
GCCATCGGTATTATGGCGCGCTTTTTCAGAGTTGGAAGACCTGGACGATGTTCTGACATCACTTCT
CGTCTTGGACAAAGACACCATCGGCGCGCTGCACTGGGCGGCCAGGACAATGTGACGTCATCAACCA
ACACGTTGATGTCGTCAGAACACCGGCGCCCTCATTCGCACCATCGCTGCCCGGCTACTGTCTACTA
AAAAATGAGGATCACTGCCCTCTGAACAGCCGAAATTTGTGTGTGTCATGTTGGAAGATGCTGGCCCTC
GTCTGTGTTGGTCCCAATGCTGCTGATCAGGGTTGCAATAACGGCACCTTGGCTGCTGGATGGGGATC
TGGCACCGCCAGTTTCCCTTATCTCACTCCTGATAGTGTCTTTCAGTTTCAAGCCGTTTCGGATGGC
TCGCGATACGAAAGCATCTCAGCAACTGGGATTATGAGCGCACAGAGGCTTGGTTTCCAGGGCGGATG
CTACTGCTCTGGTTTTCTGTCATGCAAACTCTGGCGAAGGATATATCAGCGTTGATGGAAACGAAGGTGA
TCGCAAGAACCTCACTCTCTGGAATGGAGGAGACGAGCTTATTCACAGAGTCGCTGCGGCCAACAACAAC
ACCATCGTCATCATCATTCGGTTGGTCCCCTTCTAGTCACTGACTGGTACGAGAATCCCAATATCACGG
CTATCATCTGGGCGCGCTTACCCGACAGGAGTCTGGCAACTCTATCGCCGATATTCTTTACGGCCGCGT
GAACCCGTGGTGGCAAGACACCTTTACCTGGGGTCCAAGTGTGAGAGCTACGGCGTTGACGTCCTGAGA
GAGCCCAACAATGGCAATGTTGCTCCCCAGAGGATTCGACGAGGGAGTCTTACATCGATTACCGTTGGT
TTGACCGGCACTCGGCTGTTGATAACAATGCATCAGCGCCGAGGAACAGCAGCAGCAGCCACGCCCAAT
CTTCGAGTTTGGCTATGGCCTTTTCGTACACAACCTTTGAATTTCTCCAATCTTCAGATTGAGAGGCATGAC
GTTACAGATTACGTCCTTACCACCTGGGACAGACGAGCCCTGCGCCGAGATTGGTGCTAACTACAGTACGA
ACTACGACGACTACGCTTTTCCCGAGGGCGAAATCCGTTACATCTATCAACACATCTACCCATACCTCAA
TTCTCAGACCCAAAGGAGGCAATTTGGCTGATCCTAAATACGGCCAAACTGCAGAAGAGTTCTTCCAGAG
GGCGCTCTTGATGCCCTACCGCAGCCTAGGCTCCCAGCTTCTGGAGGGCCCGGAGGCAACCCAATGCTTT
GGGACGTCATATTACGCTACCCGCGACCGGTGACCAACACGGGTAAAGTTGCTGGGACGAAGTGGCACA
GCTTTACGTTTCTTGGTGGACCTGACGATCCGATTTCGAGTCCCTCGGTGGGTTTCAGCCGATTCACATC
GCGCCTGGAGCCTCGCAAACTTCCGTGCGGAACCTCACGCGCCGGGACCTCAGCAACTGGGATGTTGTCA
CGCAAAATTTGTTTCATCAGCCAGTACGAAAGACGGTCTTTGTCGGGAGCTCATCCCGAAACCTCCCTCT
CAGCACTCGCCTCGAATAG

도면40b

서열 번호 76

버티실리움 달리아에 유래의 GH3 패밀리 β -글루코시다제, Vd3A의 단백질 서열

mkltlatalaasgcvsagqpkklkhpqrqtnssselafspphyppspwmpqatgwedayararevveqmt
llekvnlttgvgwsgdlcvngvgsiprigwrglclqdgppqgirfadyvvsyftssqtagatwdrgllyqra
haigaegvakgvdvvlgaipglgrlpaggrnwefavdpylsgvavaesvrgiqdagaiavkhyivne
qehfrqageaggygydvdealssnvddkthelylwpfadavragagsvmcsyqqinnsygcqnsiling
llkdelgfggfvlslsdwqaghagaatavagldmampgdtrfntgvafwganltnailngtvpeyrlddmam
rimaaffkvgtklldvdpinfsswtkdtigplhwaagdnvqvinqhvdvrqdhgalirtiaargtvlkn
egslplnkpkkfvavigedagrpvpngcpdqgcnnngtlaagwsgstasfpylitpdsalqfnavsdgsr
yesilsnwdyertealvsqadatalvfvnansgegyisvdgnegdrknltlwnngdeliqrvaannnti
viihsvgpvlvtdwyenpnitaiiwaglpqgesgnsiadilygrvnpqgktpftwqptvesygydvrep
nngngapqsddegfvidyrfwfdrgsgvnnasaprnssssshapifeffgyglsttfeffsnlqierhdvh
dyvpttggtspaprfganystnyddyvfpegeiriyqhiypylnssdpkealadpkygqtaeeflpega
ldaspqprlpasggpggnpmldwiftvtatvtntgkvagdevaqlvyslpgpddpirvlrgfdrihiap
gasqtfraeltrrdlsnwdvvtqnwfisgyektvfvgsrrnplstrle

도면41a

서열 번호 77

포도스포라 안세리나 유래의 GH3 패밀리 β -글루코시다제, Pa3G의 뉴클레오타이드 서열

ATGAAACTCAATAAGCCATTTCCTGGCCATTTATTTGGCTTTCAACTTGGCCGAGGCTTCGAAAACTCCGG
ATTGCATCAGTGGTCCGCTGGCAAAGACCTTGGCATGTGATACAACGGCGTCACCTCCTGCGCGAGCAGC
TGCTCTTGTGCAGGCTTTAAATATCACGAAAAAGCTTGTGAATCTAGTGGAGTATGTCAAGTCAAGAGAA
GCTCCTTTAGGGATTTCATTCAGCTAATCACTCCTCATAGCATGAGCCTCGGTGCAGAAAGGATCGGCC
TTCACGCTTATGCTTGGTGGAAACGAAGCTCTTCATGGTGTGGCCGCTCGCCTGGGGTCTCCTTCAATCA
GGCCGGACAAGAATTCTCACACGCTACTTCATTTGCGAATACTATTACGCTAGCAGCCGCCCTTGACAAT
GACCTGGTTTACGAGGTGGCGGATACCATCAGCACTGAAGCGCGAGCGTTTACGAATGCCGAGCTCGCTG
GACTGGATTACTGGACGCCTAACATCAACCCGTACAAAGATCCGAGATGGGGGAGGGGCCATGAGGTTTG
TTACCTTAGCCTTCTTTTCGCTGCCGTGCGAGTTGCTGAGAACTCAAAGACACCCGGAGAAGATCCGGTA
CACATCAAAGGCTACGTCCAAGCACTTCTCGAGGGTCTAGAAGGGAGAGACAAGATCAGAAAGGTGATTG
CCACTTGTAAACACTTTGCAGCCTATGATTTGGAGAGATGGCAAGGGGCTCTTAGATACAGGTTCAATGC
TGTGTGACCTTCGACGATCTTTTCGGAGTACTACCTCCAACCGTTTCAACAATGCGCTCGAGACAGCAAG
GTCGGGTCTTTCATGTGCTCATATAATGCGCTCAACGGAACACCGGCATGTGCAAGCACGTATTTGATGG
ACGACATCCTTCGAAAACTGGAATTTGAGCCGAGCACAACAATATATAACGAGCGACTGTAATGCTAT
TCAGGACTTCTCCTCCCACTTTCACAACCTTCAGCCAACTCCAGCTCAAGCCGCGCTGATGCTTATAAC
GCCGGTACAGACACCGTCTGTGAGGTGCTGGATACCCCCACTCACAGATGTAATCGGAGCATACAATC
AGTCTCTGCTGTGAGGAAATTTATCGACCGAGCACTTCGCAGATTATACGAAGGCCTTATCCGAGCTGG
CTATCTCGACTCAGCCTCCCACTACCATACACCAAAATCTCATGGTCCCAAGTAAACACCCCAAAGCC
CAAGCCCTGGCTCTCCAGTCCGCCACCGACGGGATAGTCTTCTCAAAAACAACGGCCTCCTTCCCTTAG
ACCTCACCAACAAAACCATAGCCCTCATAGGCCACTGGGCCAATGCAACCGCCAAATGCTAGGCGGCTA
CAGCGGTATCCCCCTTACTACGCCAACCCAACTCATATGCAGCCACCCAGCTCAACGTCACTTTTCATCAC
GCCCCAGGACCGGTGAACCACTCATCTCCCTCCACAAATGACACCTGGACCTCCCCCGCCCTCTCCGCGG
CTTCCAAATCGGATATCATCTCTTACCTCGGCGGCACCGACCTCTCCATCGCAGCCGAAGACCGAGACAG
AGACTCCATCGCCTGGCCATCCGCTCAACTTTCTTGTAACTCCCTCGCCAGATGGGAAAACCCACA
ATCGTAGCAAGACTAGGCGACCAAGTAGACGACACCCCCCTGCTCTCCAACCCAAACATCTCTCCATCC
TATGGGTAGGCTACCCAGGCCAATCAGGCGGAACAGCCCTCTTGAACATCATCACCAGGAGTCAAGCTCCCC
CGCCGCTCGACTGCCCGTCAAGTCTACCCAGAACTTACACCTCCCTCATCCCCCTGACAGCCATGTCC
CTCCGCCCAACCTCCGCCCGCCAGGCGGACTTACAGGTGGTACCCCTCCCCGTGCTCCCTTCCGCGC
ACGCCCTCCACTACACAACCTTTACCGCCAAATTCGGCGTCTTTGAGTCCCTCACCATCAACATTGCCGA
ACTCGTTTCCAACGTAAACGAACGATACCTCGACCTCTGCCGGTTCCCGCAGGTGTCCTGCTGGGTGTCG
AATACGGGAGAACTCAAATCTGACTATGTGCGCTTGTGTTTGTACAGGGGTGAGTACGGACCGGAGCCGT
ACCCGATCAAGACGCTGGTGGGTACAAGCGGATAAGGGATATCGAGCCGGGACTACGGGGGCGCGGCC
GGTGGGGGTGGTGGTGGGGATTTGGCTAGGGTGGATTTGGGGGGGAATAGGGTTTGTTCGGGGGAAG
TATGAGTTTCTGCTGATGTGGAGGGGGGAGGGATAGGGTTGTGATCGAGTTGGTTGGGGAGGAGGTGG
TGTTGGAGAAGTCCCTCAGCCGCTGCGCGGGTGA

도면41b

서열 번호 78

포도스포라 안세리나 유래의 GH3 패밀리 β -글루코시다제, Pa3G의 단백질 서열

mklnkpflaiylafnlaeaasktpdcisgplaktlacdttaspparaaalvqalniteklvnlveyvksre
aplgisiqlitphsmstlgaerigipayawwnealhgvaaspgvsnagagqefshatsfantitlaaafdn
dlvyevadtistearafснаaelagldywtpninpykdpwrwgrghevacylsllfravqllrtqktgpedpv
hikgyvqalleglegrdkirkviatckhfaaydlerrwqgalryfnavvtsqdlseyyllqpfgqccardv
vgstmcynalngtpacastylmddllrkhwntehnnysltdrcnaigldfllpnhrnsftqpaqaadaayn
agdtvtcevpgpypltdvigayngslseeiidralrrlyegiliragylldsasphpytkiswsvntpka
qalalqsatdgivllknngllpldltntktialighwanatrqmllggysgipppyanpiyaatqlnvtfhh
apgvnqsspstndtwtspalsaskdiilylggtdlsiaaedrdrdsiawpsaglslltsiaqmgkpt
ivarlgdqvddtpllssnpnissilwvgypqsggtallniitgvsspaarlptvtypetytslipltams
lrptsarprgrtyrwpssvpvlpfghlyhyttfktakgfvdesltniaaelvsncneryldlcrfpqgsvvws
ntgesksdyvalvyfvrgeygeepypkltlvtykardliefpgttgaapgvvvgpdlhrfdlgggnrvlfpgk
yeflldveggdrdrvvielvgpeevvlekfpqppaaq

도면42

서열 번호 79

써모토가 네아폴리타나 유래의 GH3 패밀리 β -글루코시다제, Tn3B의 단백질 서열

MEKVNEILSQLTLEEKVKLVGVGLPGLFGNPHSRVAGAAGETHPVPRVGLPAFVLADGFPAGLRINPTRE
NDENTYTYTAFPPVEIMLASTWNRELLLEEVGKAMGEEVREYGVVDVLLAPAMNIHRNPLCGRNFEYYSBDPV
LSGEMASSFVKGVQSQGVGACIKHFVANNQETNRMVVDITVSERALREIYLRGFEIAVKKSPKWSVMSAY
NKLNGKQYCSQNEWLLKKVLKEWGFEGFVMSDYGADNPVEQLKAGNDLIMPCKAYQVNTERRDEIEIM
EALKEGKLESEEVLDCEVRNLIKLVNAPSFKNRYSNPKDLEKHAKVAYEAGAEQVNVLLRNEEALPSEN
SKIALFGTGQIETIKGGTGSGDTHPRYAISILEGIKERGLNFDEELAKTYEDYIKKMRTEEYKPRRDSW
GTIIKPKLPENFLSEKEIHKLAKKNDAVIVISISGBGYDRKPKVGDFYLSDDETDLIKTVSREFHEQG
KKVIVLLNIGSPVEVSVWRDLVDGILLVWQAGQETGRIVADVITGRINPSGKLPPTFFPRDYSVPSPSTFF
GEPKDNQPKQVVEEDYVGYRYDYTFGVEPAYEFGYGLSYTTFEYSDLNVSFDDGETLRVQYRIENTGGRA
GKVSQVYIKAPKGIKDPPGELKAFHKTRLLNPNGESEEVLEIPVRDLASFNGEEWVEAGEYEVRVGA
SSRNLIKLGTFVSVEERRFKP

도면43a

도 43a-1	도 43a-2	도 43a-3	도 43a-4	도 43a-5
<p>본 발명의 실시예에 따른 도 43a-1은, 일련의 비트 스트림을 처리하는 장치의 블록 다이어그램이다. 이 장치는 입력 데이터 스트림을 수신하고, 이를 처리하여 출력 데이터 스트림을 생성한다. 주요 구성 요소는 다음과 같다:</p> <ul style="list-style-type: none"> 입력 데이터 스트림 (Input Data Stream): 데이터의 시작을 나타내는 비트 스트림이다. 처리 블록 (Processing Block): 입력 데이터를 처리하는 핵심 부분으로, 여러 하위 블록으로 구성된다. 출력 데이터 스트림 (Output Data Stream): 처리된 데이터의 결과물이다. <p>이 도면은 데이터 처리의 초기 단계를 보여준다.</p>	<p>본 발명의 실시예에 따른 도 43a-2는, 도 43a-1의 처리 블록을 더 상세히 설명한다. 이 블록은 여러 하위 구성 요소를 포함하고 있다:</p> <ul style="list-style-type: none"> 하위 블록 1 (Sub-block 1): 입력 데이터를 특정 방식으로 변환한다. 하위 블록 2 (Sub-block 2): 변환된 데이터를 다시 처리한다. 하위 블록 3 (Sub-block 3): 최종 처리 결과를 출력한다. <p>이 도면은 데이터 처리의 중간 단계를 보여준다.</p>	<p>본 발명의 실시예에 따른 도 43a-3은, 도 43a-2의 하위 블록을 더 상세히 설명한다. 이 블록은 여러 하위 구성 요소를 포함하고 있다:</p> <ul style="list-style-type: none"> 하위 블록 1 (Sub-block 1): 입력 데이터를 특정 방식으로 변환한다. 하위 블록 2 (Sub-block 2): 변환된 데이터를 다시 처리한다. 하위 블록 3 (Sub-block 3): 최종 처리 결과를 출력한다. <p>이 도면은 데이터 처리의 후기 단계를 보여준다.</p>	<p>본 발명의 실시예에 따른 도 43a-4는, 도 43a-3의 하위 블록을 더 상세히 설명한다. 이 블록은 여러 하위 구성 요소를 포함하고 있다:</p> <ul style="list-style-type: none"> 하위 블록 1 (Sub-block 1): 입력 데이터를 특정 방식으로 변환한다. 하위 블록 2 (Sub-block 2): 변환된 데이터를 다시 처리한다. 하위 블록 3 (Sub-block 3): 최종 처리 결과를 출력한다. <p>이 도면은 데이터 처리의 후기 단계를 보여준다.</p>	<p>본 발명의 실시예에 따른 도 43a-5는, 도 43a-4의 하위 블록을 더 상세히 설명한다. 이 블록은 여러 하위 구성 요소를 포함하고 있다:</p> <ul style="list-style-type: none"> 하위 블록 1 (Sub-block 1): 입력 데이터를 특정 방식으로 변환한다. 하위 블록 2 (Sub-block 2): 변환된 데이터를 다시 처리한다. 하위 블록 3 (Sub-block 3): 최종 처리 결과를 출력한다. <p>이 도면은 데이터 처리의 후기 단계를 보여준다.</p>

도면43b

도 43b-1

7a3a	-----MKGILLVAVLAA-----VANNGKAT	24
7c3b	-----MTLVFAALLAAVAAVPP-----PNSQAT	30
Fv3c	-----MELMVAALLGAAUTGVAALAAVPPVLAASVAAQGVVDTUAT	50
Fv3c/7c3b	MELMVAALLGAAUTGVAALAAVPPVLAASVAAQGVVDTUAT	50
Fv3c/7a3a/7c3b	MELMVAALLGAAUTGVAALAAVPPVLAASVAAQGVVDTUAT	50
7a3a	SPFFPTFNNWQDQ-DMASAVAVAPPSQLTIAEKVLTITVQMGQDC	73
7c3b	SPFFPTFNNWQDQ-DMASAVAVAPPSQLTIAEKVLTITVQMGQDC	80
Fv3c	SPFFPTFNNWQDQ-DMASAVAVAPPSQLTIAEKVLTITVQMGQDC	100
Fv3c/7c3b	SPFFPTFNNWQDQ-DMASAVAVAPPSQLTIAEKVLTITVQMGQDC	100
Fv3c/7a3a/7c3b	SPFFPTFNNWQDQ-DMASAVAVAPPSQLTIAEKVLTITVQMGQDC	100
7a3a	VQGVSEIPLGPGQLNQDSFSGTGTWRSAPVACVAAVAAVAAVAAV	123
7c3b	VQGVSEIPLGPGQLNQDSFSGTGTWRSAPVACVAAVAAVAAVAAV	130
Fv3c	VQGVSEIPLGPGQLNQDSFSGTGTWRSAPVACVAAVAAVAAVAAV	150
Fv3c/7c3b	VQGVSEIPLGPGQLNQDSFSGTGTWRSAPVACVAAVAAVAAVAAV	150
Fv3c/7a3a/7c3b	VQGVSEIPLGPGQLNQDSFSGTGTWRSAPVACVAAVAAVAAVAAV	150
7a3a	ROVAKNEERERQGVVQVGLGPGVAGVAGVAGVAGVAGVAGVAGV	173
7c3b	ROVAKNEERERQGVVQVGLGPGVAGVAGVAGVAGVAGVAGVAGV	180
Fv3c	ROVAKNEERERQGVVQVGLGPGVAGVAGVAGVAGVAGVAGVAGV	200
Fv3c/7c3b	ROVAKNEERERQGVVQVGLGPGVAGVAGVAGVAGVAGVAGVAGV	200
Fv3c/7a3a/7c3b	ROVAKNEERERQGVVQVGLGPGVAGVAGVAGVAGVAGVAGVAGV	200
7a3a	SGLMETFEKRSIGLALGATGATGATGATGATGATGATGATGATG	220
7c3b	SGLMETFEKRSIGLALGATGATGATGATGATGATGATGATGATG	230
Fv3c	SGLMETFEKRSIGLALGATGATGATGATGATGATGATGATGATG	250
Fv3c/7c3b	SGLMETFEKRSIGLALGATGATGATGATGATGATGATGATGATG	250
Fv3c/7a3a/7c3b	SGLMETFEKRSIGLALGATGATGATGATGATGATGATGATGATG	250
7a3a	STTGTGAGVACAKHPTLVQGRFPGAGQ---GSDVIAEISAGAGQ	270
7c3b	STTGTGAGVACAKHPTLVQGRFPGAGQ---GSDVIAEISAGAGQ	280
Fv3c	STTGTGAGVACAKHPTLVQGRFPGAGQ---GSDVIAEISAGAGQ	300
Fv3c/7c3b	STTGTGAGVACAKHPTLVQGRFPGAGQ---GSDVIAEISAGAGQ	300
Fv3c/7a3a/7c3b	STTGTGAGVACAKHPTLVQGRFPGAGQ---GSDVIAEISAGAGQ	300
7a3a	7TIEVTPQVQVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAV	320
7c3b	7TIEVTPQVQVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAV	330
Fv3c	7TIEVTPQVQVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAV	350
Fv3c/7c3b	7TIEVTPQVQVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAV	350
Fv3c/7a3a/7c3b	7TIEVTPQVQVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAVAV	350
7a3a	QSPNEMAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV	370
7c3b	QSPNEMAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV	380
Fv3c	QSPNEMAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV	400
Fv3c/7c3b	QSPNEMAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV	400
Fv3c/7a3a/7c3b	QSPNEMAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV	400

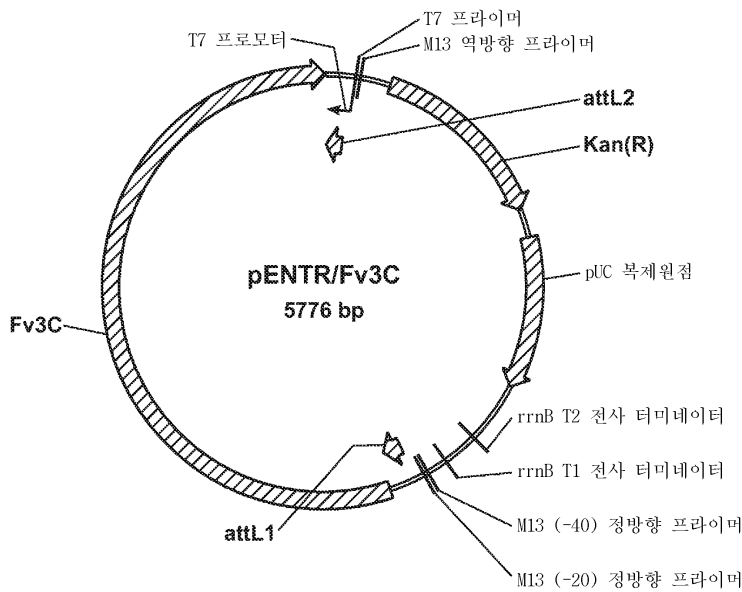
도 43b-2

7a3a	STREAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	360
7c3b	STREAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	370
Fv3c	STREAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	390
Fv3c/7c3b	STREAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	390
Fv3c/7a3a/7c3b	STREAVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	390
7a3a	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	410
7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	420
Fv3c	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	440
Fv3c/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	440
Fv3c/7a3a/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	440
7a3a	DAGNFWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	460
7c3b	DAGNFWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	470
Fv3c	DAGNFWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	490
Fv3c/7c3b	DAGNFWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	490
Fv3c/7a3a/7c3b	DAGNFWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	490
7a3a	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	510
7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	520
Fv3c	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	540
Fv3c/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	540
Fv3c/7a3a/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	540
7a3a	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	560
7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	570
Fv3c	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	590
Fv3c/7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	590
Fv3c/7a3a/7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	590
7a3a	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	610
7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	620
Fv3c	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	640
Fv3c/7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	640
Fv3c/7a3a/7c3b	MTLQAGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	640
7a3a	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	660
7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	670
Fv3c	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	690
Fv3c/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	690
Fv3c/7a3a/7c3b	QVTKNEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ	690

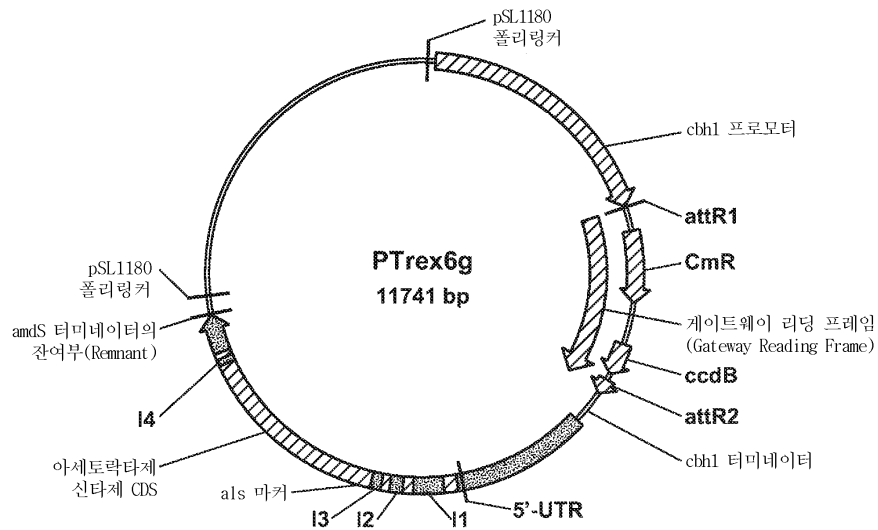
도 43b-3

7a3a	LVQ	700
7c3b	LVQ	710
Fv3c	LVQ	730
Fv3c/7c3b	LVQ	730
Fv3c/7a3a/7c3b	LVQ	730
7a3a	LVQ	750
7c3b	LVQ	760
Fv3c	LVQ	780
Fv3c/7c3b	LVQ	780
Fv3c/7a3a/7c3b	LVQ	780
7a3a	LVQ	800
7c3b	LVQ	810
Fv3c	LVQ	830
Fv3c/7c3b	LVQ	830
Fv3c/7a3a/7c3b	LVQ	830
7a3a	LVQ	850
7c3b	LVQ	860
Fv3c	LVQ	880
Fv3c/7c3b	LVQ	880
Fv3c/7a3a/7c3b	LVQ	880
7a3a	LVQ	900
7c3b	LVQ	910
Fv3c	LVQ	930
Fv3c/7c3b	LVQ	930
Fv3c/7a3a/7c3b	LVQ	930

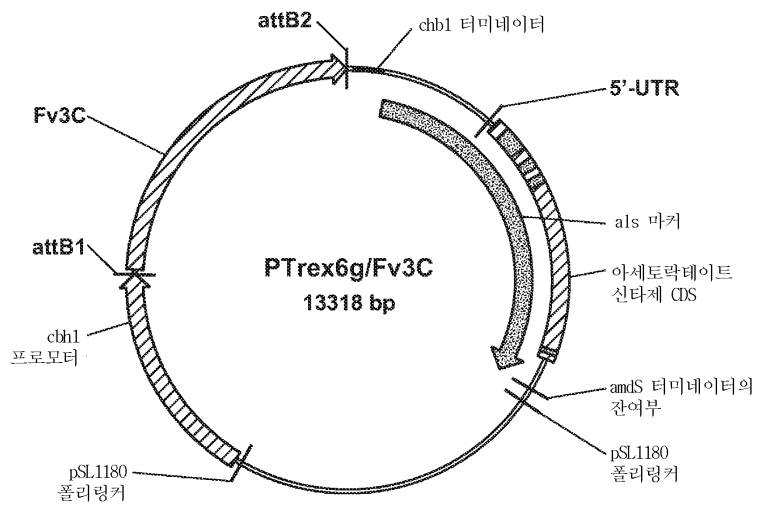
도면44



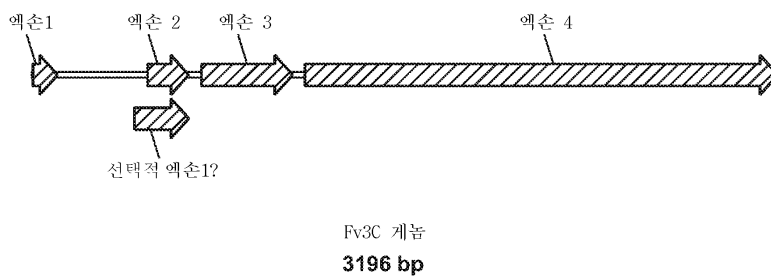
도면45a



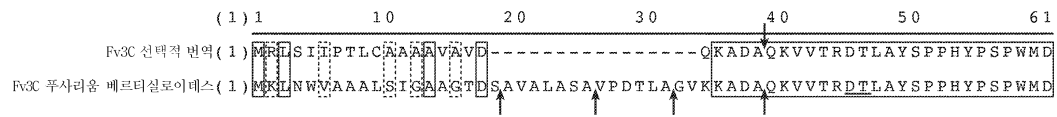
도면45b



도면46a



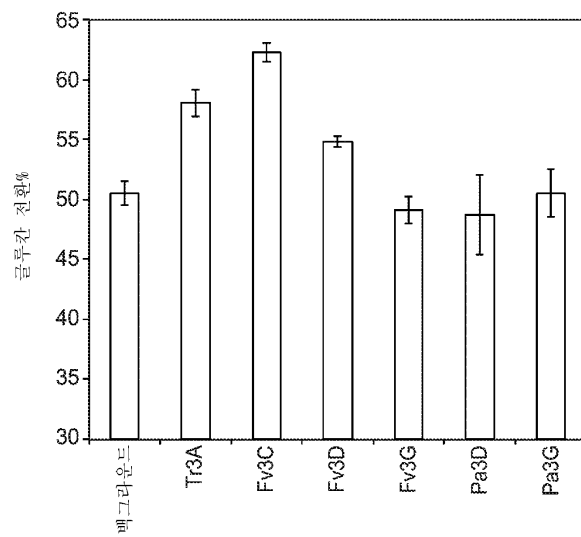
도면46b



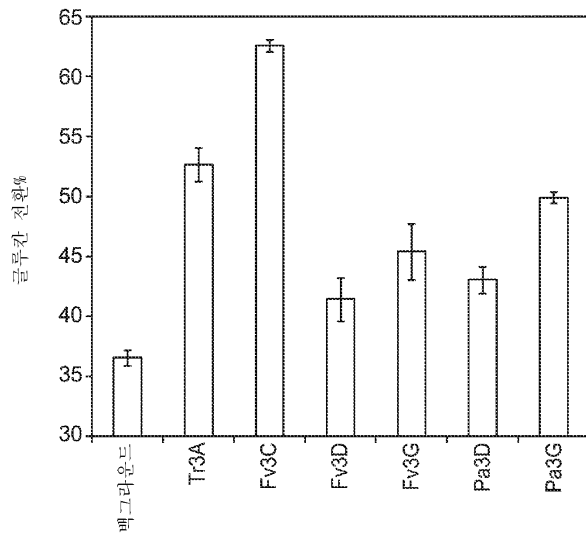
도면46c



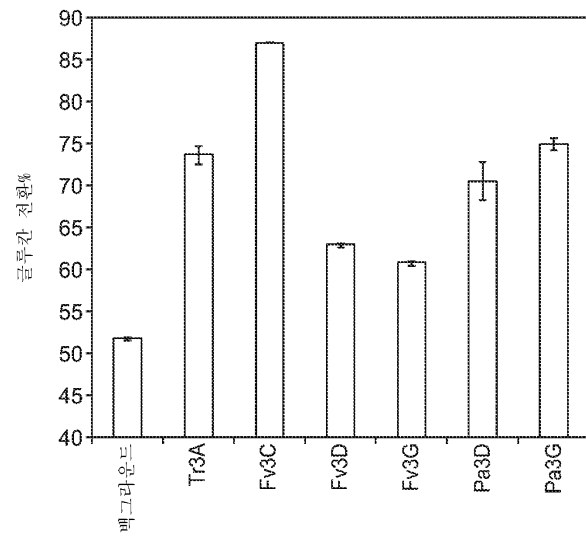
도면47



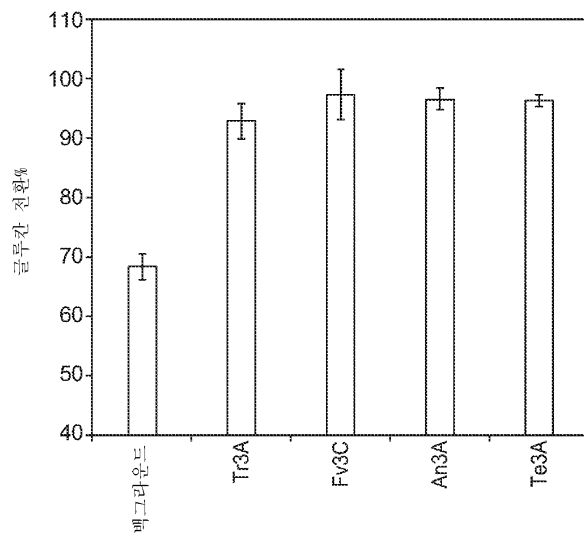
도면48



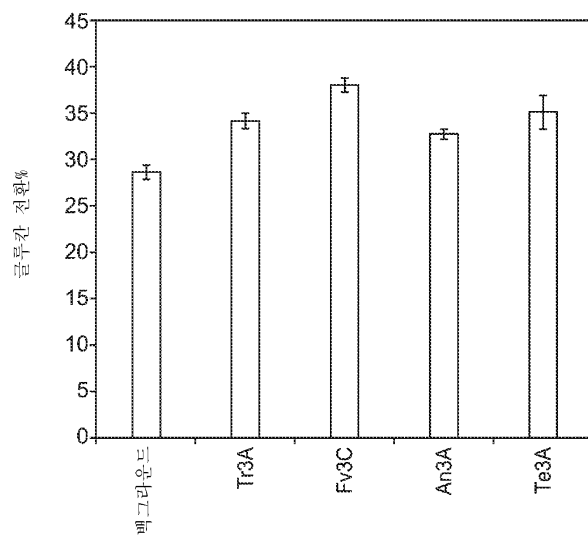
도면49



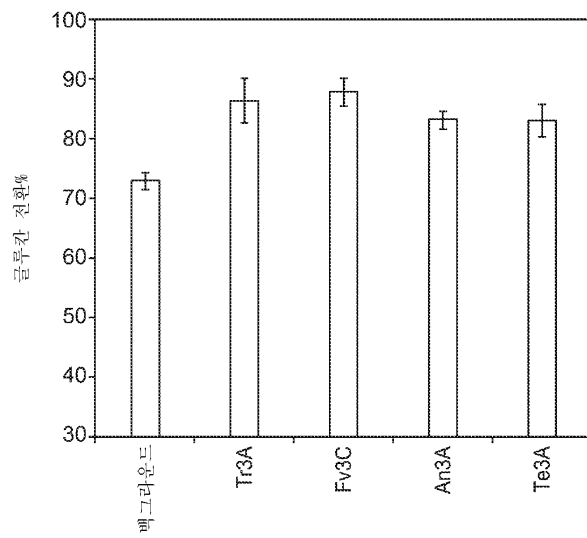
도면50



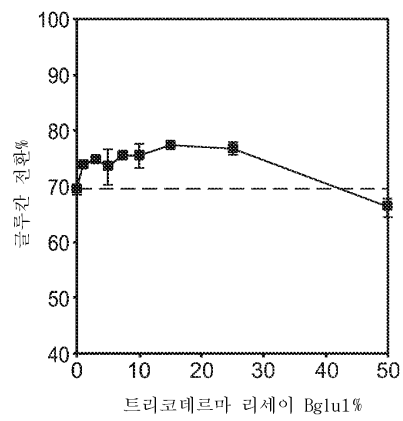
도면51



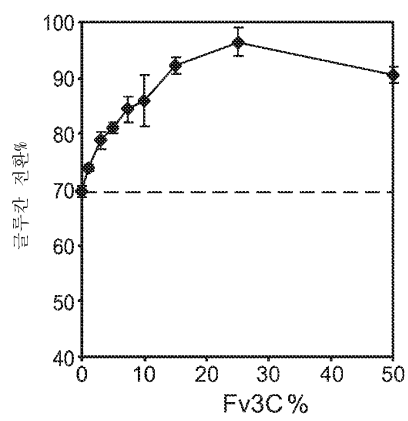
도면52



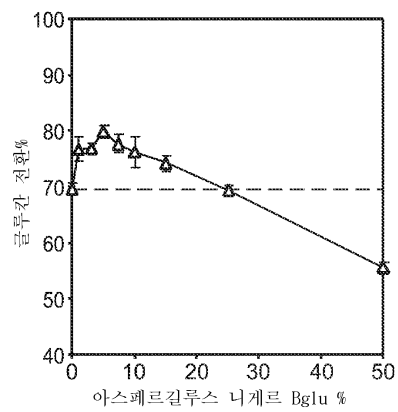
도면53a



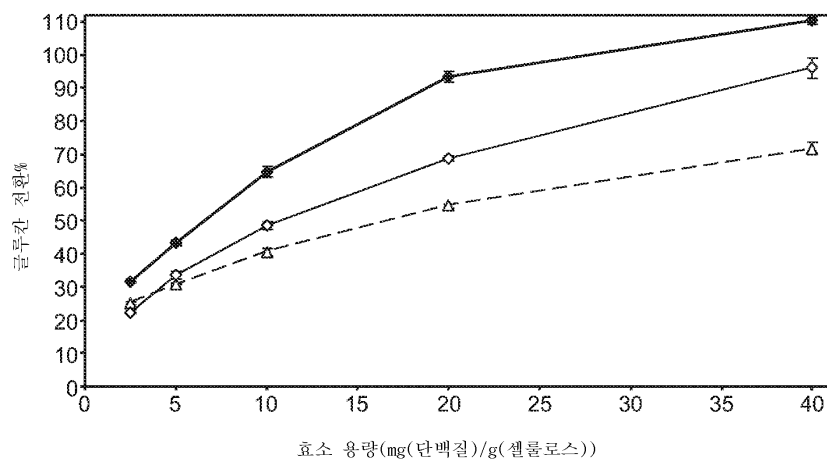
도면53b



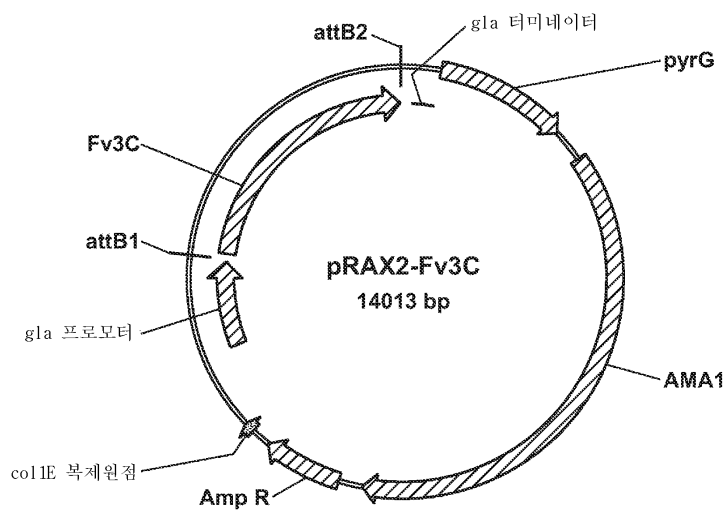
도면53c



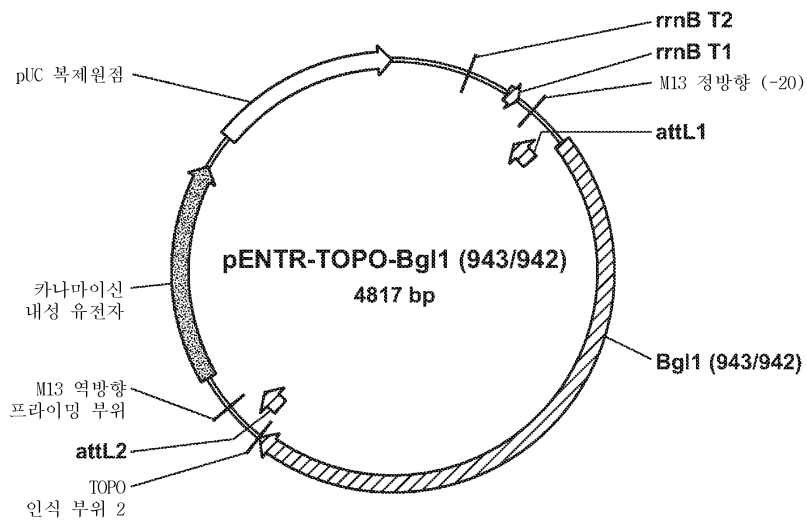
도면54



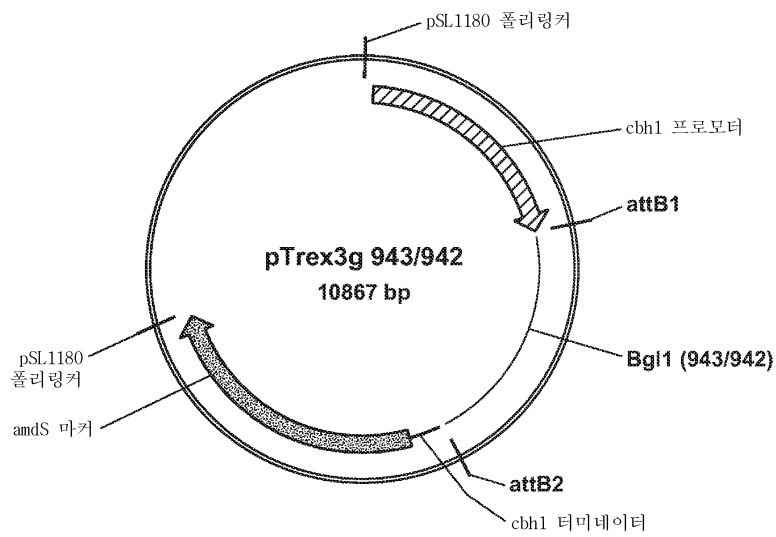
도면55a



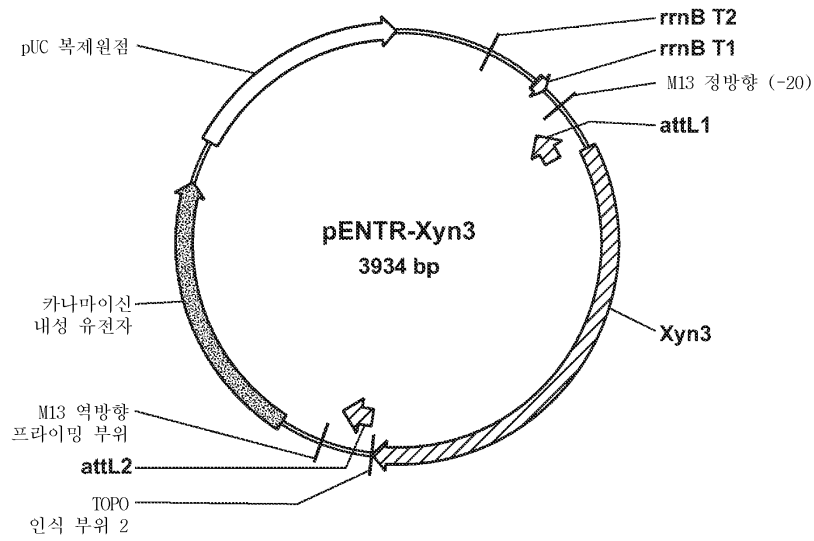
도면55b



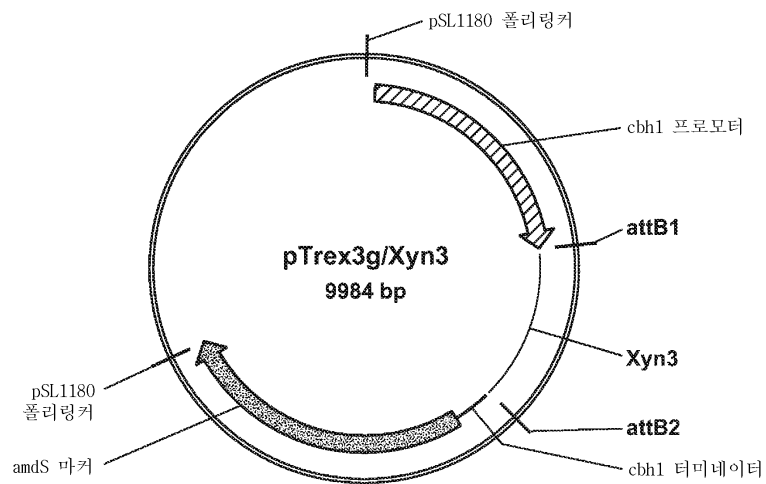
도면55c



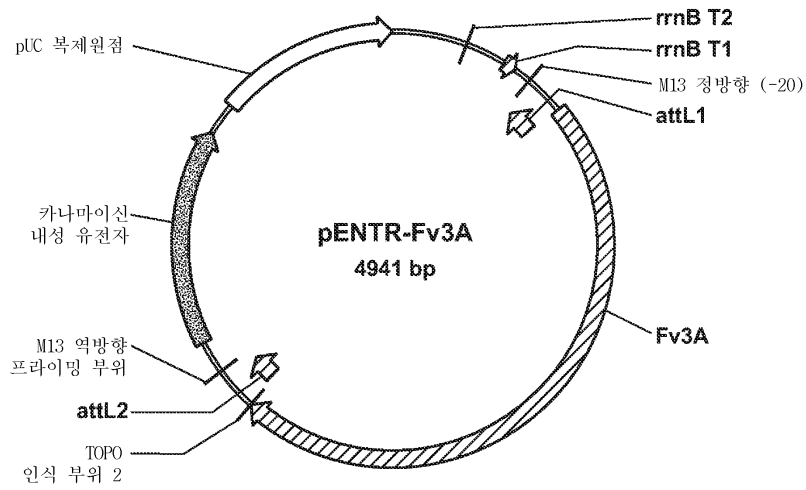
도면55d



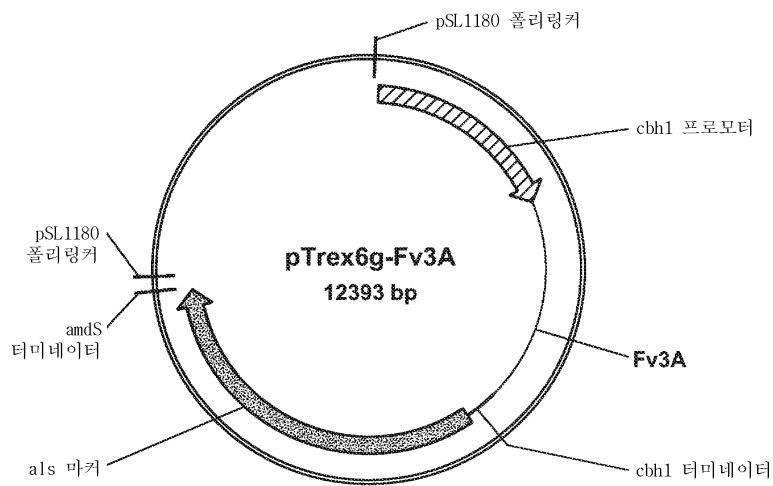
도면55e



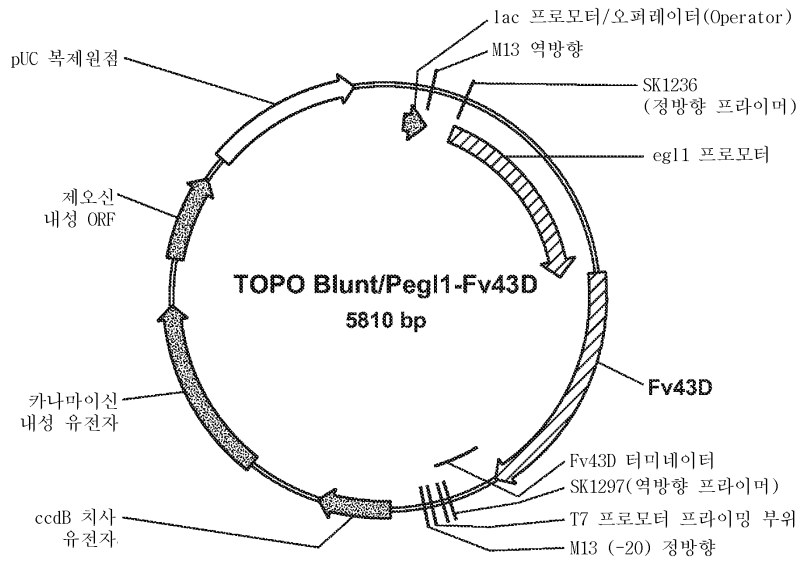
도면55f



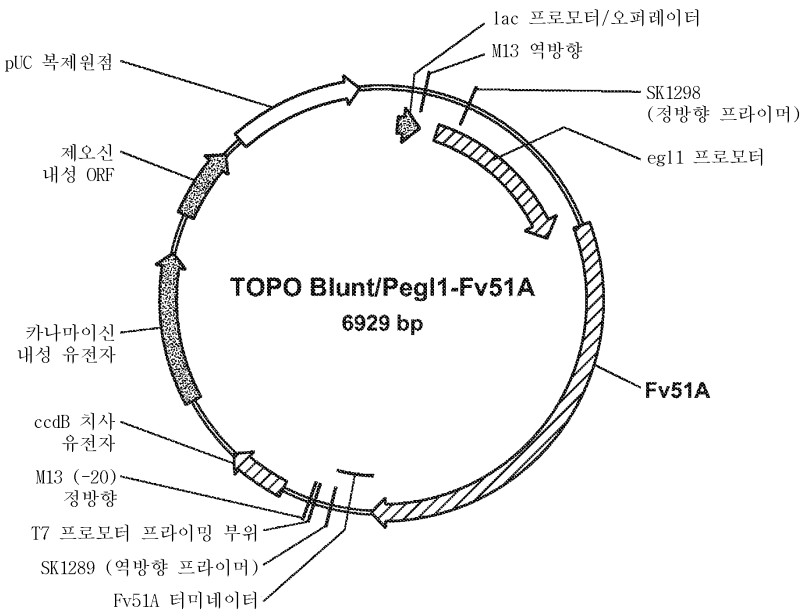
도면55g



도면55h



도면55i



도면56

```

gi|reesei|Bx11 -MVNNAALLAALSALLPTALAQNNQTYANYSAQCQPDLYPETLATLTLSPDCEHGPLKN
gi|fvert| Fv3A MLNLQVAASALSLSLGLAEATPYT-----LPDCTKGPLSK
      ::* . :*** * .** : * : :*** :***:
gi|reesei| NLVCDSSAGYVERAQUALISLFTLEELILNTQNSGPGVPRGLPNYQVWNEALHGLDRAN-
gi|fvert| NGICDTSLSPAKRAAALVAALTPEEKVGNLVSNTGAPRIGLPRYNWNEALHGLAGSPG
      * :*: * . :*: ** : :* * : * . . . . * .** :*** :* : ***** :
gi|reesei| --FATKGGQFEWATSFPMPIILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNV
gi|fvert| GREFADTP-PYDAATSFPMPLLMAAAFDDDLIHDIGNVVGTEARAFNTGGWRGVDFWTPNV
      ** . : : ***** : * :*: : *** : . : : * :*** :* * * :* :***
gi|reesei| NGFRSPLWGRGQETPGEDAFFLSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENW
gi|fvert| NPFKDPWRGRGSETPGEDALHVS-RYARYIVRGLEG--DKEQRIVATCKHYAGNDFEDW
      * * : * **** :***** : : * * : : * : * * : : : * * * * * :* : *
gi|reesei| NNQSRIGFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCAYNSVNGVPCANSFFLQTL
gi|fvert| GGFTRHDFDAKITPQDLAEYYVRPFQECTRDAKVGSIKAYNAVNGIPACANSYLQETIL
      . . : * .** * * * :*** : * . : * * * :***** :*** :* :*** : : * :
gi|reesei| RESWGFPEWGYVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNE
gi|fvert| RGEHNWTRDNWITSDCGAMQDIWQNHKYVKTNAEGAQVAFENGMDSSCEYTTTSDVSDS
      * * : . : : :*** : : : : * . . . . : * . : . * * * * . : : *
gi|reesei| FVAGEVSRGEIERSVTRLNLYANLVRLGYFD-KKNQYRSLGWKDVVKTDAWNISYEAAVEGI
gi|fvert| YKQGLLETKLMDRSLKRLFEGLVHTGFFDGAQWNSLSFADVNTKEAQLALRSAVEGA
      : * : . : :*** :** : ** : *** * * :*** : * * . : * : : :****
gi|reesei| VLLKNDGTLPKSKVR-SIALIGPWANATTQMGGNYGPAPYLISPLEAAKKAGYHVNF
gi|fvert| VLLKNDGTLPKLLKKKDSVAMIGFWANDTSKLOGGYSGRAPFLHSPLYAAEKLGLDTNVA
      ***** * : * : * : * * * * : : * * * * * * * * * * * * * . . .
gi|reesei| LGTEIAGNSTTG--FAKAIAAAKSDAIIYLGIDNTIEQEGADRTDIWPGNQDLKIQ
gi|fvert| WGPTELQSSSHDNWTTNAVAAAKSDYILYFGGLDASAAGEDRDRENLDWPESQLTLLQK
      * . : . : * : : : :***** * :*** : * : * . * : : * * * * : :
gi|reesei| LSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRL
gi|fvert| LSSLGKPLVVIQLG-DQVDDTALLKNNKINSILWVNPQDGGTAVMDLLTGRKSPAGRL
      ** :***** :* * * : : * * :*** :*** : * * * * * : : * : : :*****
gi|reesei| VTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWTGKPVYEFSGGLFYTTFKETLASHP
gi|fvert| PVTQYPSKYIEQIGMTDMDLRPT-KSLPGRTYRWYS-TPVLPYGGLHYTKFQAKFKSN-
      .*** : : . : : ** :*** * * * * * * : * * : * * * * : : * :
gi|reesei| KSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSNAGPAPYP
gi|fvert| -KLTFDIQKLLG--CSAQYSDTCALPPIQVSVKNTGRIISDFVSLVFIKS-EVGPKPY
      . * : : : . . . * : * : : : :*** : * : : : : : : * * * *
gi|reesei| NKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNNRIVYPGKYELALNTDESVKLEF
gi|fvert| LKTLAAYGRLEDVAPSTKDISLEWTLDNIAARRGENDLVVYPGTYTLLLEPTQAKIQV
      * * . . : * * : * . : : : * : : * * . : : : * * * * : * : : :
gi|reesei| ELVGEEVTIENWPLEEQQIKDATPDA
gi|fvert| TLTGKRAILLKWPQDPKSA-----

```

도면57a

Fv43D ---MQLKFLSSALLSLTGNCAAQDTNDIPPLITDLWSADPSAHVFEGKLWVYPSHDIEA
Fo43A ---MQLKFLSSALLSLTGNCAAQDTNDIPPLITDLWSADPSAHVFEGKLWVYPSHDIEA
Gz43A ---MKSLLFP---LLSFVG---QSLATNDCCPLITSRWTADPSAHVFNDTLWLYPSHDIDA
Pf43A --MLQRFAYILPLALLSVG--VKADN----PFVQSIYTA^DPFAPMVYNDRVYVFMHDNDTG
Fv43A -MWLTSPLLFASTLLGLTGVALADN-----PIVQDIYTA^DPFAPMVYNGRVYLFTHGHDNDG
Fv43B --MRFSWLLCPLLAMGSALPETKTDVSTYTNFVLPGWHS^DPFSC-IQKDGFLFCVTSTFIS
Af43A -----MAAPSLSYPTGTQSYTNPLFPGWHS^DPFSCAYVAEQDTFFCVTSTFI
Pf43B ----MSRSILPYASVFALLGGATAEP----FLVLNSDFF^DPFSLIETSSGYAFGTTGNGV
Fv43E MKVYWLVAWATSLTPALAGLLGHRRATTFNNPIIYSDFF^DNDVFLGPDNYYYFSASNFHF

Fv43D NVVNGTGGAQYAMR^DYHTYSMKTIYKDPVIDHGVALSVDDVPWAKQQMWAPDAAYK--N
Fo43A NVVNGTGGAQYAMR^DYHTYSMKSIYKDPVVDHGVALSVDDVPWAKQQMWAPDAAHK--N
Gz43A GFENDPDGGQYAMR^DYHVYSIDKIYGLP-VDHGTALSVEDVPWASRQMWAPDAAHK--N
Pf43A -----ATYNNMT^DWHLFSSADMANWQD---HGIPMSLANFTWANANAWAPQVIPR--N
Fv43A -----STDFNMT^DWRLFSSADMVNWQH---HGVPMSLKTFSWANSRAWAGQVVAR--N
Fv43B FP---GLPVYASR^DLVNWRILISHVWNRE---KQLPGISWKTAGQQQGMYPYTRYH--K
Af43A AFP---GLPLYASR^DLQNWKLASNIFNRP---SQIPDLR-VTDGQSGIYAPTTRYH--E
Pf43B N-----AQVASSP^DFNWTWLLSGT-----DALPGFPFSWVASSPQIWAPDVLVKA-D
Fv43E SP----GAPVLKSK^DLLNWDLLGHSLPRLNFGDGYDLPPGSRYYRG-GTWASSLRYRKS

Fv43D GKYYLYFPAK-DK-DEIFRIGVAVSNKPSGPFK---ADK-SWIPGTYS^{ID}PASYVDTNGE
Fo43A GKYYLYFPAK-DK-DEIFRIGVAVSNKPSGPFK---ADK-SWIPGTYS^{ID}PASYVDTDNE
Gz43A GKYYLYFPAK-DK-DIFRIGVAVSPTPGGFV---PDK-SWIPHTFS^{ID}PASFVDDDDR
Pf43A GQFYFYAPVR-HN-DGSMAGVGVSTITGPYH---DAIGKPLVENNE^{ID}PTVFIDDDGQ
Fv43A GKFYFYVVRNAK-TGGMAGVGVSTNILGPYT---DALGKPLVENNE^{ID}PTVYIDTDGQ
Fv43B GTYVYVCEYLGVG-DIIGVIFKTTNPWDESSWS---DPV---TFKPNH^{ID}PDLEFWDGDKG
Af43A GQFYLYVSYLGP--QTKGLFTSSDPYDDAAWS---DPL---EFAVHG^{ID}PDIFWDHDTG
Pf43B GTYVMYFSASASDSGKHCVGAATATSPEGPYTPVDSAVACPLDQGGAI^{ID}DANGFIDTDGT
Fv43E GQWYWIGCIN-----FWQTVVYTASSPEGPWY---NKGNEGDNNCY^{ID}NGILIDDDDT

Fv43D AYLIWGGI-WGQQLQAWQDHKTFNESWLGDKAAPNGTNALSPQIAKLSKDMHKITETPRD
Fo43A AYLIWGGI-WGQQLQAWQDKKNFNESWIGDKAAPNGTNALSPQIAKLSKDMHKITETPRD
Gz43A AYLAWGGI-MGQQLQRWQDKKNYNES--GTEPG-NGTAALSPQIAKLSKDMHTLAEKPRD
Pf43A AYLYWG-----NPDLWYVKLNQDMISYSGSPTQ
Fv43A AYLYWG-----NPGLYYVKLNQDMISYSGSINK
Fv43B VYCATHG----ITLQEI^DLETGELSPELNIWNGTGGVWPEGPHIYKRDGYYYLMIAEGGT
Af43A VYVTS^{AED}-QMIKQYTLDLKTGAIGPVDYLWNGTGGVWPEGPHIYKRDGYYYLMIAEGGT
Pf43B IYVVYKID-----GNSLDGDGTHPTPIMLQOMEADGT
Fv43E MYVVYGSGEVKVSQLSQDGSQVKSQVVFKN^DTIGVQDLEGNRM^YKING-----LYYI

Fv43D LVILAPETGKPLQAEDNKRFFE^{EGP}-----WVHKRGKLYLYMYSTG-----
Fo43A LVILAPETGKPLQAEDNKRFFE^{EGP}-----WIHKRGKLYLYMYSTG-----
Gz43A MLILDPKTGKPLLS^EDEDRFFE^{EGP}-----WIHKRNKIYYLYSTG-----
Pf43A IPLTTAGFGTRTGNAQRPTT^{FE}^{EGP}-----WVYKRN^{GI}YIAYAAD-----
Fv43A VSLTTAGFGSRENN^{AQRPTT}^{FE}^{EGP}-----WLYKRN^{LI}YMIYAAN-----
Fv43B ----AEDHAIT^{IARARKITGPY}^{EAYNNNPILTN}RGTS^{EYFQTVGHGDLFQD}TKGNWWGLC
Af43A ----ELGHSET^{MARSRT}RG^{WEPYPHNPLLSN}KGTSEYFQTVGHAD^{LFQD}GNGNWWAVA
Pf43B --TPTGSP^{IQLIDRSDL}DG^{PLI}^{EAP}-----SLLLSNGIY^{LS}FS^{SN}-----
Fv43E LNDSPSGSQ^{TWIKSKSPWGPY}^{ESKVLADKVTPPISGGNSPHQGS}LIKTPNGGWY----

Fv43D -DTHFLVYATSKN---IYGPYT-----YQGKILDPVDG-----WTTHG
Fo43A -DTHFLVYATSKN---IYGPYT-----YRGKILDPVDG-----WTTHG
Gz43A -TTHYL^{VYAT}SKT---PYGPYT-----YQGRILEPVDG-----WTTHS
Pf43A CCS^{EDIRYSTGTS}---ATGPWT-----YRGVIMPTQ^{SS}-----FTNHE

도면57b

Fv43A CCSEDIRYSTGPS---ATGPWT-----YRGVVMNKAGRS-----FTNHP
Fv43B LATRITAQGVSPMGREAVLFNGTWNKGWPKLQPVGRMPGNLLPKPTRN-----VPGD
Af43A LSTRSGPAWKNYPMGRETVLAPAAWEKGEWPVIQPVRGQMGG-PFPPPNKR-----VPRGE
Pf43B YYNTNYDYTSYAYASSITGPWT-----KQSAPYAPLLVTGT-----ETSND
Fv43E FMSFTWAYPAGRLPVLAPITWG-----SDGFPILVKGANGGWSSSYPTLPGT

Fv43D SIVEYKQWWLFFAD-AHTSGKDYLRQVKARKIWDKDG-----KILLTRPKI-----
Fo43A SIVEYKQWWLFFAD-AHTSGKDYLRQVKARKIWDKDG-----KILLHRP-----
Gz43A SIVKYQQWWLFYHD-AKTSGKDYLRQVKAKKIWDYDSKG-----KILTCKP-----
Pf43A GIIDFQNNSYFFYHNGALPGGGGYQRSVCVEQFKYNADG-----TIPTIEMTTAG-----
Fv43A GIIDFENNSYFFYHNGALDGGSGYTRSVAVESFKYGS DG-----LIPEIKMTTQG-----
Fv43B GPFNADPDNYNLKTKKIPPHFVHHRVPRDGAFSLSSKG---LHIVPSRNNVTGSLVLP
Af43A GGWIQPKDKVDFRPGSKI PAHFQYWRYPKTEDFTVSPRGHENTLRLTPSFYNLTG-----
Pf43B GALSAPGGADESVDTGKMLFHANLNGQDISGGRALFAAS-----ITEASDVVTLQ-----
Fv43E DGVTKNWTRTDTFRGTSLAPSWEWNNHNPVNSFTVNNGLTLRTASITKDIYQARN-----

Fv43D -----
Fo43A -----
Gz43A -----
Pf43A -----PAQIGTLNPPYVRQEAETAAWSSGITTEVCSEGGIDVGFINNG
Fv43A -----PAQLKSLNPPYVRQEAETIAWSEGIETEVCSSEGGNLNVAFIDNG
Fv43B DEIELSGQRLAFIGRRQTHTLFKYSVDIDFKPKSDDQEAGITVFRTQFDHIDLGIIVRLP
Af43A -TADFKPDDGSLVMRKQTDTLFTYTVDVSDPKVADEEAGVTVFLTQQQHIDLGIIVLLQ
Pf43B -----
Fv43E -----TLSHRTHGDHPTGIVKIDFSPMKDGDRAGLSAFRDQSAIYIGIHRDNGK

Fv43D -----
Fo43A -----
Gz43A -----
Pf43A DYIK-----VKGVAFGS-GAHSFSARVASANSGGTIAIHLGSTTGTLVGTCTV
Fv43A DYIK-----VKGVDFGSTGAKTFSARVASNSSGGKIELRLGSKTGKLVGTCTV
Fv43B TNQGSNNKSKLAFRRFRATGAQNVPAKP---VVPVPDGEWKGVISLHIEAANATHYNLGAS
Af43A TTEG---LSLSFRFRVEGRGNYEGPLPEATVPVPKEWCGQTIRLEIQAVSDTEYVFAAA
Pf43B -----
Fv43E FTIAT---KHGMNMDEWNGTTTDLGQIKATANVPSGRTKIWLRLQLDTPAGTGNTIFS

Fv43D -----
Fo43A -----
Gz43A -----
Pf43A PSTGGWQTWTTVTCSVSGASGTQ-----DVYFVFGSGGTGYLFN-----FDYWQFA
Fv43A TTTGNWQTYKTVDCPVSGATGTS-----DLFFVFTGSGSGSLFN-----FNWWQFS
Fv43B --SHRGKTLDIATASASLVSGGTGSGFVGSLLGPYATCNGKSGVECPKGGDVYVTQWYK
Af43A PARHPAQRQIIISRANSLIVSGDTGRFTGSLVGVYATSNNG-GAGSTP-----AYISRWRYE
Pf43B -----
Fv43E YSWDGVKYETLGFNFKLYNG-----WAFFIAYRFGIFNFAETALGGSIKVESFT

Fv43D -----
Fo43A -----
Gz43A -----
Pf43A -----
Fv43A -----
Fv43B PVAQEIDHGVFVKSEL
Af43A GRGQMIDFGRVVPY-
Pf43B -----
Fv43E AA-----

도면58

Pa51A MIHLKPALAALLALSTQCVAIDL FVKSSGGNKTTDIMYGLMHEDINNSGDGGIY AELISN
Fv51A MVRFPSSILAAAACF--VAVESVNIKVDSKGGNATSGHQYGFLEHEDINNSGDGGIY AELIRN
Pf51A MGKMWHISILVVLGLLSVGHAITINVSQSGGNKTSPLQYGLMFEHEDINHGGDGGLY AELVRN

Pa51A RAFQGSSEKFPSNLDNWSVPVGGATLTQLKAKPLSSALPYSVNVANPKEGKGKDKTKGKK
Fv51A RAFQYSKKYPVSLSGWRPINDAKLSLNRLDTPLSDALFVSMNVK---PGKGK-----AKE
Pf51A RAFQGSTVYPANLDGYDSVNGAILALQNLTNPLSPSPSSLNVA-----KGS-----NNGS

Pa51A VGLANAGFWGMDVKRQKYTG SFHVTGEYKGD FEVSLRSAITGETFGKKVVKGSGKKGKWT
Fv51A IGFLENEG YWGMVKKQKYTG SFVWK GAYKGHFTASLRSNLTDDVFGSVKVKSKANKKQWV
Pf51A IGFANEGWGWIEVKPQRYAGSFYVQGDYQGGDFDISLQSKLTQEVFATAKVRSSGKHEDWV

Pa51A EKEFELVPFKDAPNSNNTFVVQWDAE GAKDGS LLDNLISLFPPTFKGRKNGLRIDLAQTM
Fv51A EHEFVLTPKNAPNSNNTFAITYDPKGA-DGALDFNLISLFPPTYKGRKNGLRVDLAEAL
Pf51A QYKYELVPKKAASNTNNTLTITFDSKGLKDGSLNFNLISLFPPTYNNRPNGLRIDLVEAM

Pa51A VELKPTFLRFPGGNMLEGNTLDTWKKWYETIGPLKDRPGMAGVWEYQOTLGLGLV EYMEW
Fv51A EGLHPSLLRFPGGNMLEGNTNKTWWDWKDTLGPLRNRPGFEGVWNYQQT HGLGILEYLQW
Pf51A AELEGKFLRFPGGSDVEGVQAPYWKWNETVGD LKDRYSRPSAWTYEESNGIGLIEYMNW

Pa51A ADDMNEPIVGVFAGLALDGSFVPESEM GWVIQQALDEIEFLTGD AKTTKWGA VRAKLGH
Fv51A AEDMNEIIVGVYAGLSLDGSVTPK DQLQPLID DALDEIEFIRG-PVTSKWGKKRAELGH
Pf51A CDDMGLEPILAVWDGHYLSNEVIS ENDLQPYIDDTLNQLEFLMG-APDTPYGSWRASLG Y

Pa51A PKPWKVKWVEIGNEDWLAGRPAGFESYIN YRFPMMMKAFNEKYPDIKIIASPSIFD----
Fv51A PKPFRLSVVEVGNEDWLAGYPTGWNSYKEYRFP MFLEAIKKAHPDLTVISSGASIDPVGK
Pf51A PKPWTINVVEIGNEDNLYG---GLETYIAYRFQAYYDAITAKYPHMTVMESLTEMPG---

Pa51A ---NMTIPAGAAGDHHHPYLT PDEFVERFAKFDNLSKDNVTLIGEA AASTHPNG---GIAWE
Fv51A KDAGFDIPAPGIGDYHPYREPDVLVEEFNLFDNNKYG--HIIGEVASTHPNG---GTGWS
Pf51A -----PAAAASDYHQYSTPDGFVSQFN YFDQMPVTNRTLNGE IATVYPNNPSNSVAWG

Pa51A GDLMLPLPWGGSVAEAI FLISTERNGDKII GATYAPGLRSLDRWQWSMTWVQH AADPALT
Fv51A GNLMPPYWWISGVGEAVALCGYERNADRI PGTFYAPILKNENRWQWAITMIQFAADSAMT
Pf51A SPFPPLYPWWIGSVAEAVFLIGEERNSPKII GASYAPMFRNINNQQWSPTLIAFDADSSRT

Pa51A TRSTSWYVWRILAHHIIRETLPVDAPAGKPNFDPLFYVAGKSES-GTGIFKAAVYNSTES
Fv51A TRSTSWYVWSLFGHPMTHTLP TTA-----DFDPLYVAGKNEDKGT LIWKGAA YNTTKG
Pf51A SRSTSWHVIKLLSTNKITQNLPTTWSGG--DIGPLYWVAGRNDNTGSNIFKAAVYNSTSD

Pa51A --IPVSLKFDGLNEGAVANLT VLTGPE-DPYGYNDPFTGINVVKERTTFIKAGKGGKFTF
Fv51A ADVPVSLSPFKGVKPGAQAE LTLTNKEKDPFAFNDPHKGNVVDTKKTVLKADGKGAFNF
Pf51A --VPVTVQFAGCN-AKSANLTI LSSDD--PNASNYPG-GPEVVKTEIQSVTANAHGAFEF

Pa51A TLPGLSVAVLETADAVKGGKGKKGKKGKN
Fv51A KLPNLSVAVLET LK-----KGKPYSS
Pf51A SLPNLSVAVLKTE-----

도면59a

```

      *           20           *           40           *
xyn3   : -MKANVILC--LLAPLVAAAPTETIHLDPELAALRANLTERTADLWDRQA : 47
P56588 : -----QA : 2
P23360 : MVRPTILLTSLLLAPFAAASPI-----LEERQA : 28

      60           *           80           *           100
xyn3   : SQSIDQLIKRKGLYFGTATDRGLLQRE-KNAAIQADLGQVTPENSMKW : 96
P56588 : SVSIDAKFKAHGKKYLGTIGDQYTLTKNTKNPAIIKADFGQLTPENSMKW : 52
P23360 : AQSVDQLIKARGKVYFGVATDQNRLTTG-KNAAIQADFGQVTPENSMKW : 77

      *           120          *           140          *
xyn3   : QSLENNQGQLNWGDADYLVNFAQQNGKSIRGHTLIWHSQLPAWVNNINNA : 146
P56588 : DATEPNRGQFTFSGSDYLVNFAQSNGLIRGHTLVWHSQLPGWVSSITDK : 102
P23360 : DATEPSQGNFNFAGADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDK : 127

      160          *           180          *           200
xyn3   : DTLRQVIRTHVSTVVGRYKGGKIRAWDVVNEIFNEDGTLRSSVFSRLLGEE : 196
P56588 : NTLISVLKNHITTVMTRYKGGKIYAWDLNEIFNEDGSLRNSVFYINVIGED : 152
P23360 : NTLTNVMKNHITTLMTRYKGGKIRAWDVVNEAFNEDGSLRQTVFELNVIGED : 177

      *           220          *           240          *
xyn3   : FVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPI : 246
P56588 : YVRIAFETARSVDPNAKLYINDYNLDSAGYSKVNCGMVSHVKKWLAAGIPI : 202
P23360 : YIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNRVKQWRAAGVPI : 227

      260          *           280          N           *           300
xyn3   : DGIGSQSHLSGGGSGTGLGALQQLATVPVTELAITELDIQGAPTTDYTVQV : 296
P56588 : DGIGSQTHLGAGAGSAVAGALNALASAGTKEIAITELDIAGASSTDYVNV : 252
P23360 : DGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITELDVAGASPTDYVNV : 277

      *           320          *           340          *
xyn3   : VQACLSVSKCVGITVWVGISDKDSWRASTNPLLF DANFNPKPAYNSIVGIL : 346
P56588 : VNACLNQAKCVGITVWGVADPDSWRSSSSPLLF DGNYNPKAAYNAIANAL : 302
P23360 : VNACLVQSCVGITVWGVADPDSWRASTTPLLF DGNFNPKPAYNAIVQDL : 327

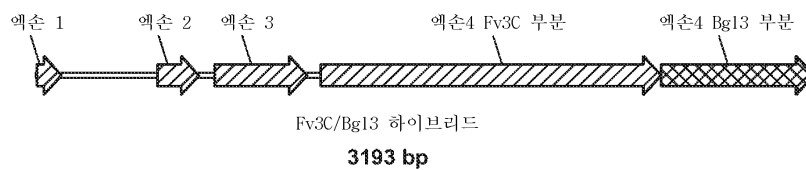
xyn3   : Q- : 347
P56588 : -- : -
P23360 : QQ : 329

```

도면59b

1	M V S F S Y L L L A C S A I G - A L A A P V E P E T T S F N	AfuXyn2
1	M I S I S S L S F G L A A I A G A Y A L P S D - K S V S L A	AfuXyn5
1	M V S F T S L L A A S P P S R A S C R P A A E V E S V A V E	Xyn2
30	E T A L H E F A E R A G T P S S T G W N N G Y Y Y S F W T D	AfuXyn2
30	E R - - - - - Q T I T T S Q T G T N N G Y Y Y S F W T N	AfuXyn5
31	K R - - - - - Q T I Q P G - T G Y N N G Y F Y S Y W N D	Xyn2
60	G G G D V T Y T N G A G G S Y S V N W R N - - V G N F V G G	AfuXyn2
53	G A G S V Q Y T N G A G G E Y S V T W A N Q N G G D F T C G	AfuXyn5
53	G H G G V T Y T N G P G G Q F S V N W S N - - S G N F V G G	Xyn2
88	K G W N P G S A - R T I N Y G G S F N P S G N G Y L A V Y G	AfuXyn2
83	K G W N P G S D - H D I T F S G S F N P S G N A Y L S V Y G	AfuXyn5
81	K G W Q P G T K N K V I N F S G S Y N P N G N S Y L S V Y G	Xyn2
117	W T T N P L I <u>N</u> E Y Y V V <u>E</u> S Y G T Y N P G S G G T F R G T V	AfuXyn2
112	W T T N P L V <u>E</u> Y Y I L <u>E</u> N Y G S Y N P G S G M T H K G T V	AfuXyn5
111	W S R N P L I <u>E</u> Y Y I V <u>E</u> N F G T Y N P S T G A T K L G E V	Xyn2
147	N T D G G T Y N I Y T A V R Y N A P S I E G T K T F T Q Y W	AfuXyn2
142	T S D G S T Y D I Y E H Q Q V N Q P S I V G T A T F N Q Y W	AfuXyn5
141	T S D G S V Y D I Y R T Q R V N Q P S I I G T A T F Y Q Y W	Xyn2
177	S V R T S K R T G G T V T M A N H F N A W S R L G M N L G T	AfuXyn2
172	S I R Q N K R S S G T V T T A N H F K A W A S L G M N L G T	AfuXyn5
171	S V R R N H R S S G S V N T A N H F N A W A Q Q G L T L G T	Xyn2
207	H N Y Q I V A T <u>A</u> <u>E</u> G Y Q S S G S A S I T V Y	AfuXyn2
202	H N Y Q I V S T <u>E</u> <u>E</u> G Y E S S G T S T I T V S S G G S S S G G	AfuXyn5
201	M D Y Q I V A V <u>E</u> <u>E</u> G Y F S S G S A S I T V S	Xyn2
228		AfuXyn2
232	S G G S S S T T S S G S S P T G G S G S C S A L W G Q C G G	AfuXyn5
222		Xyn2
228		AfuXyn2
262	I G - W S G P T C C S S G T C Q V S N S Y Y S Q C L	AfuXyn5
222		Xyn2

도면60a



도면60b

도 60b-1

서열 번호 82
Fv3C/Bg13를 암호화하는 뉴클레오타이드 서열

atgaagctgaattgggtcgccgacgacctgtctctataggtgctgctggcactgacagcgca
gttgcctctgtctctgcagttccagacactttggctgggtgtaaaagtcagtttttttca
ccatttctcgtcttaactcgaacctgttgccatctgcacctgttgcgtcggaagccac
gcaacagatcgcgatcatcttccctctgcaacctgttgcctcttaagacttccctcc
gcaattatcagcgcccttagtctacacaaaaacccccgagacagctcttctatgagtttg
tcgacatcaagttgctctctcaactgtgcatcttggctggctgtctacttctgcctctagac
aaccaaatctggcgcaattgaccgctcaaaccttgttcaaaataaccttttttattcgag
acgcacatttaataattggcgctttcaataataacgactttatggcgcgcggtgctgtg
ggcggttgatcagaaaagctgacgctcaaaaggttgtcacgagagatacactcgcatactc
gccgctcatcttacttcaacatggatggacacctaatgctgttggttgaggaggaagctta
cgccaaagccaagagctttgtgtcccaactcaactctcatggaaaaggtcaacttgaccac
tggtgttggttaagcagctccttgcaaacagggtatctcaatccccctcagctaacacactt
ctcagatggcgaagcggaagctgtgtgaggaacgtgggatcaattcctcgtctcggtag
cgaggtctctgtctccaggttggtcctcttggaaattcgctgtcgcgactacaacagcgct
tttccgctcgccacacagctggtgctcttggagcaagctctctctggtatgagagaggt
ctcctgtggtgacactgagttcaaggagaaggttatcgatatcgctcttggctcgtact
ggacctcttggtcgactgtgctgtgggacgaaactgggaaggttccacgttgatcct
tatatggctggccagccactggcgcagggccgtcaagggtattcaagacgcaggtgtcatt
gcttgtgtcaagcattacatcgcaaacgagcaggttaagccacttgagcgtttgaggaa
ttgacagagaactgacctcttcttagagcaacttcgcagacagatggcgaggtccagtcgcc
caagtaacaactctccgagctctctctctccaaactggatgacaagactatgcacgagct
ctacgctggcccttcgctgacgcgctccgcgcggcggtcggttcgctcatgtgctcgta
caaccagatcaacaactcgtacggttgccagaactccaagctcctcaacggtatcctcaa
ggacgagatgggcttccagggtttcgtcatgagcgtatggcgccgccaagctacacggtgc
cgctctcgcgctcgtcgtctcgatatgacatgctcgttgacactgcttcgacagcg
atacagcttctggggcggaacttgactctggtcgtcatcaacggaactgttccgccctg
gcgagttgatgacatggctctgcgaatcatgtcgtccttctcaagttggaagacgat
agaggatcttcccgacatcaacttctcctcctggaccccgacacacttcggtctgtgca
tactttgtcgaagagaacgcgagcaggtcaactttggagtaacgctccagcagaccca
caagagccacatccgtgagggcgctgccaaagggaagcgtcgtgctcaagaacacccgggtc
cctccccctcaagaacccaagttcctcgtctcatgtgtgaggaagcggttcccaacc
tgctggaccacaatggtttggttgacctggtttgcgataatggtaccctggctatggcttg

도 60b-2

gggctcgggaacttcccaattcccttaacttgatcaccocgatacaagggtctctataatcg
agctactcaagacgggaactcgatatgagagcatotttgaccaacaaggaatgggtctcagt
acaagctotttgtagcagcagctaaactgacccgatacgttttgcgaatgccgactctgg
tgagggtatcattgaagtcgacgggaactttggtgatcgcaagaacctcaacctctggca
cgaggagacgagctcatcaagaacgtgtcgtccatgcccccaacacattgtagtct
gcacacgctcgccctgtcctatcgtccgactacgagaagaaccccaacatcactgcoact
cgtctgggtgggtcttcccgcccaagagtcaagcaatgcaatcgctgatctcctctacgg
caaggtcagccctggcgatctcccttcaacttggggcgcaacccgagagactacggtac
tgaggttctttatgagcgacaacacggcggtggcgctcctcaggtgacttctctgaggg
tgtcttcatcgactacgctcaactcgacccgacgactctcaagcaccgatggaagagctc
tcccaacaacacgctcgtcctctctacgagttcggtcaggtctatcttggtcgagctt
caagttctccaacctccacatccagaagaacatgtcgcccatgagccgcccccaagcg
caagcagattggcgctcctctctggcgagcttcaagcaaaccttaaggactatggctt
cccccaagaacgttcgcgcgcatcaaggagtttatctacccctacctgagcaccactacctc
tggcaaggagggcgctcggttgacgctcaactacgcccagactgcgaagaggttctccccgc
cggtgccttggaacgcagccctcagcctcgtctcgcggtcttgccgaacccggcgga
ccgcccagctgacacatctctacacgtgacgcccacattaccaacacgggctcggt
catggacgacgacgcttccccagctgtacctgagcaccgcggttcccaacgagccgccaa
gggtcgtcggtggcttcgacccgcatcgagcgcttgcctccggccagagcgtcaagttcaa
ggcagacctgacgcgcggtgacgtgtccaactgggaacagaagaagcagcagtggtcat
taccgactaccccaagactgtgtacgtggggcagctcctcgcgacactgccgctgagcgc
ccgctgcactga

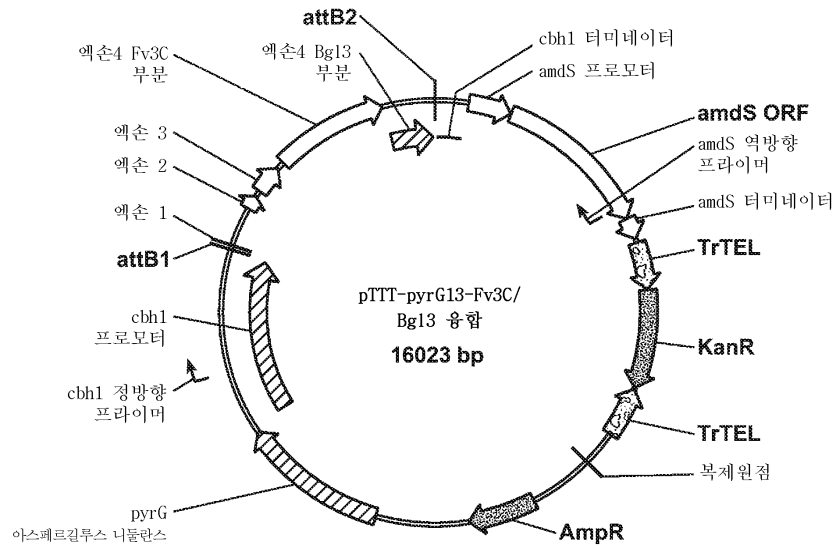
도면60c

서열 번호 159

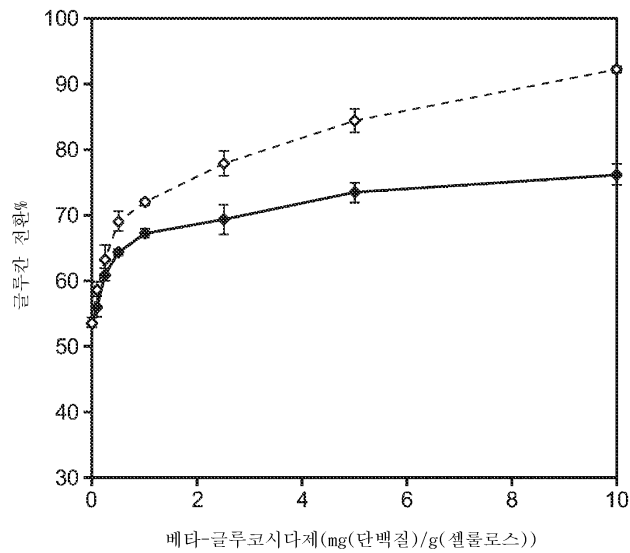
Fv3C/Bg13 카메라 폴리펩티드 서열(Bg13 카메라 부분은 볼드체와 대문자로 되어 있음)

mklnwaaalsigaagtdsavalasavpdtlagvkkadaqkvvtrdltlayspphypspswmdpna
vgweeayakaksfvsqltlmekvnlttgvgwgqgercvgngvgsiprlgmrglclqdgplgirisd
ynsafpagttagawskslwergllmgtefkekgidialgpattgplgrtaaggrnwegftvdp
ymaghamaeavkigiqdagviaacakhylaneghehfrqsggevqsrkyniseslssnlddktmhely
awpfadavragvgsvmscynqinnsygcqnsklngilkdemgfggfvmstdwaaqhtgaasava
gldmsmpgdtafdsygswggntltlavingtvpawrvddmalrimsaffkvgtiedlpdinfs
swtrdtfgfvhtfagenreqvnfgvnnvqhdkshireaaakgsvvlkntgslplknpkflavig
edagpnpagpngcgdrgcdngtlamawgsqsfpyltpdgglsnratqdgtryesilttnnew
asvqalvsgpntaivfanadsggyievdgnfgdrknltlwqggdeliknvssicpntiivlh
tvvgpvladyeknpnitaivwaglpqggesgnaiadllygkvspgrspftwgrtresygtevlye
anngrgapqdddfsegvfidyrhfdrrspstdgksspnntaaplyefghgls**WSTFKFSNLHIQK**
NNVGPMSPPNKGTIAAPSLGFSKLNLDYGFKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQT
AKEFLPAGALDGSPPQRSASGEPPGNNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNE
PPKVLRGFDRIERIAPEGQSVTFKADLTRRDLNWDTKKQQWVITDYPKTVYVGSSSRDLPISARLP

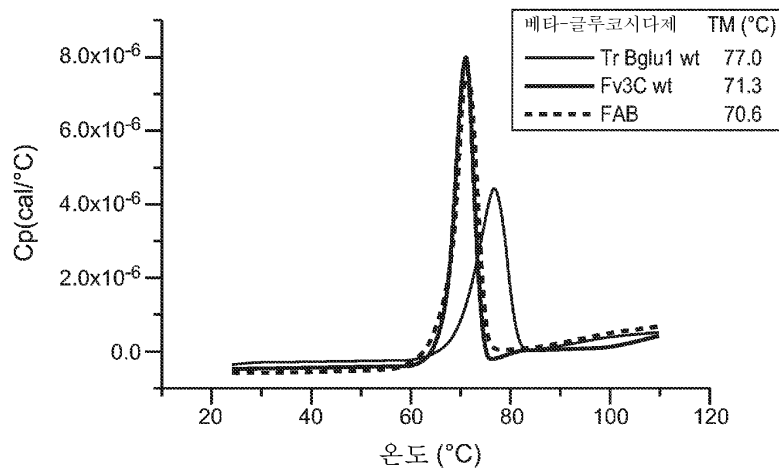
도면61



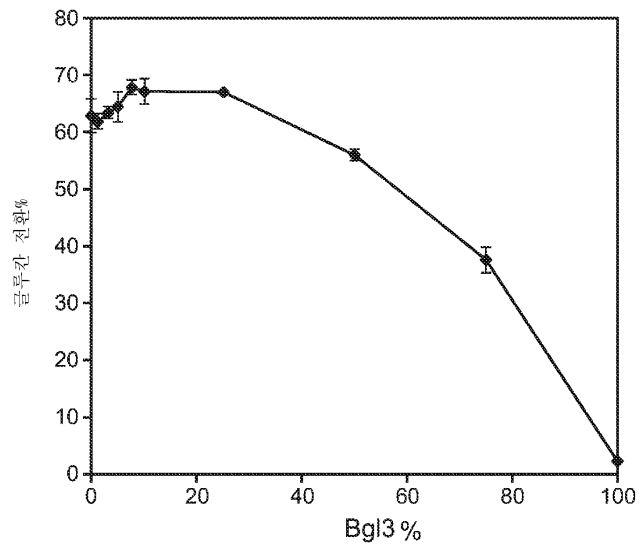
도면62



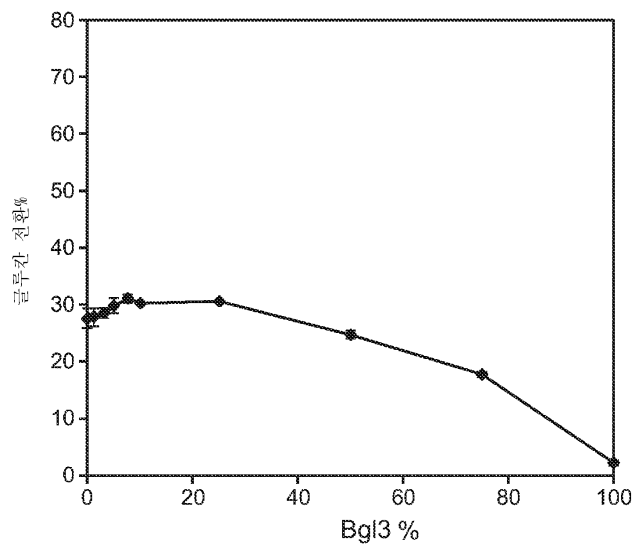
도면63



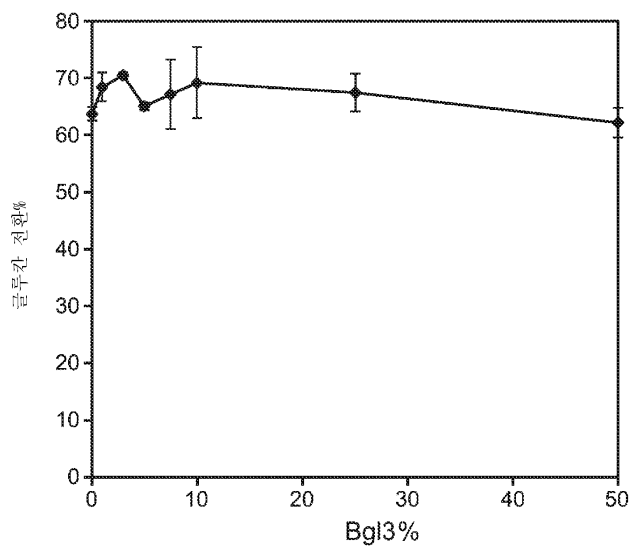
도면64a



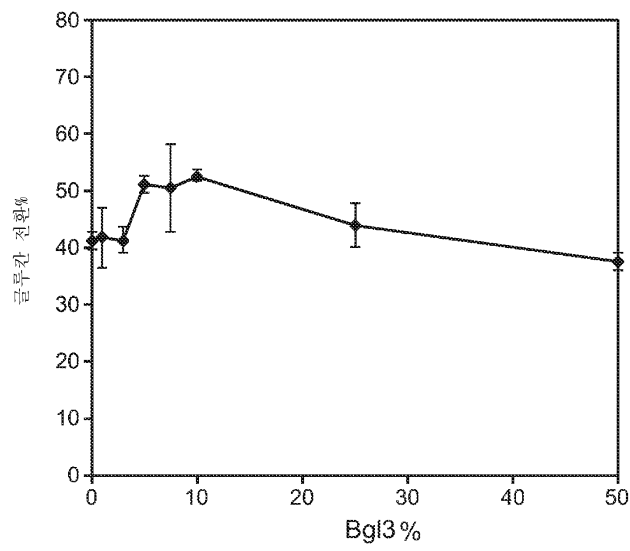
도면64b



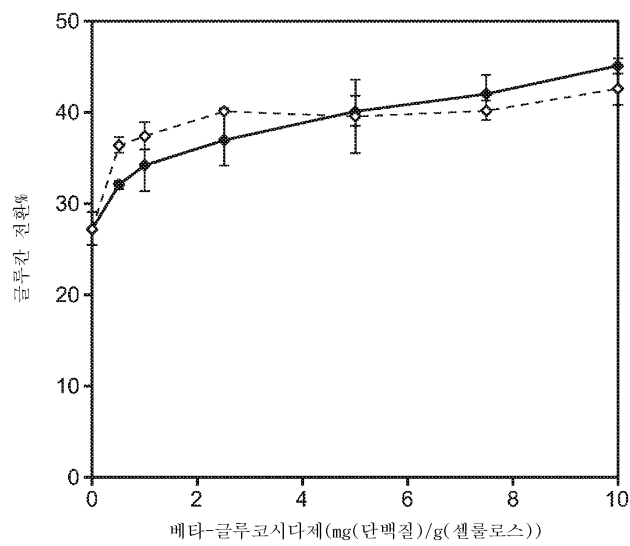
도면64c



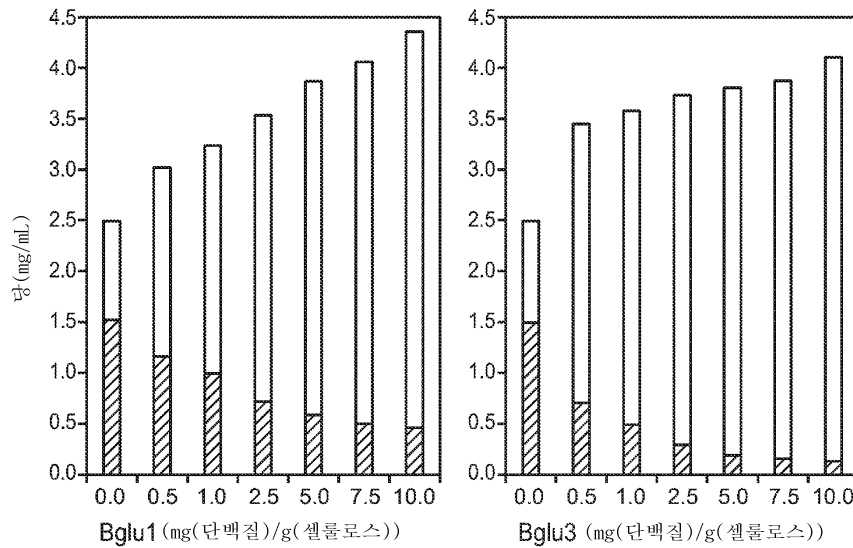
도면64d



도면65a



도면65b



도면66a

서열 번호	프라이머	서열
92	정방향 SK943	5'- CACCATGAGATATAGAACAGCTGCCGCT-3'
93	역방향 SK941	5'-CGACCGCCCTGCGGAGTCTTGCCAGTGGTCCCGCGACAG-3'
94	정방향 (SK940)	5'-CTGTCGCGGGACCACTGGGCAAGACTCCGCAGGGCGGTCG-3'
95	역방향 (SK942)	5'- CCTACGCTACCGACAGAGTG-3'
96	정방향 SK771	5' - GTCTAGACTGGAAACGCAAC -3'
97	역방향 SK745	5' - GAGTTGTGAAGTCGGTAATCC -3'
98	정방향 xyn3F-2	5'-CACCATGAAAGCAAACGTCATCTTGTGCCTCCTGG-3
99	역방향 (xyn3R-2)	5'-CTATTGTAAGATGCCAACAATGCTTATATGCCGGCTTGGGG-3'
100	정방향 SK745	5' - GAGTTGTGAAGTCGGTAATCC-3'
101	역방향 SK822	5' - CACGAAGAGCGGCGATTG-3'
102	정방향 MH124	5' - CAC CCA TGC TGC TCA ATC TTC AG -3'
103	역방향 MH125	5' - TTA CGC AGA CTT GGG GTC TTG AG -3'
104	정방향 SK1334	5' - GCTTGAGTGTATCGTGTAAG -3'
105	정방향 프라이머 SK1335	5' - GCAACGGCAAAGCCCCACTTC -3'
106	역방향 SK1299	5' - GTAGCGGCCGCCCTCATCTCATCTCATCCATCC -3'
107	정방향 SK1322	5'-CACCATGCAGCTCAAGTTTCTGTGC-3
108	역방향 SK1297	5'-GGTTACTAGTCAACTGCCCGTTCTGTAGCGAG-3'
109	정방향 SK1236	5'-CATGCGATCGCGACGTTTTGGTCAGGTCG-3
110	역방향 SK1321	5'-GACAGAAACTTGAGCTGCATGGTGTGGGACAACAAGAAGG-3'
111	정방향 SK1159	5'-CACCATGGTTGCGCTTCAGTTCAATCCTAG-3'
112	역방향 SK1289	5'-GTGGCTAGAAGATATCCAACAC-3'
113	정방향 SK1236	5'-CATGCGATCGCGACGTTTTGGTCAGGTCG-3'
114	역방향 SK1262	5'-GAAGTGAAGCGAACCATGGTGTGGGACAACAAGAAGGAC-3'
115	정방향 SK1298	5'-GTAGTTATGCGCATGCTAGAC-3'
116	정방향 MH234	5'-CACCATGAAGCTGAATTGGGTCGC-3'
117	역방향 MH235	5'-TTACTCCAACCTTGGCGCTG-3'

도면66b

서열 번호	프라이머	서열
118	MH255	5'-AAGCCAAGAGCTTTGTGTCC-3'
119	MH256	5'-TATGCACGAGCTCTACGCCT-3'
120	MH257	5'-ATGGTACCCTGGCTATGGCT-3'
121	MH258	5'-CGGTCACGGTCTATCTTGGT-3'
122	pDonor 정방향	5'- GCTAGCATGGATGTTTTCCAGTCACGACGTTGTAACGACGGC- 3'
123	Fv3C/Bgl3 역방향	5'-GGAGGTTGGAGAACTTGAACGTCGACCAAGATAGACCGTGACCGAAC TCGTAG-3'
124	pDonor 역방향	5'-TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG-3'
125	Fv3C/Bgl3 정방향	5'- CTACGAGTTCGGTCACGGTCTATCTTGGTCGACGTTCAAGTTCCTCA ACCTCC-3
126	Att L1 정방향	5'-TAAGCTCGGGCCCCAAATAATGATTTTATTTTACTGATAGT-3'
127	AttL2 역방향	5'-GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTACTGATA-3'
128	Cbh1 정방향	5'-GAGTTGTGAAGTCGGTAATCCCGCTG-3'
129	AmdS 역방향	5'-CCTGCACGAGGGCATCAAGCTCACTAACC-3'
130	Pr Cbh1 정방향	5'-CGGAATGAGCTAGTAGGCAAAGTCAGC-3'
131	725/751 역방향	5'-CTCCTTGATGCGGCGAACGTTCTTGGGGAAGCCATAGTCCTTAAGGTTCTT GCTGAAGTTGCCAGAGAG-3'
132	725/751 정방향	5'-GGCTTCCCCAAGAAGCTTCGCCGCATCAAGGAGTTTATCTACCCCTACC TGAACACCACTACCTC-3'
133	Ter Cbh1 역방향	5'-GATACACGAAGAGCGGCGATTCTACGG-3'
134	Forward MH234	5'-CACCATGAAGCTGAATTGGGTCGC-3'
160	Te3A 역방향	5'-GATAGACCGTGACCGAACTCGTAGATAGGCGTGATGTT GTACTTGTGAAG TGACGGTAGTCGATGAAGAC-3'
161	Te3A2 정방향	5'-GTCTTCATCGACTACCGTCACTTCGACAAGTACAACATCAC GCCTATCTACG AGTTCGGTCACGGTCTATC-3'

도면67a

서열 번호 135

Fv3C/Te3A/트리코테르마 리세이 Bgl3 (FAB) 키메라의 아미노산 서열

mklmwaaalsigaagtdsavlasavpdtlagvkkadaqkvvtrdtlaysspshypspwmdpna
vgweeayakaksfvsqiltmekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirlsd
ynsafpagttagaswkslwyergllmgtefkekgidialgpatgplgrtaaggrnwegftvdp
ymaghamaeavkgiqdagviacakhyaneeqehfrqsgevgqsrkyniseslssnlddktmhely
awpfadavragvgsvmcsynqinnsygcqnsklngilkdemgfggfvmstdwaaghtgaasava
gldmsmpgdtafdsyfwggnltlavingtvpawrvddmalrimsaffkvgktiedlpdinf
swtrdtfgfvhtfagenreqvnfgvnhqdhkshireaaakgsvvlkntgslplknpkflavig
edagpnpagpncgdrgcdngtlamawsgtsqfpyltpdgglsnratqdgtryesiltnew
asvqalvsqpnvtaixfanadsggyievdgnfgdrknltlwqggdeliknvssicpntivvlh
tvqpvlladyeknpnitaivwaglpqgesgnaiadllygkvspgrspftwgrtresygtevlye
annrggapqdddfsegvfidyrfhd**KYNITP**IyefghglswSTFKFSNLHIQNNVGPMSPPNGK
TIAAPSLGNFSKNLKDYGFPKNVRIKEFLYPYLNTTTSKGESGDAHYGOTAKEFLPAGALDG
SPQPRSAASGEPPGNRLYDILYTVTATITNTGSMDDAVPQLYLSHGGPNEPPKVLRGFDRIE
RIAPQSVTFKADLTRRDLNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP

도면67b

서열 번호 83

Fv3C/Te3A/트리코테르마 리세이 Bg13(FAB) 키메라를 암호화하는 핵산 서열:

atgaagctgaattgggtcgccgcagccctgtctataggtgctgctggcactgacagcgagttg
ctcttgcttctgcagttccagacacatttggctgggtgtaaaggctcagtttttttaccatttcc
tcgtctaattctcagccttgttgccatatcgcccttgttcgctcggaagccacgcaccagatcgc
gatcatttccctcccttgagccttgggttctcttactgatcttccctccgcaattatcagcgccc
ttagtctacacaaaaaccccgagacagtcctttcattgagtttctogacatcaagttgcttctc
aactgtgcatttgcgtgggtgtctacttctgcctctagacaaccaaactctgggcgcaattgacc
gctcaaaccttgttcaataaccttttttattcgagacgcacatttataaatatgcgccttca
ataataccgactttatgcgcggcggtgtgtgtggcggttgatcagaaagctgacgctcaaaagg
ttgtcagcgagagatacactcgcatactcgccgctcattatccttcaccatggatggaccctaa
tgctgttggctgggaggaagcttacgccaagccaagagcttctgttcccaactcactctcatg
gaaaggtcaacttgaccactgggtgttgggtaagcagctccttcgaaacagggtatctcaatcc
cctcagctaacaacttctcagatggcaaggcgaacgctgtgtaggaaacgtgggatcaattcct
cgctcggtatgcgaaggtctctgtctccaggatgggtcctcttggaaattcgtctgtcgcgactaca
acagcgcttttcccgctggcaccacagctgggtcttcttggagcaagtcctctctggtatgagag
aggtctcctgatgggcactgagttcaaggagaagggtatcgatatcgctcttggctcctgctact
ggacctcttggctgcactgctgtgtgtggacgaaactgggaaggcttcaccggttgatccttata
tggtcgccacgcatggccgagggcgtcaagggtattcaagacgcaggtgctcattgcttctgtgc
taagcattacatcgcaaacgagcagggtaagccacttggacgatttgaggaattgacagagaac
tgacctcttctgtagagcacttcgcagagagtggtgaggtccagtcgccgaagtacaacatctcc
gagtcctctcctccaactctggatgacaagactatgcacgagctctacgctggcccttctgtg
acgcgctccgcgcggcggtgtgtctcgtatgtctctgacaaccagatcaacaactcgtacgg
ttgacagaaactccaagctcctcaacggtatcctcaaggacgagatgggcttccagggttctgtc
atgagcgattggggcgccacagcataccgggtgcccgttctgcccgtcgctggtctcgatatgagca
tgctgtgtgacactgcttgcagacgggatacagcttctggggcggaacttgactctggtgtg
catcaacggaactgttcccgctggcgagttgatgacatggctctgcaatcatgtctgcttcc
ttcaagggttgaaagacgatagaggatcttcccgacatcaacttctcctcctggaccccgacaca
ccttcggcttctgtcatacatttgtcacaagaaacccgcgagcaggtcaacttggagtcacagct
ccagcacgacccacaagagccacatccgtgagggcgctgccaagggaagcgtcgtctcaagaac
accgggtcccttccctccaagaacccaaagtctcctcgctgtcattgggtgaggacgcccgtccca
accctgctggacccaatggttgtggtgacccgtgggttcgataatggtaccctggctatggcttg
gggtcgggaacttcccaattcccttacttgatcacccccgatcaagggtctctctaactcgagct
actcaagacggaactcgatatgagagcatcttgaccaacaacgaatgggttccagtacaagctc
ttgtcagccagcctaactgacccgtatcgttttgcgcaatgcccactctggtgagggatacat
tgaagtgcagcggaaacttgggtgatcgcaagaacctcaccctctggcagcagggagacgagctc
atcaagaacgtgtcgtccatatgccccaacaccattgtagtcttgcacaccgtcggccctgtcc
tactcgccgactacgagaagaacccccacatcactgccatcgtctgggctgggtcttcccgcca
agagtcaggcaatgccatcgctgatctcctctacggcaagggtcagccctggccgatctcccttc
acttggggcgccacccgcgagagctacggtactgaggttctttatgaggcgaacaacggccgtg
gcgctcctcaggatgacttctctgaggggtgtcttcatcgactaccgtcacttcgacaagtacaa
catcacgcctatctacgagttcggtcacgggtctatcttgggtcgacggtcaagttctccaacctc
cacatccagaagaacaatgtcggtcccatgagcccgcccaacggcaagacgatttgggtcctcc
ctctgggcaacttcagcaagaaccttaaggactatgggttccccaagaacggttcgcccgcacaa
ggagtttatctaccctacctgaacacaccactacctctggcaaggagcgctcggtgacgctcac
tacggccagactgcgaaggagttcctccccgcgggtgcccgtggacggcagccctcagcctcgt
ctgcggcctctggcgaacccggcggaacccgagctgtacgacattctctacaccgtgacggc
caccattaccaacacgggtcgggtcatggacgacgcccgttccccagctgtacctgagccacggc
ggtcccaacgagccgcccagggtgctgctggttctcgacgcgcatcgagcgcatttctcccgcc
agagcgtcacgttcaaggcagacctgacgcgcccgtgacctgtccaactgggacacgaagaagca
gcagtggtcattaccgactacccaagactgtgtacgtgggcagctcctcgcgacccgtgcgc
ctgagcgccccgctgccatga

도면68a

서열 번호	말단	서열 모티프
136	N	A-x-S-P-P-x-Y-P-S-P-W-M-D-P-x-A-x-G-W-E-x-A-Y
137	N	A-K-x-F-V-S-x-x-T-L-x-E-K-V-N-L-T-T-G-V-G-W-x-G-E-x-C-V-G-N-V-G
138	N	P-R-x-G-M-R-x-L-C-x-Q-D-G-P-L-G-x-R
139	N	Y-N-S-A-F-x-x-G-x-T-A-x-A-S-W-S
140	N	G-x-I-A-C-A-K-H-x-x-x-N-E-Q-E-H-x-R-Q
141	N	L-S-S-N-x-D-D-K-T-x-H-E-x-Y-x-W-P-F-x-D-A-V-x-A-G-V-G
142	N	M-C-S-Y-x-Q-x-N-N-S-Y-x-C-Q-N-S-K-L-x-N-G
143	N	G-F-Q-G-F-V-M-S-D-W-x-A-Q-H-x-G-x-A-x-A-V-A-G-L-D-M-x-M-P-G-D-T
144	N	N-L-T-L-A-V-x-N-G-T-V-P-x-W-R-x-D-D-M
145	N	P-x-F-L-x-V-x-G-E-D-A-G-x-N-P-A-G-P-N-G-C-x-D-R-G-C
146	N	G-T-L-A-M-x-W-G-S-G-T-x-F-P-Y-L
147	N	A-I-V-F-A-N-x-x-S-G-E-G-Y-I-x-V-D-G-N-x-G-D-R-K-N-L-T-L-W
148	N	D-x-L-Y-G-K-x-S-P-G-R-x-P-F-T-W-G
149	C	P-x-Y-E-F-G-x-G-L-S-W-x-T-F-x-x-S-x-L
150	C	L-x-D-Y-x-F-P
151	C	E-F-L-P-x-x-A-L-x-G-S-x-Q-P-R
152	C	S-G-x-P-G-G-N-x-x-L-x-D
153	C	Y-T-V-x-A-x-I-T-N-T-G
154	C	V-L-R-G-F-x-R-x-E-x-I-A-P-G-x-S
155	C	T-R-R-D-L-S-N-W-D-x-x-x-Q-x-W-V-I-T-D
156	C	V-G-S-S-R-x-L-P-L-x-A-x-L

도면68b

본 발명의 적절한 β -글루코시다제 폴리펩티드 하이브리드/키메라를
설계하는데 사용되는 아미노산 서열 모티프

서열 번호	말단	서열 모티프
164	N	Y-P-S-P-W-M-D-P
165	N	E-K-V-N-L-T-T-G-V-G-W
166	N	K-G-(I/V)-D-(V/I)
167	N	C-Q-N-S-K-L-x-N-G
168	N	N-L-T-L-A-V-(L/I/V)-N-G-(S/T)-(V/I)-P-x-W
169	N	S-W-(T/S)-x-D-T-(Y/F)-G
170	C	E-F-L-P-x-x-A-L-x-G-S-x-Q-P-R

도면68c

본 개시내용의 GH61 엔도글루카나제 모티프:

GH61 패밀리 엔도글루카나제의 모티프 1:

서열 번호 84: (I/L/M/V)-P-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-a-(H/N/Q)

GH61 패밀리 엔도글루카나제의 모티프 2:

서열 번호 85: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-a-(H/N/Q)

GH61 패밀리 엔도글루카나제의 모티프 3:

서열 번호 86: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-A-(H/N/Q)

GH61 패밀리 엔도글루카나제의 모티프 4:

서열 번호 87: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-A-(H/N/Q)

GH61 패밀리 엔도글루카나제의 모티프 5:

서열 번호 88: (F/W)-(T/F)-K-(A/I/V)

GH61 패밀리 엔도글루카나제의 모티프 6:

서열 번호 89: H-a-a-G-P-a-a-a-(Y/W)-(A/I/L/M/V)

GH61 패밀리 엔도글루카나제의 모티프 7:

서열 번호 90: H-a-G-P-a-a-a-(Y/W)-(A/I/L/M/V)

GH61 패밀리 엔도글루카나제의 모티프 8:

서열 번호 91: (E/Q)-a-Y-a-a-C-a-(E/H/Q/N)-(F/I/L/V)-a-(I/L/V)

도면69a

서열 번호 81

포도스포라 안세리나 Pa3C 뉴클레오타이드 서열

atggcataccgctcatttagctcttggggcgcttcgcctccacctctcttggccgagcgctcgtgacgcct
cgagatcctgttccgcttgattcgtcgtcgcctccatactatccagcgctcctcgtgaggtgggtcgt
tcgtgggaagaggtcttacagcaaggccgaagccttggctcgcagatgaccttggctgaaaagaccaac
atcacctcaggcattggcatctttatgggtgagttattaaccagacatggcttatataaaagcacaaga
gactgactgacatgtgaatagggtcagtgccaccaccctaatagagacgtttttctgattttgactaaca
catgatacgcctagtcctatgcgtaggaatactggaagcgcagaaagattgggttcccgcgcatgtgtc
ttcaggactctgcgttgggtgtgtcgtcggctgacaacgtcactgcgtttcctgctggcatcaccactg
gtgcaacgtttgacaagaagctgatctatgctcgtggtgttgcatttgggtgaagagcatcgcgcaagg
gcacaaatgtctatctgggtccttccgtagggcctcttgggcggaagccttgggtggcgcaactggg
agggtcttggatctgaccagttcttcaagccaaggctgctgccctgacgatcaaggcggttcaggaac
aaggcatcattgtactatcaagcatctgatcggcaacgagcaggagatgtatagaatgtacaaccct
tcagcctggatatacgccaatatttggtagtgagactcttgcctctttgacggactaaaaggctgactc
cccacagatgatcggaactctgcacgagctctacctgtggcctttgcccgaatccgtccatgcgggtgtt
gggtcggcaatgacagcttacaatgctgtaaacgggtctgcttgcctcagcacagctatctcatcaac
ggtattttgaaggatgagcttggattccagggtctcgtcatgtctgactggctgtcccacatctccgga
gtcgaactccggttggcaggtctcgacatgaacatgccaggtgacaccaacattcccctatttgggttc
agcaactggcactatgagctcagcagatcggttctcaacgggtctgtgctcttgacagactgaacgac
atggtcaccagaatcgtcgcgacatggtaacagttcggtcaggatagggaccaccaaggcctaacttc
tcgtcaaacaccggtgacgctgacgggtctgctttatcctgcagctctcttctccccaagggtcagggtg
aactggtttgtcaatgttcaggtgatcatttattgatcgccagagaggtcgccaggtgcatcacc
cttctcaagaacaatgggagcttcttccctgacgacttcgcagctctctccatgtcttcggtagctgt
gccaggtcaaccccgatgggcccacgcttgcatgaaccgcgctgcaacaaaggaacacttggcatg
ggctgggttctggttgcgattatccttacttggatgacccgatctcggctatcagggaagcgggtt
ccgcagctcaagttcttcaacaccgacggcttcccttggttccacccataccgctcgcggatgacgtt
gccatcgtgttcatcactccgatgctggagagaactcgttcactgttgagggaacaaacgggtgatcgc
aacagtgccaagctggctgctgtggcataacgggtgacgagctggtcaggaagactgcgagaagtacaac
aacgttatgtggtagctcaaacctgcccctctcgatctcgaatcctggatcgacaacccctcgcgtc
aagggcgctcctgtttcagcaccttcccggtcaagaagcgggcgagtcgttggccaacattctcttggc
gatgtctcccttagcgttcaacttccctactccatcacaagcgcgccaacgacttcccgacagcatc
gccaacctccgtggcttggcttgggtcaggtccaggacagctacagcgagggtgtacattgactac
cgctggctcaacaaggagaagatcagggccccttggcttggccaaggtctcagctacaccaacttc
tcgtttgatgccaccatcgagctctgctcactccactgtctctggttctcctcctgcccgtgccccagggtc
tcaacgcgggtgactcgaccgaatcccccccgctcagaggcggtactggccggaagggttcaacagg
atctgcgggtacctctactcctggtcacaagaacgacgcggataaacgctacgctgttgggtatcgc
gggtgagaagaagtataactatcccgctgggtacagcaccgcccagaagccgggtcccgacgcgggtggtc
ggggagggggglaactcctgcgcttgggatattgtcttccgtgtccagttacgggtcaagaacactggg
gatacgttctcgggacgggtcctcggtgacggcttatgttcagtatcctgaggggataccggtatgatac
cctgttgtgcagctgagggaactttgagaagacgaggttttggctccgggggaggaggagacgggtgacg
gttgagctgaccaggaaggacttgagcgtgtgggacacggagctgcagaactgggttgtgccgggggtt
ggggggaagaggtatacgttttgattggggagggcagcgataggttggttacggttgtttatacggat
acgggggtttgtgaggggggagggtgcccgtgtttaa

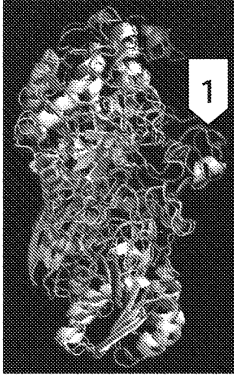
도면69b

서열 번호 80

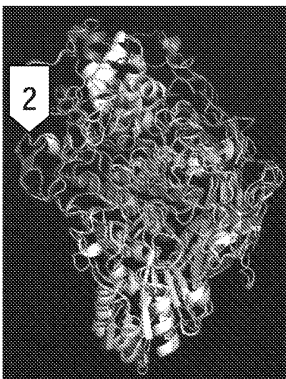
포도스포라 안세리나 Pa3C 단백질 서열

MAYRSLVLGFAFASTSLAASVVTPRDPVPPGFVAAPIYPAPHGGWVASWEEAYSKAELVSQMTLAEKTN
ITSGIGIFMGNITGSAERLGFPRMCLQDSALGVSSADNVTAFPAGITTGATFDKKLIYARGVAIGEEHRG
KGTNVYLGPSVGLGRKPLGGRNWEGFGSDPVLQAKAAALTIKGVQEQGIATIKHLIGNEQEMYRMYN
PFQPGYSANIDRTLHELILWPFABESVHAGVGSAMTAYNAVNGSACSQHSYLINGILKDELGFQGFVMS
DWLSHISGVDLSALAGLDMNMPGDTNIPLFGFSNWHYELSRSVLNGSVPLDRINDMVTTRIVATWYKFGQD
RDHPRPNFSSNTRDRDGLLYPAALFSPKQVNWVFNVDHYLIAREVAQDAITLLKNNGSFLPLTTSQ
SLHVFGTAAQVNPDPNACMNRACNKGTLGMGWGSGVADYPYLDPISAIRKRPDVKFFNTDGFPFWFH
PTPSPDDVAIVFITS DAGENSFTVEGNGDRNSAKLAWHNGDELVRKTAEKYNNVIVVAQTGVLDLLE
SWIDNPRVKGVLFQHLPGQEAGESLANILFGDVSFSGHLPSITKRANFPDSIANLRGFAFGQVQDTY
SEGLYIDYRWLNKEKIRPRFAFGHLSYTNFSFDATIESVTPLSLVPPARAPKGSTPVYSTEIPFASEA
YWPEGFNRIWRLYSLWLNKDADNAYAVGIAGVKKYNYPAGYSTAQKPGPAAGGGEGGNPALWDIAFRV
PVTVKNTGDTFSGRASVQAYVQYPEGIPYDTPVVQLRDFEKTRVLAPGEEETVTVELTRKDLVWDTTEL
QNWVVPGVGGKRYTVWIGEASDRLETTACTTDTGVCEGGRVPPV*

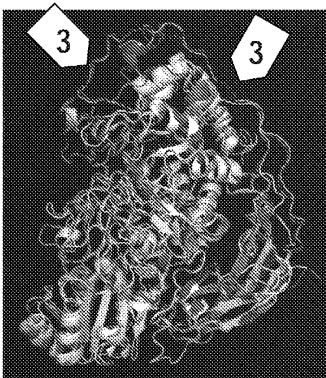
도면70a



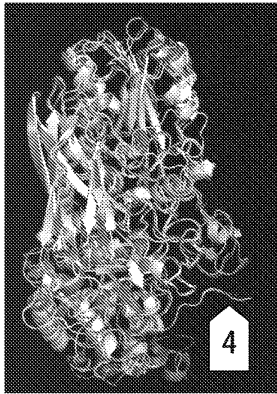
도면70b



도면70c



도면70d



도면70e

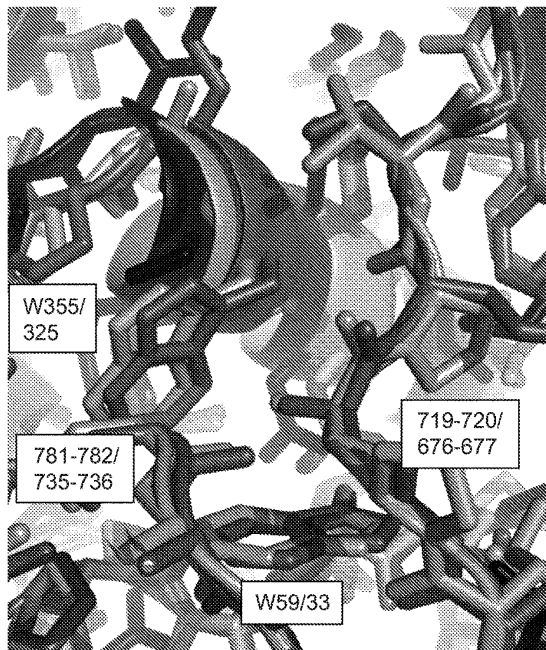
도 70e-1

트리코테트라 리세이 BGL1, Te3A 및 Fv3C의 정렬. 성숙 트리코테트라 리세이 BGL1 서열이 나타나 있으며, 성숙 Fv3C 및 ABG2의 시작은 'Mat'로 표시되어 있음								
	*	20	*	40	*	60	*	
Q12715_TRI :		-----		-----		VVPFAGTFWGT :	12	
ABG2_T_eme :		-----		-----		MRNGLKVAALAAASAVNGENLAYSPFFYPSPWAN	43	
Fv3C :		MKLNWVAALSIGAAGTDSAVLASAVPDTLAGVKKADAQKVTRD		TLAYSPPHYPSFPMD		PNAVWGEEA :	70	
				Mat		Ins4		
		80	*	100	*	120	*	140
Q12715_TRI :		YDKAKAALAKLNLQDKVIVSGVWNGGPCVGNTPSPAKISYPSLCLQDGLGVRYSTGSTAFTPGVQAA :						82
ABG2_T_eme :		YQKAVQFVSQTLAEKVNLTGTGWEQDRCVGQVGSIPRLGFPGLCMQDSPLGVRDTDYNSAFFAGVNV :						113
Fv3C :		YAKAKSVFSQTLMEKVNLTGTVGQGERCVGNVGSIPRLGMRGLCLQDGLGRLSDYNSAFFAGTTAG :						140
		*	160	*	180	*	200	*
Q12715_TRI :		STWDVNLIRERGQFIEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFVDPYLTGIAMGQTINGIQSVG :						152
ABG2_T_eme :		ATWDRNLAYRRGVAMGEEHKGKGVQVQLGPVAGPLGRSPDAGRNEGFADPDVLTGNMMASTIQGIQDAG :						183
Fv3C :		ASWSKSLWYERGLMGTEFKKGIDIALGPATGLGRTAAGGRNWEGFTVDPYMGHAMAEAVKGIQDAG :						210
		220	*	240	*	260	*	280
Q12715_TRI :		VQATAKHYYILNEQELNR-----ETISSNPDDRTLHELYTWFFADAVQANVASVMCSYNKVNNT :						209
ABG2_T_eme :		VIACAKHFILYEQEHFRQGAQ---DGYDISDSISANADDKTMHELYLWPFADAVRAGVGSVMCSYNQVNN :						250
Fv3C :		VIACAKHYIANEQEHFRQSGEVQSRKYNISELSNLDKTMHELYAWPFADAVRAGVGSVMCSYNQINN :						280
		삽입 1						
		*	300	*	320	*	340	*
Q12715_TRI :		TWACEDQYTLQTVLKDQLGFPYVMTDWNQHTTVQSANSGLDMSMPG-TDFNGNNRLWGPALTNVAVNS :						278
ABG2_T_eme :		SYACSNYSYTMNKLKSELGPGGFVMTDWGGHSGVGSALAGLDMSMPGDIADFSGTSFWGTNLTAVLVNG :						320
Fv3C :		SYGQNSKLLNGILKDEMGFGFVMSDWAQAHTGAASAVAGLDMSMPGDTAFDSGYSFWGGNLTAVLVNG :						350
		360	*	380	*	400	*	420
Q12715_TRI :		QVPTSRVDDMVTRILAAYLITGQDQAGYPSFNISR-----NVQGNHKTNVNR :						324
ABG2_T_eme :		SIPFWRVDDMAVRIMSAYYKVRDRYSVP-INFDSWTLDTYCPPEHYAVGQCGQTKINEHV						389
Fv3C :		TVPFWRVDDMALRIMSAPFFVKGKTIEDLPDINFSSWTRDTFGFVHTFAQENREQVNFVNVQHDHKS						420
		삽입 2						
		*	440	*	460	*	480	*
Q12715_TRI :		AIARDGIVLLKNDANIILPK-KPASIAVVGSAAIIGNHARNSPSCNDKCGDGLGMGWGSGAVNYPYFV :						393
ABG2_T_eme :		EIGAASAVLLKKNKG-LPLTGTTERFVGVFGKDA--GSNPGVCVNGCSDRGCDNGTLAMGWGSGTANFPYLV :						456
Fv3C :		EAAAKGSVVLKNTGS-LPLK-NPKFLAVIGEDA--GPNPAGPNGCGRGCDNGTLAMAWGSGTSQFPYLI :						486

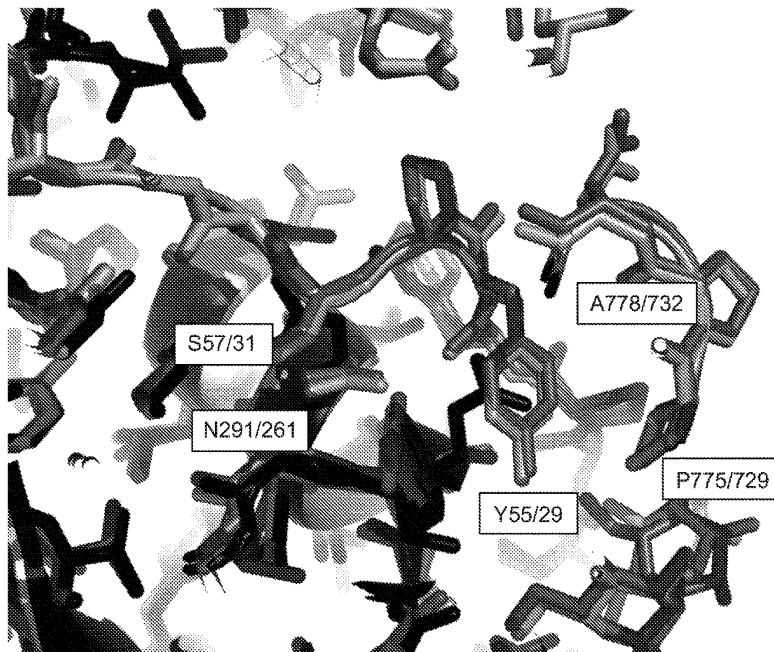
도 70e-2

트리코테트라 리세이 BGL1, Te3A 및 Fv3C의 정렬. 성숙 트리코테트라 리세이 BGL1 서열이 나타나 있으며, 성숙 Fv3C 및 ABG2의 시작은 'Mat'로 표시되어 있음					
	500	520	540	560	
Q12715_TRI	: APYDAINTRASSQGTQVT--LSNTDNTSSGASAGKDVAVFITADSGEGYITVEGNAGDRNNLDPWHN	:	461		
ABG2_T_eme	: TPEQAIQREVLNRNGTFTG-ITDNGALAEMAAAASQADTCLVFANADSGEGYITVDGNEGDRKNLTWQG	:	525		
Fv3C	: TPDQGLSNRATQDGTREYSILTNNEWASVQALVSQPNVTAIVFANADSGEGYIEVDGNFGDRKNLTWQG	:	556		
	580	600	620		
Q12715_TRI	: GNALVQAVAGANSNVIVVHVSVAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGVDSVPSGKLVYT	:	531		
ABG2_T_eme	: ADQVIHNVSANCNNTVVVLTHTVGPVLIDDWDHPNVTAIWLWAGLPGQESGNLVDVLYGRVNP-GKTPFT	:	594		
Fv3C	: GDELKNVSSICPNTIVVLHTVGPVLLADYEKNPNITAIVWAGLPGQESGNALVDVLYGRVNP-GRSFT	:	625		
	640	660	680	700	
Q12715_TRI	: IAKSPNDYNTRIVSGGSDS-----FSEGLFIDYKHFDAN-----ITPRYEFGLSYTKF	:	582		
ABG2_T_eme	: WGRARDDYGAPLIVKPNNGKGAQQDFTEGIFIDYRRFDKYN-----ITPIYEFGLSYTTF	:	652		
Fv3C	: WGRTRSYGTEVLYEANNRGAQQDFTEGSEGVFIDIRHFDRRSPSTDGKSSFNNTAALYEFHGLSWSTF	:	695		
	720	740	760		
Q12715_TRI	: NYSRLSV-----LSTAKSGPA	:	598		
ABG2_T_eme	: EFSQLNVQPINAPPYTPASGFTKAAQSFQO-PSNASDNLVPSDIERVLYIYPWLN--STDLKASANDPD	:	719		
Fv3C	: EYSDLNIQKNVENPYSPPAGQOTIPAPTFGNFSKNLNDYVFPKGVRYTYKFIYFPFLNTSSSASEASNDGGQ	:	765		
삽입 3					
	780	800	820	840	
Q12715_TRI	: TCAVV-----FGGPSDLFQNVATVTVDIANSQGVGTGAEVAQLYITYPSSAPRT	:	646		
ABG2_T_eme	: YGLPTEKYVPPNATNGDPQPIDPAGGAPGNGPSLYEPVARVTIITNTGKVTGDEVPLQYVSL--GGPDD	:	787		
Fv3C	: FGKTAQEEFLPPNALNGSAQPRLPASGAPGNGNPQLWDILYTVTATITNTGNATSDIPLQYVSL--GGENE	:	833		
부표 3					
	860	880	900		
Q12715_TRI	: PPKQLRGFAKL-NLTPGQSGTATFNIRRRDLSYWDTSQKWVPSGSGFISVGASSRDIRLTSTLSVA--	:	713		
ABG2_T_eme	: APKVLRGFDRI-TLAPGQYLWTTTLTRRDISNWDPTQNWVVTNYTKTIYVGNSSRNPLQAPLKPYPG	:	856		
Fv3C	: PIRVLRGFDRIENAPGQSAIFNAQLTRRDLNWDNTAQNWVITDHPKTVVWGSSSRKLPLSAKLE----	:	899		
Q12715_TRI	: - : -				
ABG2_T_eme	: I : 857				
Fv3C	: - : -				

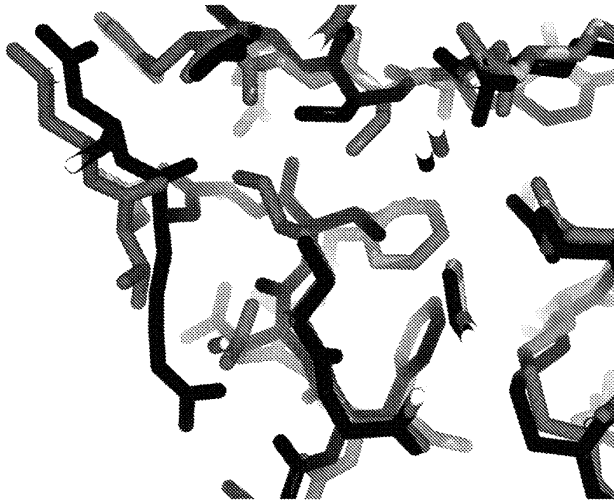
도면70f



도면70g

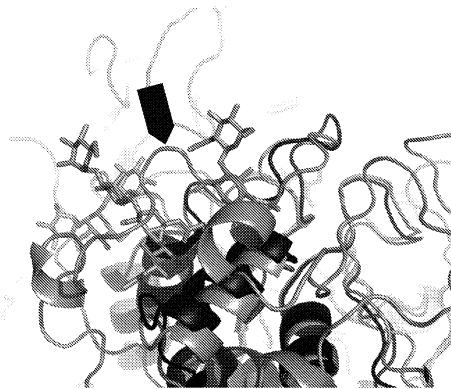


도면70h

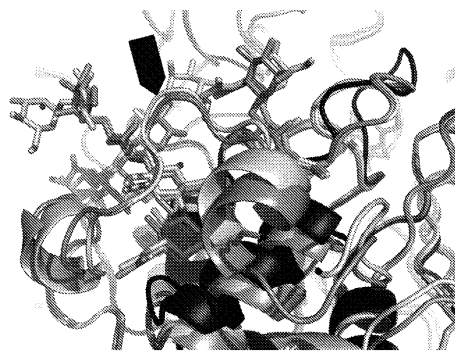


도면70i

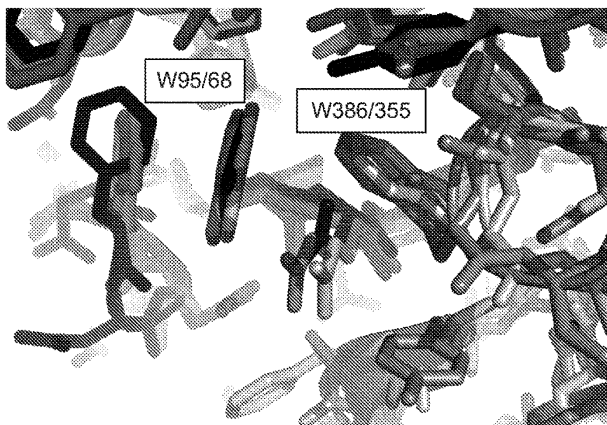
도 70i(a)



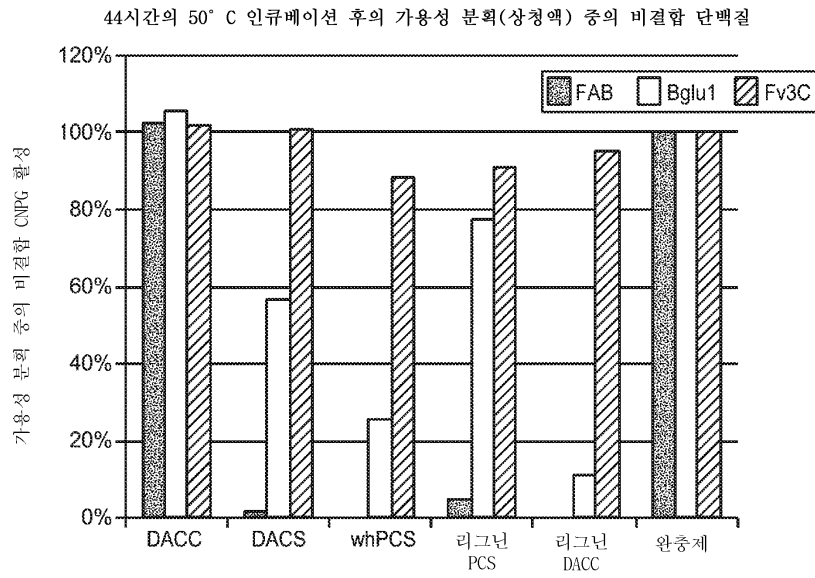
도 70i(b)



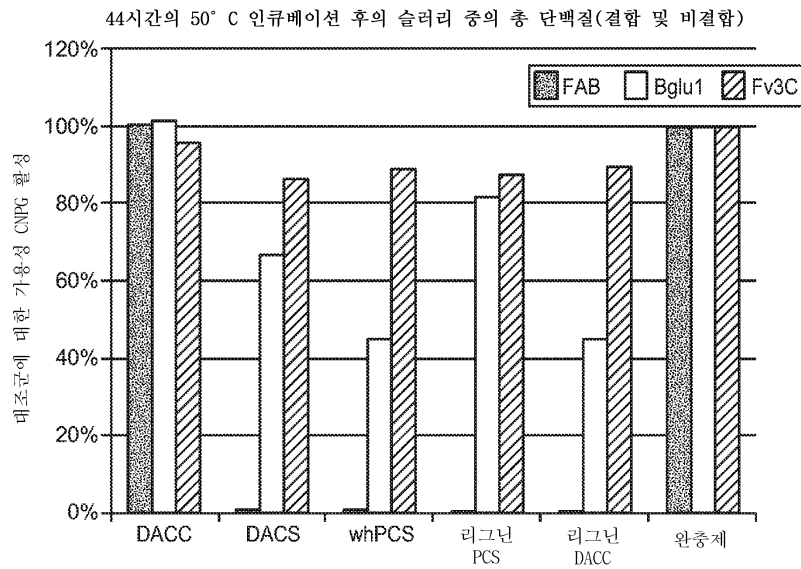
도면70j



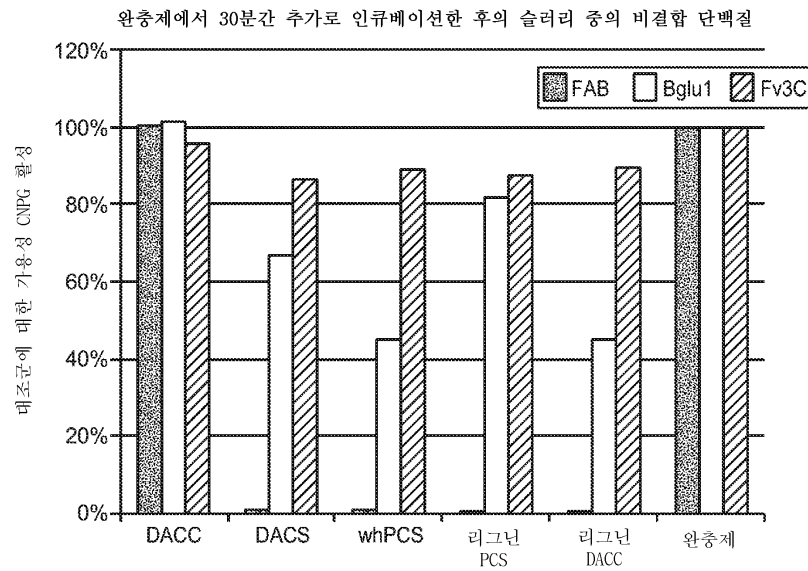
도면71a



도면71b



도면71c



서열목록

SEQUENCE LISTING

- <110> Danisco US Inc.
Kaper, Thijs
Nikolaev, Igor
Lantz, Suzanne
Fujdala, Meredith K.
Hsi, Megan Y.
- <120> Cellulase Compositions and Methods of Using the Same for Improved Conversion of Lignocellulosic Biomass into Fermentable Sugars
- <130> 31517-WO
- <140> PCT/US12/29498
- <141> 2012-03-16
- <150> US 61/453,918
- <151> 2011-03-17
- <160> 178
- <170> PatentIn version 3.5
- <210> 1
- <211> 2358
- <212> DNA

<213> Fusarium verticillioides

<400> 1

atgctgctca atcttcaggt cgctgccagc gctttgtcgc tttctctttt aggtggattg	60
gctgaggctg ctacgccata tacccttccg gactgtacca aaggaccttt gagcaagaat	120
ggaatctgcg atacttcgtt atctccagct aaaagagcgg ctgctctagt tgctgctctg	180
acgcccgaag agaaggtggg caatctggtc aggtaaaata tcccccccc cataatcact	240
attcggagat tggagctgac ttaacgcagc aatgcaactg gtgcaccaag aatcggactt	300
ccaaggtaca actggtggaa cgaagccctt catggcctcg ctggatctcc aggtggtcgc	360
tttgccgaca ctctcccta cgacgcggcc acatcatttc ccatgcctct tctcatggcc	420
gctgctttcg acgatgatct gatccacgat atcggaacg tcgtcggcac cgaagcgcgt	480
gcgttacta acggcgggtg gcgcggagtc gacttctgga caccacacgt caaccctttt	540
aaagatcctc gctggggctg tggctccgaa actccagggtg aagatgcctt tcatgtcagc	600
cggatgctc gctatatcgt caggggtctc gaaggcgata aggagcaacg acgtattgtt	660
gctacctgca agcactatgc tggaaacgac tttagaggact ggggaggctt cagcgtcac	720
gactttgatg ccaagattac tctcaggac ttggctgagt actacgtcag gcctttccag	780
gagtgcacc gtgatgaaa ggttggttcc atcatgtgcg cctacaatgc cgtgaacggc	840
attcccgcat gcgcaaactc gtatctgcag gagacgatcc tcagagggca ctggaactgg	900
acgcgcgata acaactggat cactagtgat tgtggcgcca tgcaggatat ctggcagaat	960
cacaagtatg tcaagaccaa cgctgaaggt gcccaggtag cttttgagaa cggcatggat	1020
tctagctgcg agtatactac taccagcgtat gtctccgatt cgtacaagca aggcctcttg	1080
actgagaagc tcatggatcg ttcggtgaag cgccttttcg aagggtctgt tcatactggt	1140
ttctttgacg gtgcccgaag gcaatggaac tcgctcagtt ttgcggatgt caacaccaag	1200
gaagctcagg atcttgcaat cagatctgct gtggagggtg ctgttcttct taagaatgac	1260
ggcactttgc ctctgaagct caagaagaag gatagtgttg caatgatcgg attctgggcc	1320
aacgatattt ccaagctgca ggggtggttac agtggacgtg ctccgttctt ccacagcccc	1380
ctttatgcag ctgagaagct tggctctgac accaacgtgg cttgggggtcc gacactgcag	1440
aacagctcat ctcatgataa ctggaccacc aatgctgttg ctgcggcgaa gaagtctgat	1500
tacattctct acittgggtg tcttgacgcc tctgctgctg gcgaggacag agatcgtgag	1560
aaccttgact ggcctgagag ccagctgacc cttcttcaga agctctctag tctcggaag	1620
ccactgggtg ttatccagct tggatgataa gtcgatgaca ccgtctttt gaagaacaag	1680
aagattaaca gtattctttg ggtcaattac cctggtcagg atggcggcac tgcagtcagt	1740

gacctgctca ctggacgaaa gagtccctgct ggccgactac ccgtcacgca atatcccagt 1800
aaatacactg agcagattgg catgactgac atggacctca gacctacaa gtcgttgcca 1860
gggagaactt atcgttggtta ctcaactcca gttcttccct acggcttttg cctccactac 1920
accaagttcc aagccaagtt caagtcacaac aagttgacgt ttgacatcca gaagcttctc 1980
aagggtgca gtgctcaata ctccgatact tgcgcgtgc ccccatcca agttagtgc 2040

aagaacaccg gccgcattac ctccgacttt gtctctctgg tctttatcaa gagtgaagtt 2100
ggacctaagc cttaccctct caagaccctt gcggcttatg gtcgcttgca tgatgtcgcg 2160
ccttcacga cgaaggatat ctactggag tggacgttg ataacattgc gcgacgggga 2220
gagaatggtg atttggttgt ttatcctggg acttacactc tgttgctgga tgagcctacg 2280
caagccaaga tccaggttac gctgactgga aagaaggcta ttttgataa gtggcctcaa 2340
gacccaagt ctgcgtaa 2358

<210> 2

<211> 766

<212> PRT

<213> Fusarium verticillioides

<400> 2

Met Leu Leu Asn Leu Gln Val Ala Ala Ser Ala Leu Ser Leu Ser Leu

1 5 10 15

Leu Gly Gly Leu Ala Glu Ala Ala Thr Pro Tyr Thr Leu Pro Asp Cys

20 25 30

Thr Lys Gly Pro Leu Ser Lys Asn Gly Ile Cys Asp Thr Ser Leu Ser

35 40 45

Pro Ala Lys Arg Ala Ala Ala Leu Val Ala Ala Leu Thr Pro Glu Glu

50 55 60

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

65 70 75 80

Leu Pro Arg Tyr Asn Trp Trp Asn Glu Ala Leu His Gly Leu Ala Gly

85 90 95

Ser Pro Gly Gly Arg Phe Ala Asp Thr Pro Pro Tyr Asp Ala Ala Thr

100 105 110

Ser Phe Pro Met Pro Leu Leu Met Ala Ala Ala Phe Asp Asp Asp Leu

115	120	125
Ile His Asp Ile Gly Asn Val Val Gly Thr Glu Ala Arg Ala Phe Thr		
130	135	140
Asn Gly Gly Trp Arg Gly Val Asp Phe Trp Thr Pro Asn Val Asn Pro		
145	150	155
Phe Lys Asp Pro Arg Trp Gly Arg Gly Ser Glu Thr Pro Gly Glu Asp		
165	170	175
Ala Leu His Val Ser Arg Tyr Ala Arg Tyr Ile Val Arg Gly Leu Glu		
180	185	190
Gly Asp Lys Glu Gln Arg Arg Ile Val Ala Thr Cys Lys His Tyr Ala		
195	200	205
Gly Asn Asp Phe Glu Asp Trp Gly Gly Phe Thr Arg His Asp Phe Asp		
210	215	220
Ala Lys Ile Thr Pro Gln Asp Leu Ala Glu Tyr Tyr Val Arg Pro Phe		
225	230	235
Gln Glu Cys Thr Arg Asp Ala Lys Val Gly Ser Ile Met Cys Ala Tyr		
245	250	255
Asn Ala Val Asn Gly Ile Pro Ala Cys Ala Asn Ser Tyr Leu Gln Glu		
260	265	270
Thr Ile Leu Arg Gly His Trp Asn Trp Thr Arg Asp Asn Asn Trp Ile		
275	280	285
Thr Ser Asp Cys Gly Ala Met Gln Asp Ile Trp Gln Asn His Lys Tyr		
290	295	300
Val Lys Thr Asn Ala Glu Gly Ala Gln Val Ala Phe Glu Asn Gly Met		
305	310	315
Asp Ser Ser Cys Glu Tyr Thr Thr Thr Ser Asp Val Ser Asp Ser Tyr		
325	330	335
Lys Gln Gly Leu Leu Thr Glu Lys Leu Met Asp Arg Ser Leu Lys Arg		
340	345	350
Leu Phe Glu Gly Leu Val His Thr Gly Phe Phe Asp Gly Ala Lys Ala		
355	360	365

Gln Trp Asn Ser Leu Ser Phe Ala Asp Val Asn Thr Lys Glu Ala Gln
370 375 380

Asp Leu Ala Leu Arg Ser Ala Val Glu Gly Ala Val Leu Leu Lys Asn
385 390 395 400

Asp Gly Thr Leu Pro Leu Lys Leu Lys Lys Asp Ser Val Ala Met
405 410 415

Ile Gly Phe Trp Ala Asn Asp Thr Ser Lys Leu Gln Gly Gly Tyr Ser
420 425 430

Gly Arg Ala Pro Phe Leu His Ser Pro Leu Tyr Ala Ala Glu Lys Leu
435 440 445

Gly Leu Asp Thr Asn Val Ala Trp Gly Pro Thr Leu Gln Asn Ser Ser
450 455 460

Ser His Asp Asn Trp Thr Thr Asn Ala Val Ala Ala Ala Lys Lys Ser
465 470 475 480

Asp Tyr Ile Leu Tyr Phe Gly Gly Leu Asp Ala Ser Ala Ala Gly Glu
485 490 495

Asp Arg Asp Arg Glu Asn Leu Asp Trp Pro Glu Ser Gln Leu Thr Leu
500 505 510

Leu Gln Lys Leu Ser Ser Leu Gly Lys Pro Leu Val Val Ile Gln Leu
515 520 525

Gly Asp Gln Val Asp Asp Thr Ala Leu Leu Lys Asn Lys Lys Ile Asn
530 535 540

Ser Ile Leu Trp Val Asn Tyr Pro Gly Gln Asp Gly Gly Thr Ala Val
545 550 555 560

Met Asp Leu Leu Thr Gly Arg Lys Ser Pro Ala Gly Arg Leu Pro Val
565 570 575

Thr Gln Tyr Pro Ser Lys Tyr Thr Glu Gln Ile Gly Met Thr Asp Met
580 585 590

Asp Leu Arg Pro Thr Lys Ser Leu Pro Gly Arg Thr Tyr Arg Trp Tyr
595 600 605

Ser Thr Pro Val Leu Pro Tyr Gly Phe Gly Leu His Tyr Thr Lys Phe

610 615 620
 Gln Ala Lys Phe Lys Ser Asn Lys Leu Thr Phe Asp Ile Gln Lys Leu
 625 630 635 640

 Leu Lys Gly Cys Ser Ala Gln Tyr Ser Asp Thr Cys Ala Leu Pro Pro
 645 650 655
 Ile Gln Val Ser Val Lys Asn Thr Gly Arg Ile Thr Ser Asp Phe Val
 660 665 670
 Ser Leu Val Phe Ile Lys Ser Glu Val Gly Pro Lys Pro Tyr Pro Leu
 675 680 685
 Lys Thr Leu Ala Ala Tyr Gly Arg Leu His Asp Val Ala Pro Ser Ser
 690 695 700

Thr Lys Asp Ile Ser Leu Glu Trp Thr Leu Asp Asn Ile Ala Arg Arg
 705 710 715 720
 Gly Glu Asn Gly Asp Leu Val Val Tyr Pro Gly Thr Tyr Thr Leu Leu
 725 730 735
 Leu Asp Glu Pro Thr Gln Ala Lys Ile Gln Val Thr Leu Thr Gly Lys
 740 745 750
 Lys Ala Ile Leu Asp Lys Trp Pro Gln Asp Pro Lys Ser Ala
 755 760 765

<210> 3

<211> 1338

<212> DNA

<213> Penicillium funiculosum

<400> 3

atgcttcagc gatttgctta tattttacca ctggctctat tgagtgttg agtgaaagcc 60
 gacaaccct ttgtgcagag catctacacc gctgatccgg caccgatggt atacaatgac 120
 cgcgtttatg tcttcatgga ccatgacaac accggagcta cctactaaa catgacagac 180
 tggcatctgt tctcgtcagc agatatggcg aattggcaag atcatggcat tccaatgagc 240
 ctggccaatt tcacctgggc caacgcgaat gcgtgggccc cgcaagtcac ccctcgcaac 300
 ggccaattct acttttatgc tctgtgccga cacaacgatg gttctatggc tatcggtgtg 360

 ggagtgagca gcaccatcac aggtccatac catgatgcta tcggcaaacc gctagtagag 420

aacaacgaga ttgatccac cgtgttcac gacgatgacg gtcaggcata cctgtactgg 480
 ggaaatccag acctgtggta cgtcaaattg aaccaagata tgatatcgta cagcgggagc 540
 cctactcaga ttccactcac caccgctgga ttgggtactc gaacgggcaa tgctcaacgg 600
 ccgaccactt ttgaagaagc tccatgggta tacaacgca acggcatcta ctatatcgcc 660
 tatgcagccg attgtttgtc tgaggatatt cgctactcca cggaaccag tgccactggt 720
 ccgtggactt atcgaggcgt catcatgccg acccaaggta gcagcttcac caatcacgag 780

ggtattatcg acttccagaa caactcctac tttttctatc acaacggcgc tcttcccggc 840
 ggaggcggct accaacgacg tgtatgtgtg gagcaattca aatacaatgc agatggaacc 900
 attccgacga tcgaaatgac caccgccggt ccagctcaaa ttgggactct caacccttac 960
 gtgcgacagg aagccgaaac ggcggcatgg tcttcaggca tcactacgga ggtttgtagc 1020
 gaaggcggaa ttgacgtcgg gtttatcaac aatggcgatt acatcaaagt taaaggcgta 1080
 gctttcgggt caggagccca tttttctca gcgcgggttg cttctgcaaa tagcggcggc 1140
 actattgcaa tacacctcgg aagcacaact ggtacgtcgt tgggcacttg tactgtcccc 1200

agcactggcg gttggcagac ttggactacc gttacctgtt ctgtcagtgg cgcactctggg 1260
 acccaggatg tgtattttgt tttcgggtgg agcgaacag gatactgtt caactttgat 1320
 tattggcagt tcgcataa 1338

<210> 4

<211> 445

<212> PRT

<213> Penicillium funiculosum

<400> 4

Met Leu Gln Arg Phe Ala Tyr Ile Leu Pro Leu Ala Leu Leu Ser Val

1 5 10 15

Gly Val Lys Ala Asp Asn Pro Phe Val Gln Ser Ile Tyr Thr Ala Asp

20 25 30

Pro Ala Pro Met Val Tyr Asn Asp Arg Val Tyr Val Phe Met Asp His

35 40 45

Asp Asn Thr Gly Ala Thr Tyr Tyr Asn Met Thr Asp Trp His Leu Phe

50 55 60

Ser Ser Ala Asp Met Ala Asn Trp Gln Asp His Gly Ile Pro Met Ser

65 70 75 80

Leu Ala Asn Phe Thr Trp Ala Asn Ala Asn Ala Trp Ala Pro Gln Val

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

Glu Val Cys Ser Glu Gly Gly Ile Asp Val Gly Phe Ile Asn Asn Gly
340 345 350

Asp Tyr Ile Lys Val Lys Gly Val Ala Phe Gly Ser Gly Ala His Ser
355 360 365

Phe Ser Ala Arg Val Ala Ser Ala Asn Ser Gly Gly Thr Ile Ala Ile
370 375 380

His Leu Gly Ser Thr Thr Gly Thr Leu Val Gly Thr Cys Thr Val Pro
385 390 395 400

Ser Thr Gly Gly Trp Gln Thr Trp Thr Thr Val Thr Cys Ser Val Ser
405 410 415

Gly Ala Ser Gly Thr Gln Asp Val Tyr Phe Val Phe Gly Gly Ser Gly
420 425 430

Thr Gly Tyr Leu Phe Asn Phe Asp Tyr Trp Gln Phe Ala
435 440 445

<210> 5

<211> 1593

<212> DNA

<213> *Fusarium verticillioides*

<400> 5

atgaaggat acigcgtcgt ggcgtgggcc acttctttga cgccggcact ggctggcttg 60
attggacacc gtcgcgccac cacttcaac aatcctatca tctactcaga ctttcagat 120
aacgatgtat tctcgggtcc agataactac tactacttct ctgcttccaa cttccacttc 180

agcccaggag caccggtttt gaagtctaaa gatctgctaa actgggatct catcgcccat 240
tcaattcccc gctgaactt tggcgacggc tatgatcttc ctctggctc acgttattac 300
cgtggaggta ctggggcatc atccctcaga tacagaaaga gcaatggaca gtgtactgg 360
atcgggtgca tcaacttctg gcagacctgg gtataactg cctcatcgcc ggaaggtcca 420
tggtacaaca agggaaactt cggtgataac aattgctact acgacaatgg catactgatc 480
gatgacgatg ataccatgta tgtctgatac ggttccggtg aggtcaaagt atctcaacta 540
tctcaggacg gattcagcca ggtcaaatct caggtagttt tcaagaacac tgatattggg 600

gtccaagact tggagggtta ccgcatgtac aagatcaacg ggctctacta tatectaaac 660
gatagcccaa gtggcagtca gacctggatt tggaagtcga aatcacctg gggcccttat 720

gagtctaagg tctcgcgca caaagtcacc ccgcctatct ctggtggtaa ctgcccgcac 780
cagggtagtc tcataaagac tcccaatggt ggctgggtact tcatgtcatt cacttggggc 840
tatactgccg gccgtcttcc ggttcttgca ccgattacgt ggggtagcga tggtttcccc 900
attcttgtca aggggtgctaa tggcggatgg ggatcatctt acccaacact tcctggcacg 960
gatggtgtga caaagaattg gacaaggact gataccttcc gcggaacctc acttgctccg 1020

tcctgggagt ggaaccataa tccggacgtc aactccttca ctgtcaacaa cggcctgact 1080
ctccgcactg ctagcattac gaaggatatt taccaggcga ggaacacgct atctcaccga 1140
actcatggtg atcatccaac aggaatagtg aagattgatt tctctccgat gaaggacggc 1200
gaccggggccg ggctttcagc gtttcgagac caaagtgcac acatcggtat tcatcgagat 1260
aacggaaagt tcacaatcgc tacgaagcat gggatgaata tggatgagtg gaacggaaca 1320
acaacagacc tgggacaaat aaaagccaca gctaattgtc cttctggaag gaccaagatc 1380
tggtcgagac ttcaacttga taccaaccca gcaggaactg gcaacactat cttttcttac 1440

agttgggatg gagtcaagta tgaaacactg ggtcccaact tcaaactgta caatggttgg 1500
gcattcttta ttgcttaccg attcggcatc ttcaacttgc ccgagacggc tttaggaggc 1560
tcgatcaagg ttgagtcctt cacagctgca tag 1593

<210> 6

<211> 530

<212> PRT

<213> *Fusarium verticillioides*

<400> 6

Met Lys Val Tyr Trp Leu Val Ala Trp Ala Thr Ser Leu Thr Pro Ala

1 5 10 15

Leu Ala Gly Leu Ile Gly His Arg Arg Ala Thr Thr Phe Asn Asn Pro

20 25 30

Ile Ile Tyr Ser Asp Phe Pro Asp Asn Asp Val Phe Leu Gly Pro Asp

35 40 45

Asn Tyr Tyr Tyr Phe Ser Ala Ser Asn Phe His Phe Ser Pro Gly Ala

50 55 60

Pro Val Leu Lys Ser Lys Asp Leu Leu Asn Trp Asp Leu Ile Gly His

65 70 75 80

Ser Ile Pro Arg Leu Asn Phe Gly Asp Gly Tyr Asp Leu Pro Pro Gly

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

Ser Leu Ala Pro Ser Trp Glu Trp Asn His Asn Pro Asp Val Asn Ser
340 345 350

Phe Thr Val Asn Asn Gly Leu Thr Leu Arg Thr Ala Ser Ile Thr Lys
355 360 365

Asp Ile Tyr Gln Ala Arg Asn Thr Leu Ser His Arg Thr His Gly Asp
370 375 380

His Pro Thr Gly Ile Val Lys Ile Asp Phe Ser Pro Met Lys Asp Gly
385 390 395 400

Asp Arg Ala Gly Leu Ser Ala Phe Arg Asp Gln Ser Ala Tyr Ile Gly
405 410 415

Ile His Arg Asp Asn Gly Lys Phe Thr Ile Ala Thr Lys His Gly Met
420 425 430

Asn Met Asp Glu Trp Asn Gly Thr Thr Thr Asp Leu Gly Gln Ile Lys
435 440 445

Ala Thr Ala Asn Val Pro Ser Gly Arg Thr Lys Ile Trp Leu Arg Leu
450 455 460

Gln Leu Asp Thr Asn Pro Ala Gly Thr Gly Asn Thr Ile Phe Ser Tyr
465 470 475 480

Ser Trp Asp Gly Val Lys Tyr Glu Thr Leu Gly Pro Asn Phe Lys Leu
485 490 495

Tyr Asn Gly Trp Ala Phe Phe Ile Ala Tyr Arg Phe Gly Ile Phe Asn
500 505 510

Phe Ala Glu Thr Ala Leu Gly Gly Ser Ile Lys Val Glu Ser Phe Thr
515 520 525

Ala Ala
530

<210> 7

<211> 1374

<212> DNA

<213> Fusarium verticillioides

<400> 7

atgcactacg ctaccctcac cactttgggtg ctggctctga ccaccaacgt cgctgcacag 60

caaggcacag caactgtcga cctctccaaa aatcatggac cggcgaaggc ccttggttca 120

ggcttcatat acggctggcc tgacaacgga acaagcgtcg acacctccat accagatttc 180

ttggttaactg acatcaaatt caactcaaac cgcggcgggtg gcgcccaaat cccatcactg 240

ggttgggcca gaggtggcta tgaaggatac ctggccgct tcaactcaac cttatccaac 300

tatcgacca cgcgcaagta taacgtgac tttatcttgt tgcctcatga cctctggggt 360

gcggatggcg ggcagggttc aaactccccg tttctggcg acaatggcaa ttggactgag 420

atggagttaa tctggaatca gcttgtgtct gacttgaagg ctcataatat gctggaaggt 480

cttgtgattg atgtttggaa tgagcctgat attgatatct tttgggatcg cccgtggtcg 540

cagtttcttg agtattacaa tcgcgcgacc aaactacttc ggtgagtcta ctactgatcc 600

atacgtatatt acagtgagct gactggtcga attagaaaaa cacttcccaa aactcttctc 660

agtggcccag ccatggcaca ttctccatt ctgtccgatg ataaatggca tacctggctt 720

caatcagtag cgggtaacaa gacagtcct gatatttact cctggcatca gattggcgct 780

tgggaacgtg agccggacag cactatcccc gactttacca ccttgcgggc gcaatatggc 840

gttcccgaga agccaattga cgtcaatgag tacgctgcac gcgatgagca aaatccagcc 900

aactccgtct actacctctc tcaactagag cgtcataacc ttagaggctct tcgcgcaaac 960

tggggtagcg gatctgacct ccacaactgg atgggcaact tgatttacag cactaccggt 1020

acctcgagg ggacttacta ccctaattgt gaatggcagg cttacaagta ctatgcggcc 1080

atggcagggc agagacttgt gaccaaagca tcgtcggact tgaagtttga tgtctttgcc 1140

actaagcaag gccgtaagat taagattata gccggcacga ggaccgttca agcaaagtat 1200

aacatcaaaa tcagcggttt ggaagtagca ggacttccta agatgggtac ggtaaaggtc 1260

cggacttata ggttcgactg ggctgggccg aatggaaagg ttgacgggcc tgttgatttg 1320

ggggagaaga agtatactta ttcggccaat acggtgagca gccctctac ttga 1374

<210> 8

<211> 439

<212> PRT

<213> Fusarium verticillioides

<400> 8

Met His Tyr Ala Thr Leu Thr Thr Leu Val Leu Ala Leu Thr Thr Asn

1 5 10 15

Val Ala Ala Gln Gln Gly Thr Ala Thr Val Asp Leu Ser Lys Asn His

20 25 30
 Gly Pro Ala Lys Ala Leu Gly Ser Gly Phe Ile Tyr Gly Trp Pro Asp
 35 40 45

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

 Leu Leu Pro His Asp Leu Trp Gly Ala Asp Gly Gly Gln Gly Ser Asn
 115 120 125
 Ser Pro Phe Pro Gly Asp Asn Gly Asn Trp Thr Glu Met Glu Leu Phe
 130 135 140
 Trp Asn Gln Leu Val Ser Asp Leu Lys Ala His Asn Met Leu Glu Gly
 145 150 155 160
 Leu Val Ile Asp Val Trp Asn Glu Pro Asp Ile Asp Ile Phe Trp Asp
 165 170 175

 Arg Pro Trp Ser Gln Phe Leu Glu Tyr Tyr Asn Arg Ala Thr Lys Leu
 180 185 190
 Leu Arg Lys Thr Leu Pro Lys Thr Leu Leu Ser Gly Pro Ala Met Ala
 195 200 205
 His Ser Pro Ile Leu Ser Asp Asp Lys Trp His Thr Trp Leu Gln Ser
 210 215 220
 Val Ala Gly Asn Lys Thr Val Pro Asp Ile Tyr Ser Trp His Gln Ile
 225 230 235 240

 Gly Ala Trp Glu Arg Glu Pro Asp Ser Thr Ile Pro Asp Phe Thr Thr
 245 250 255
 Leu Arg Ala Gln Tyr Gly Val Pro Glu Lys Pro Ile Asp Val Asn Glu
 260 265 270

Tyr Ala Ala Arg Asp Glu Gln Asn Pro Ala Asn Ser Val Tyr Tyr Leu
 275 280 285
 Ser Gln Leu Glu Arg His Asn Leu Arg Gly Leu Arg Ala Asn Trp Gly
 290 295 300

 Ser Gly Ser Asp Leu His Asn Trp Met Gly Asn Leu Ile Tyr Ser Thr
 305 310 315 320
 Thr Gly Thr Ser Glu Gly Thr Tyr Tyr Pro Asn Gly Glu Trp Gln Ala
 325 330 335
 Tyr Lys Tyr Tyr Ala Ala Met Ala Gly Gln Arg Leu Val Thr Lys Ala
 340 345 350
 Ser Ser Asp Leu Lys Phe Asp Val Phe Ala Thr Lys Gln Gly Arg Lys
 355 360 365

 Ile Lys Ile Ile Ala Gly Thr Arg Thr Val Gln Ala Lys Tyr Asn Ile
 370 375 380
 Lys Ile Ser Gly Leu Glu Val Ala Gly Leu Pro Lys Met Gly Thr Val
 385 390 395 400
 Lys Val Arg Thr Tyr Arg Phe Asp Trp Ala Gly Pro Asn Gly Lys Val
 405 410 415
 Asp Gly Pro Val Asp Leu Gly Glu Lys Lys Tyr Thr Tyr Ser Ala Asn
 420 425 430

 Thr Val Ser Ser Pro Ser Thr
 435

 <210> 9
 <211> 1350
 <212> DNA
 <213> *Fusarium verticillioides*
 <400> 9

 atgtggctga cctccccatt gctgttcgcc agcaccctcc tgggcctcac tggcgttgct 60
 ctagcagaca accccatcgt ccaagacatc tacaccgcag acccagcacc aatggctctac 120
 aatggccgcg tctacctctt cacaggccat gacaacgacg gctctaccga cttcaacatg 180
 acagactggc gtctcttctc gtcagcagac atggtcaact ggcagcacca tgggtgtcccc 240

atgagcttaa agaccttcag ctgggccaac agcagagcct gggctgttca agtcgttgcc 300

cgaaacggaa agttttactt ctatgttcct gtcgtaatg ccaagacggg tggaatggct 360

attggtgtcg gtgttagtac caacatcctt gggccctaca ctgatgccct tggaaagcca 420

ttggtcgaga acaatgagat cgacccaact gtctacatcg aactgatgg ccaggcctat 480

ctctactggg gcaaccctgg attgtactac gtcaagctca accaagacat gctctctac 540

agtggtagca tcaacaaagt atcgtcaca acagctggat tcggcagccg cccgaacaac 600

gcgcagcgtc ctactacttt cgaggaagga ccgtggctgt acaagcgtgg aaatctctac 660

tacatgatct acgcagccaa ctgctgttcc gaggacattc gctactcaac tggaccagc 720

gccactggac ctiggactta ccgcggtgtc gtgatgaaca aggcgggtcg aagcttcacc 780

aaccatcctg gcatcatcga ctttgagaac aactcgtact tctttacca caatggcgct 840

cttgatggag gtagcggtta tactcggctc gtggctgtcg agagcttcaa gtatggttcg 900

gacggctcga tccccgagat caagatgact acgcaaggcc cagcgagct caagtctctg 960

aaccatgatg tcaagcagga ggccgagact atcgccctgg ctgagggtat cgagactgag 1020

gtctgcagcg aaggtggctc caacgttgc ttcacgcaca atggtgacta catcaagtc 1080

aaggagtcg actttggcag caccggtgca aagacgttca gcgcccgtgt tgcttccaac 1140

agcagcggag gcaagattga gcttcgactt ggtagcaaga ccggttaagt ggttggtacc 1200

tgcacggtaa cgactacggg aaactggcag acttataaga ctgtggattg ccccgtcagt 1260

ggtgctactg gtacagcga tctattcttt gtcttcacgg gctctgggtc tggtctctg 1320

ttcaacttca actggtggca gtttagctaa 1350

<210> 10

<211> 449

<212> PRT

<213> *Fusarium verticillioides*

<400> 10

Met Trp Leu Thr Ser Pro Leu Leu Phe Ala Ser Thr Leu Leu Gly Leu

1 5 10 15

Thr Gly Val Ala Leu Ala Asp Asn Pro Ile Val Gln Asp Ile Tyr Thr

20 25 30

Ala Asp Pro Ala Pro Met Val Tyr Asn Gly Arg Val Tyr Leu Phe Thr

35 40 45

Gly His Asp Asn Asp Gly Ser Thr Asp Phe Asn Met Thr Asp Trp Arg

50 55 60
 Leu Phe Ser Ser Ala Asp Met Val Asn Trp Gln His His Gly Val Pro
 65 70 75 80

 Met Ser Leu Lys Thr Phe Ser Trp Ala Asn Ser Arg Ala Trp Ala Gly
 85 90 95
 Gln Val Val Ala Arg Asn Gly Lys Phe Tyr Phe Tyr Val Pro Val Arg
 100 105 110
 Asn Ala Lys Thr Gly Gly Met Ala Ile Gly Val Gly Val Ser Thr Asn
 115 120 125
 Ile Leu Gly Pro Tyr Thr Asp Ala Leu Gly Lys Pro Leu Val Glu Asn
 130 135 140

 Asn Glu Ile Asp Pro Thr Val Tyr Ile Asp Thr Asp Gly Gln Ala Tyr
 145 150 155 160
 Leu Tyr Trp Gly Asn Pro Gly Leu Tyr Tyr Val Lys Leu Asn Gln Asp
 165 170 175
 Met Leu Ser Tyr Ser Gly Ser Ile Asn Lys Val Ser Leu Thr Thr Ala
 180 185 190
 Gly Phe Gly Ser Arg Pro Asn Asn Ala Gln Arg Pro Thr Thr Phe Glu
 195 200 205

 Glu Gly Pro Trp Leu Tyr Lys Arg Gly Asn Leu Tyr Tyr Met Ile Tyr
 210 215 220
 Ala Ala Asn Cys Cys Ser Glu Asp Ile Arg Tyr Ser Thr Gly Pro Ser
 225 230 235 240
 Ala Thr Gly Pro Trp Thr Tyr Arg Gly Val Val Met Asn Lys Ala Gly
 245 250 255
 Arg Ser Phe Thr Asn His Pro Gly Ile Ile Asp Phe Glu Asn Asn Ser
 260 265 270

 Tyr Phe Phe Tyr His Asn Gly Ala Leu Asp Gly Gly Ser Gly Tyr Thr
 275 280 285
 Arg Ser Val Ala Val Glu Ser Phe Lys Tyr Gly Ser Asp Gly Leu Ile
 290 295 300

Pro Glu Ile Lys Met Thr Thr Gln Gly Pro Ala Gln Leu Lys Ser Leu
 305 310 315 320
 Asn Pro Tyr Val Lys Gln Glu Ala Glu Thr Ile Ala Trp Ser Glu Gly
 325 330 335

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

Cys Thr Val Thr Thr Thr Gly Asn Trp Gln Thr Tyr Lys Thr Val Asp
 405 410 415
 Cys Pro Val Ser Gly Ala Thr Gly Thr Ser Asp Leu Phe Phe Val Phe
 420 425 430
 Thr Gly Ser Gly Ser Gly Ser Leu Phe Asn Phe Asn Trp Trp Gln Phe
 435 440 445
 Ser

<210> 11

<211> 1725

<212> DNA

<213> Fusarium verticillioides

<400> 11

atgcgcttct cttagctatt gtgccccctt ctagegatgg gaagtgtctt tcttgaaacg 60

aagacggatg tttcgacata caccaaccct gtccttcag gatggcactc ggatccatcg 120

tgtatccaga aagatggcct ctttctctgc gtcacttcaa cattcatctc cttcccaggt 180

cttcccgctt atgcctcaag ggatctagtc aactggcgtc tcatcagcca tgtctggaac 240

cgcgagaaac agttgcctgg cattagctgg aagacggcag gacagcaaca gggaatgtat 300

gcaccaacca ttcgatacca caagggaaca tactacgtca tctgcgaata cctgggcgtt 360

ggagatatta ttggtgtcat cttcaagacc accaatccgt gggacgagag tagctggagt 420

gaccctgtta ccttcaagcc aaatcacatc gaccccgatc tgttctggga tgatgacgga 480

aaggtttatt gtgctaccca tggcatcact ctgcaggaga ttgatttgga aactggagag 540

cttagcccg agcttaatat ctggaacggc acaggagggtg tatggcctga gggtecccat 600

atctacaagc gcgacgggta ctactatctc atgattgccg aggggtggaac tgccgaagac 660

cacgctatca caatcgctcg ggcccgaag atcaccggcc cctatgaagc ctacaataac 720

aaccaatct tgaccaaccg cgggacatct gagtacttcc agactgtcgg tcacggtgat 780

ctgttccaag ataccaaggg caactgggtgg ggtctttgtc ttgctactcg catcacagca 840

caggaggttt caccatggg ccgtgaagct gttttgttca atggcacatg gaacaagggc 900

gaatggccca agttgcaacc agtacgaggt cgcatgcctg gaaacctcct cccaaagccg 960

acgcgaaacg ttcccggaga tgggcccttc aacgtgacc cagacaacta caacttgaag 1020

aagactaaga agatccctcc tcactttgtg caccatagag tccaagaga cggtagccttc 1080

tctttgtctt ccaagggctc gcacatcgtg cctagtcgaa acaacgttac cggtagtgtg 1140

ttgccaggag atgagattga gctatcagga cagcgaggtc tagctttcat cggacgccgc 1200

caaaactaca ctctgttcaa atatagtgtt gatatcgact tcaagcccaa gtccgatgat 1260

caggaagctg gaatcacctg tttccgcacg cagttcgacc atatcgatct tggcattgtt 1320

cgtcttecta caaaccaagg cagcaacaag aaatctaagc ttgccttcg attccgggcc 1380

acaggagctc agaatgttcc tgcaccgaag gtagtaccgg tcccgatgg ctgggagaag 1440

ggcgtaatca gtctacatat cgaggcagcc aacgcgacgc actacaacct tggagcttcg 1500

agccacagag gcaagactct cgacatcgcg acagcatcag caagtcttgt gagtggaggc 1560

acgggttcat ttgttggtag ttgcttgga ccttatgcta cctgcaacgg caaaggatct 1620

ggagtggaat gtccaaggg aggtgatgtc tatgtgacct aatggactta taagcccgtg 1680

gcacaagaga ttgatcatgg tgtttttgtg aaatcagaat tgtag 1725

<210> 12

<211> 574

<212> PRT

<213> *Fusarium verticillioides*

<400> 12

Met Arg Phe Ser Trp Leu Leu Cys Pro Leu Leu Ala Met Gly Ser Ala

1 5 10 15

Leu Pro Glu Thr Lys Thr Asp Val Ser Thr Tyr Thr Asn Pro Val Leu

20

25

30

Pro Gly Trp His Ser Asp Pro Ser Cys Ile Gln Lys Asp Gly Leu Phe
 35 40 45
 Leu Cys Val Thr Ser Thr Phe Ile Ser Phe Pro Gly Leu Pro Val Tyr
 50 55 60
 Ala Ser Arg Asp Leu Val Asn Trp Arg Leu Ile Ser His Val Trp Asn
 65 70 75 80
 Arg Glu Lys Gln Leu Pro Gly Ile Ser Trp Lys Thr Ala Gly Gln Gln
 85 90 95
 Gln Gly Met Tyr Ala Pro Thr Ile Arg Tyr His Lys Gly Thr Tyr Tyr
 100 105 110
 Val Ile Cys Glu Tyr Leu Gly Val Gly Asp Ile Ile Gly Val Ile Phe
 115 120 125
 Lys Thr Thr Asn Pro Trp Asp Glu Ser Ser Trp Ser Asp Pro Val Thr
 130 135 140
 Phe Lys Pro Asn His Ile Asp Pro Asp Leu Phe Trp Asp Asp Asp Gly
 145 150 155 160
 Lys Val Tyr Cys Ala Thr His Gly Ile Thr Leu Gln Glu Ile Asp Leu
 165 170 175
 Glu Thr Gly Glu Leu Ser Pro Glu Leu Asn Ile Trp Asn Gly Thr Gly
 180 185 190
 Gly Val Trp Pro Glu Gly Pro His Ile Tyr Lys Arg Asp Gly Tyr Tyr
 195 200 205
 Tyr Leu Met Ile Ala Glu Gly Gly Thr Ala Glu Asp His Ala Ile Thr
 210 215 220
 Ile Ala Arg Ala Arg Lys Ile Thr Gly Pro Tyr Glu Ala Tyr Asn Asn
 225 230 235 240
 Asn Pro Ile Leu Thr Asn Arg Gly Thr Ser Glu Tyr Phe Gln Thr Val
 245 250 255
 Gly His Gly Asp Leu Phe Gln Asp Thr Lys Gly Asn Trp Trp Gly Leu
 260 265 270
 Cys Leu Ala Thr Arg Ile Thr Ala Gln Gly Val Ser Pro Met Gly Arg

275 280 285
 Glu Ala Val Leu Phe Asn Gly Thr Trp Asn Lys Gly Glu Trp Pro Lys
 290 295 300
 Leu Gln Pro Val Arg Gly Arg Met Pro Gly Asn Leu Leu Pro Lys Pro

 305 310 315 320
 Thr Arg Asn Val Pro Gly Asp Gly Pro Phe Asn Ala Asp Pro Asp Asn
 325 330 335
 Tyr Asn Leu Lys Lys Thr Lys Lys Ile Pro Pro His Phe Val His His
 340 345 350
 Arg Val Pro Arg Asp Gly Ala Phe Ser Leu Ser Ser Lys Gly Leu His
 355 360 365
 Ile Val Pro Ser Arg Asn Asn Val Thr Gly Ser Val Leu Pro Gly Asp

 370 375 380
 Glu Ile Glu Leu Ser Gly Gln Arg Gly Leu Ala Phe Ile Gly Arg Arg
 385 390 395 400
 Gln Thr His Thr Leu Phe Lys Tyr Ser Val Asp Ile Asp Phe Lys Pro
 405 410 415
 Lys Ser Asp Asp Gln Glu Ala Gly Ile Thr Val Phe Arg Thr Gln Phe
 420 425 430
 Asp His Ile Asp Leu Gly Ile Val Arg Leu Pro Thr Asn Gln Gly Ser

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

 500 505 510
 Ser Ala Ser Leu Val Ser Gly Gly Thr Gly Ser Phe Val Gly Ser Leu
 515 520 525

Leu Gly Pro Tyr Ala Thr Cys Asn Gly Lys Gly Ser Gly Val Glu Cys
 530 535 540
 Pro Lys Gly Gly Asp Val Tyr Val Thr Gln Trp Thr Tyr Lys Pro Val
 545 550 555 560
 Ala Gln Glu Ile Asp His Gly Val Phe Val Lys Ser Glu Leu

565 570

<210> 13

<211> 2251

<212> DNA

<213> Podospora anserina

<400> 13

atgatccacc tcaagccagc cctcggcgcg ttgttgccgc tgtcgacgca atgtgtggct	60
attgatttgt ttgtcaagtc ttcggggggg aataagacga ctgatatcat gtatggctctt	120
atgcacgagg tatgtgtttt gcgagatctc cttttgttt ttgcgcactg ctgacatgga	180
gactgcaaac aggatatcaa caactccggc gacggcggca tctacgccga gctaactctc	240
aaccgcgcgt tccaaggag tgagaagttc cctccaacc tcgacaactg gagccccgtc	300
ggtggcgcta cccttaccct tcagaagctt gccaaagccc tttcctctgc gttgccttac	360
tccgtcaatg ttgccaaccc caaggagggc aagggaagg gcaaggacac caagggaag	420
aaggttggct tggccaatgc tgggttttgg ggtatggatg tcaagaggca gaagtacact	480
ggtagcttcc acgttactgg tgagtacaag ggtgactttg aggttagctt gcgcagcgcg	540
attaccgggg agacctttgg caagaagggt gtgaagggtg ggagtaagaa ggggaagtgg	600
accgagaagg agtttgagtt ggtgcctttc aaggatgcgc ccaacagcaa caacaccttt	660
gttgtgcagt gggatgccga ggtatgtgct tctttgatat tggctgagat agaagttggg	720
ttgacatgat gtggtgcagg gcgcaaagga cggatctttg gatctcaact tgatcagctt	780
gttcctccg acattcaagg gaaggaagaa tgggctgaga attgatcttg cgcagacgat	840
ggttgagctc aagccggtaa gtcctctcta gtcagaaaag tagagccttt gttaacgctt	900
gacagacctt cttgcgttc cccgtggca acatgctcga gggtaacacc ttggacactt	960
ggtggaagtg gtacgagacc attggccctc tgaaggatcg cccgggcatg gctgggtgtct	1020
gggagtacca gcaaaccctt ggcttgggtc tggtcgagta catggagtgg gccgatgaca	1080
tgaacttgga gccagtatg tgatccatt tcttgagtg acttctcttg ctaacgtatc	1140
cacagttgtc ggtgtcttcg ctggtcttgc cctcgatggc tcgttcgttc ccgaatccga	1200

gatgggatgg gtcatccaac aggctctcga cgaaatcgag ttcctcactg gcgatgctaa 1260
gaccaccaaa tggggtgccg tccgcgcgaa gcttgggtcac cccaagcctt ggaagggtcaa 1320
gtgggttag atcggtaacg aggattggct tgccggacgc cctgctggct tcgagtcgta 1380
catcaactac cgcttcccca tgatgatgaa ggccttcaac gaaaagtacc ccgacatcaa 1440
gatcatgcc tcgcccctca tcttcgacaa catgacaatc cccgcgggtg ctgccggtga 1500
tcaccacccg tacctgactc ccgatgagtt cgttgagcga ttgcccaagt tcgataactt 1560

gagcaaggat aacgtgacgc tcatcggcga ggctgcgtcg acgcatccta acggtggtat 1620
cgcttgggag ggagatctca tgcccttgcc ttggtggggc ggagtggtg ctgaggtat 1680
cttcttgatc agcactgaga gaaacgggtga caagatcatc ggtgctactt acgcgcctgg 1740
tcttcgcagc ttggaccgt ggcaatggag catgacctgg gtgcagcatg ccgccgaccc 1800
ggccctcacc actcgtcga ccagttggtg tgcctggaga atcctgccc accacatcat 1860
ccgtgagacg ctcccggtcg atgccccggc cggcaagccc aactttgacc ctctgttcta 1920
cgttgccgga aagagcgaga gtggcaccgg tatcttcaag gctgccgtct acaactcgac 1980

tgaatcgatc ccggtgtcgt tgaagtttga tggcttcaac gagggagcgg ttgccaactt 2040
gacggtgctt actgggccgg aggatccgta tggatacaac gacccttca ctggtatcaa 2100
tgttgtcaag gagaagacca ctttcatcaa ggccggaaag ggccgcaagt tcaccttcac 2160
cctgccgggc ttgagtgtg ctgtgttga gacggccgac gcggtcaagg gtggcaaggg 2220
aaagggcaag ggcaagggaa agggtaactg a 2251

<210> 14

<211> 676

<212> PRT

<213> Podospora anserina

<400> 14

Met Ile His Leu Lys Pro Ala Leu Ala Ala Leu Leu Ala Leu Ser Thr

1	5	10	15
Gln	Cys	Val	Ala
Ile	Asp	Leu	Phe
Val	Lys	Ser	Ser
Gly	Gly	Asn	Lys
20	25	30	
Thr	Thr	Asp	Ile
Met	Tyr	Gly	Leu
Met	His	Glu	Asp
Ile	Asn	Asn	Ser
35	40	45	
Gly	Asp	Gly	Gly
Ile	Tyr	Ala	Glu
Leu	Ile	Ser	Asn
Arg	Ala	Phe	Gln
50	55	60	

Gly Ser Glu Lys Phe Pro Ser Asn Leu Asp Asn Trp Ser Pro Val Gly

65 70 75 80

Gly Ala Thr Leu Thr Leu Gln Lys Leu Ala Lys Pro Leu Ser Ser Ala

85 90 95

Leu Pro Tyr Ser Val Asn Val Ala Asn Pro Lys Glu Gly Lys Gly Lys

100 105 110

Gly Lys Asp Thr Lys Gly Lys Lys Val Gly Leu Ala Asn Ala Gly Phe

115 120 125

Trp Gly Met Asp Val Lys Arg Gln Lys Tyr Thr Gly Ser Phe His Val

130 135 140

Thr Gly Glu Tyr Lys Gly Asp Phe Glu Val Ser Leu Arg Ser Ala Ile

145 150 155 160

Thr Gly Glu Thr Phe Gly Lys Lys Val Val Lys Gly Gly Ser Lys Lys

165 170 175

Gly Lys Trp Thr Glu Lys Glu Phe Glu Leu Val Pro Phe Lys Asp Ala

180 185 190

Pro Asn Ser Asn Asn Thr Phe Val Val Gln Trp Asp Ala Glu Gly Ala

195 200 205

Lys Asp Gly Ser Leu Asp Leu Asn Leu Ile Ser Leu Phe Pro Pro Thr

210 215 220

Phe Lys Gly Arg Lys Asn Gly Leu Arg Ile Asp Leu Ala Gln Thr Met

225 230 235 240

Val Glu Leu Lys Pro Thr Phe Leu Arg Phe Pro Gly Gly Asn Met Leu

245 250 255

Glu Gly Asn Thr Leu Asp Thr Trp Trp Lys Trp Tyr Glu Thr Ile Gly

260 265 270

Pro Leu Lys Asp Arg Pro Gly Met Ala Gly Val Trp Glu Tyr Gln Gln

275 280 285

Thr Leu Gly Leu Gly Leu Val Glu Tyr Met Glu Trp Ala Asp Asp Met

290 295 300

Asn Leu Glu Pro Ile Val Gly Val Phe Ala Gly Leu Ala Leu Asp Gly

305 310 315 320
 Ser Phe Val Pro Glu Ser Glu Met Gly Trp Val Ile Gln Gln Ala Leu

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

 385 390 395 400
 Glu Lys Tyr Pro Asp Ile Lys Ile Ile Ala Ser Pro Ser Ile Phe Asp
 405 410 415
 Asn Met Thr Ile Pro Ala Gly Ala Ala Gly Asp His His Pro Tyr Leu
 420 425 430
 Thr Pro Asp Glu Phe Val Glu Arg Phe Ala Lys Phe Asp Asn Leu Ser
 435 440 445
 Lys Asp Asn Val Thr Leu Ile Gly Glu Ala Ala Ser Thr His Pro Asn

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

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

Asn Phe Asp Pro Leu Phe Tyr Val Ala Gly Lys Ser Glu Ser Gly Thr
565 570 575
Gly Ile Phe Lys Ala Ala Val Tyr Asn Ser Thr Glu Ser Ile Pro Val
580 585 590
Ser Leu Lys Phe Asp Gly Leu Asn Glu Gly Ala Val Ala Asn Leu Thr
595 600 605
Val Leu Thr Gly Pro Glu Asp Pro Tyr Gly Tyr Asn Asp Pro Phe Thr
610 615 620
Gly Ile Asn Val Val Lys Glu Lys Thr Thr Phe Ile Lys Ala Gly Lys
625 630 635 640
Gly Gly Lys Phe Thr Phe Thr Leu Pro Gly Leu Ser Val Ala Val Leu
645 650 655
Glu Thr Ala Asp Ala Val Lys Gly Gly Lys Gly Lys Gly Lys Gly Lys
660 665 670
Gly Lys Gly Asn
675
<210> 15
<211> 1023
<212> DNA
<213> *Gibberella zeae*
<400> 15
atgaagtcca agttgttatt cccactctc tctttcgttg gtcaaagtct tgccaccaac 60
gacgactgtc ctctcatcac tagtagatgg actgcggatc cttcggtca tgtctttaac 120
gacacctgtt ggctctaccc gtctcatgac atcgatgctg gatttgagaa tgatcctgat 180

ggaggccagt acgcatgag agattacat gtctactcta tcgacaagat ctacggttcc 240
ctgccggtcg atcacggtac ggccctgtca gtggaggatg tcccctgggc ctctcgacag 300
atgtgggctc ctgacgtcg ccacaagaac ggcaaatact acctatactt ccctgcaaaa 360
gacaaggatg atatcttcag aatcggcgtt gctgtctcac caacccccgg cggaccattc 420
gtccccgaca agagtggat ccctcacact ttcagcatcg accccgccag tttcgtcgat 480
gatgatgaca gacacctt ggcatgggggt ggtatcatgg gtggccagct tcaacgatgg 540
caggataaga acaagtacaa cgaatctggc actgagccag gaaacggcac cgctgccttg 600

agccctcaga ttgccaagct gagcaaggac atgcacactc tggcagagaa gcctcgcgac 660
atgtcattc ttgaccccaa gactggcaag ccgctccttt ctgaggatga agaccgacgc 720
ttcttcgaag gaccctggat tcacaagcgc aacaagattt actacctcac ctactctact 780
ggcacaaccc actatcttgt ctatgcgact tcaaagaccc cctatggtcc ttacacctac 840
cagggcagaa ttctggagcc agttgatggc tggactactc actctagtat cgtcaagtac 900
cagggtcagt ggtggctatt ttatcacgat gccaagacat ctggcaagga ctatcttcgc 960
caggtaaagg ctaagaagat ttggtacgat agcaaaggaa agatcttgac aaagaagcct 1020

tga 1023

<210> 16

<211> 340

<212> PRT

<213> Gibberella zeae

<400> 16

Met Lys Ser Lys Leu Leu Phe Pro Leu Leu Ser Phe Val Gly Gln Ser

1 5 10 15

Leu Ala Thr Asn Asp Asp Cys Pro Leu Ile Thr Ser Arg Trp Thr Ala

20 25 30

Asp Pro Ser Ala His Val Phe Asn Asp Thr Leu Trp Leu Tyr Pro Ser

35 40 45

His Asp Ile Asp Ala Gly Phe Glu Asn Asp Pro Asp Gly Gly Gln Tyr

50 55 60

Ala Met Arg Asp Tyr His Val Tyr Ser Ile Asp Lys Ile Tyr Gly Ser

65 70 75 80

Leu Pro Val Asp His Gly Thr Ala Leu Ser Val Glu Asp Val Pro Trp

85 90 95

Ala Ser Arg Gln Met Trp Ala Pro Asp Ala Ala His Lys Asn Gly Lys

100 105 110

Tyr Tyr Leu Tyr Phe Pro Ala Lys Asp Lys Asp Asp Ile Phe Arg Ile

115 120 125

Gly Val Ala Val Ser Pro Thr Pro Gly Gly Pro Phe Val Pro Asp Lys

130 135 140

Ser Trp Ile Pro His Thr Phe Ser Ile Asp Pro Ala Ser Phe Val Asp

145 150 155 160
 Asp Asp Asp Arg Ala Tyr Leu Ala Trp Gly Gly Ile Met Gly Gly Gln
 165 170 175

Leu Gln Arg Trp Gln Asp Lys Asn Lys Tyr Asn Glu Ser Gly Thr Glu
 180 185 190
 Pro Gly Asn Gly Thr Ala Ala Leu Ser Pro Gln Ile Ala Lys Leu Ser
 195 200 205
 Lys Asp Met His Thr Leu Ala Glu Lys Pro Arg Asp Met Leu Ile Leu
 210 215 220
 Asp Pro Lys Thr Gly Lys Pro Leu Leu Ser Glu Asp Glu Asp Arg Arg
 225 230 235 240

Phe Phe Glu Gly Pro Trp Ile His Lys Arg Asn Lys Ile Tyr Tyr Leu
 245 250 255
 Thr Tyr Ser Thr Gly Thr Thr His Tyr Leu Val Tyr Ala Thr Ser Lys
 260 265 270
 Thr Pro Tyr Gly Pro Tyr Thr Tyr Gln Gly Arg Ile Leu Glu Pro Val
 275 280 285
 Asp Gly Trp Thr Thr His Ser Ser Ile Val Lys Tyr Gln Gly Gln Trp
 290 295 300

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

<210> 17

<211> 1047

<212> DNA

<213> Fusarium oxysporum

<400> 17

atgcagctca agtttctgtc ttcagcattg ctgttctctc tgaccagcaa atgcgctgcg 60

caagacacta atgacattcc tccctgac accgacctct ggtccgcaga tccctcggt 120

catgttttcg aaggcaagct ctgggtttac ccatctcacg acatcgaagc caatgttgtc 180

aacggcacag gaggcgtca atacccatg agggattacc atacctactc catgaagagc 240

atctatggta aagatccgt tgtcgaccac ggcgtcgctc tctcagtcga tgacgttccc 300

tggcggaagc agcaaatgtg ggctcctgac gcagctcata agaacggcaa atattatctg 360

tacttccccg ccaaggacaa ggatgagatc ttcagaattg gagttgctgt ctccaacaag 420

cccagcggtc ctttcaaggc cgacaagagc tggatccctg gcacgtacag tatcgatcct 480

gctagctacg tcgacactga taacgaggcc tacctcatct ggggcggtat ctgggcgggc 540

cagctccaag cctggcagga taaaagaac tttaacgagt cgtggattgg agacaaggct 600

gctcctaacg gcaccaatgc cctatctcct cagatcgcca agctaagcaa ggacatgcac 660

aagatcaccg aaacaccccg cgatctcgtc attctcgccc ccgagacagg caagcctctt 720

caggctgagg acaacaagcg acgattcttc gagggccctt ggatccacaa gcgcggcaag 780

ctttactacc tcatgtactc caccggtgat acccacttcc ttgtctacgc tacttccaag 840

aacatctacg gtccttatac ctaccggggc aagattcttg atcctgttga tgggtggact 900

actcatggaa gtattgttga gtataaggga cagtgggtggc ttttctttgc tgatgcgcat 960

acgtctggta aggattacct tcgacaggtg aaggcgagga agatctggta tgacaagaac 1020

ggcaagatct tgcttcaccg tccttiag 1047

<210> 18

<211> 348

<212> PRT

<213> *Fusarium oxysporum*

<400> 18

Met Gln Leu Lys Phe Leu Ser Ser Ala Leu Leu Phe Ser Leu Thr Ser

1 5 10 15

Lys Cys Ala Ala Gln Asp Thr Asn Asp Ile Pro Pro Leu Ile Thr Asp

20 25 30

Leu Trp Ser Ala Asp Pro Ser Ala His Val Phe Glu Gly Lys Leu Trp

35 40 45

Val Tyr Pro Ser His Asp Ile Glu Ala Asn Val Val Asn Gly Thr Gly

50 55 60

Gly Ala Gln Tyr Ala Met Arg Asp Tyr His Thr Tyr Ser Met Lys Ser

65 70 75 80
 Ile Tyr Gly Lys Asp Pro Val Val Asp His Gly Val Ala Leu Ser Val
 85 90 95
 Asp Asp Val Pro Trp Ala Lys Gln Gln Met Trp Ala Pro Asp Ala Ala

 100 105 110
 His Lys Asn Gly Lys Tyr Tyr Leu Tyr Phe Pro Ala Lys Asp Lys Asp
 115 120 125
 Glu Ile Phe Arg Ile Gly Val Ala Val Ser Asn Lys Pro Ser Gly Pro
 130 135 140
 Phe Lys Ala Asp Lys Ser Trp Ile Pro Gly Thr Tyr Ser Ile Asp Pro
 145 150 155 160
 Ala Ser Tyr Val Asp Thr Asp Asn Glu Ala Tyr Leu Ile Trp Gly Gly

 165 170 175
 Ile Trp Gly Gly Gln Leu Gln Ala Trp Gln Asp Lys Lys Asn Phe Asn
 180 185 190
 Glu Ser Trp Ile Gly Asp Lys Ala Ala Pro Asn Gly Thr Asn Ala Leu
 195 200 205
 Ser Pro Gln Ile Ala Lys Leu Ser Lys Asp Met His Lys Ile Thr Glu
 210 215 220
 Thr Pro Arg Asp Leu Val Ile Leu Ala Pro Glu Thr Gly Lys Pro Leu

 225 230 235 240
 Gln Ala Glu Asp Asn Lys Arg Arg Phe Phe Glu Gly Pro Trp Ile His
 245 250 255
 Lys Arg Gly Lys Leu Tyr Tyr Leu Met Tyr Ser Thr Gly Asp Thr His
 260 265 270
 Phe Leu Val Tyr Ala Thr Ser Lys Asn Ile Tyr Gly Pro Tyr Thr Tyr
 275 280 285
 Arg Gly Lys Ile Leu Asp Pro Val Asp Gly Trp Thr Thr His Gly Ser

 290 295 300
 Ile Val Glu Tyr Lys Gly Gln Trp Trp Leu Phe Phe Ala Asp Ala His
 305 310 315 320

Thr Ser Gly Lys Asp Tyr Leu Arg Gln Val Lys Ala Arg Lys Ile Trp
325 330 335

Tyr Asp Lys Asn Gly Lys Ile Leu Leu His Arg Pro
340 345

<210> 19

<211> 1677

<212> DNA

<213> *Aspergillus fumigates*

<400> 19

```

atggcagctc caagtttatt ctacccacac ggtatccaat cgtataccaa tctctcttc      60
cctggttggc actccgatcc cagctgtgcc tacgtagcgg agcaagacac cttttctgc      120
gtgacgtcca ctttcattgc cttccccggt cttctctttt atgcaagccg agatctgcag      180
aactggaaac tggcaagcaa tattttcaat cggcccagcc agatccctga tcttcgcgtc      240
acggatggac agcagtcggg tatctatgcg cccactctgc gctatcatga gggccagttc      300
tacttgatcg ttctgtacct gggcccgagc actaagggtc tgcgtttcac ctctctgat      360
ccgtacgacg atgccgcgtg gacgatccg ctgaattcg cgtacatgg catcgaccg      420

gatatcttct gggatcacga cgggacggtc tatgtcacgt ccgccgagga ccagatgatt      480
aagcagtaca cactcgatct gaagacgggg gcgattggcc cggttgacta cctctggaac      540
ggcaccggag gagtctggcc cgaggcccg cacatttaca agagagacgg atactactac      600
ctcatgatcg cagagggagg taccgagctc ggccactcgg agaccatggc gcgatctaga      660
acccggacag gtccctggga gccatacccg cacaatccgc tcttgtcgaa caagggcacc      720
tcggagtact tccagactgt gggccatgcg gacttgttcc aggatgggaa cggcaactgg      780
tgggccgtgg cgttgagcac ccgatcaggg cctgcatgga agaactatcc catgggtcgg      840

gagacggtgc tcgccccgc cgcttgggag aagggtgagt ggctgtcat tcagcctgtg      900
agaggccaaa tgcaggggcc gtttcacca ccaataaagc gagttcctcg cggcgagggc      960
ggatggatca agcaaccga caaagtggat ttcaggcccg gatcgaagat accggcgcac      1020
ttccagtact ggcgatatcc caagacagag gattttaccg tctccctcg gggccaccg      1080
aatactcttc ggetcacacc ctctttttac aacctaccg gaactgcgga cttaagccg      1140
gatgatggcc tgtcgttgt tatgcgcaa cagaccgaca cttgtttac gtacactgtg      1200
gacgtgtctt ttgacccaa ggttgccgat gaagaggcgg gtgtgactgt ttctctacc      1260

cagcagcagc acatcgatct tggattgtc cttctccaga caaccgagg gctgtcgttg      1320

```

tccttcggt tccggtgga aggccggt aactacgaag gtcctcttcc agaagccacc 1380
 gtgcctgttc ccaaggaatg gtgtggacag accatccggc ttgagattca ggccgtgagt 1440
 gacaccgagt atgtctttgc ggctgccccg gctcggcacc ctgcacagag gcaaatac 1500
 agccgcgcca actcgttgat tgtcagtggt gatacgggac ggtttactgg ctgccttggt 1560
 ggcggtgatg ccacgtcgaa cgggggtgcc ggatccacgc ccgcataat cagcagatgg 1620
 agatacgaag gacggggcca gatgattgat ttggtcgag tggccccgag ctactga 1677

<210> 20

<211> 558

<212> PRT

<213> Aspergillus fumigates

<400> 20

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

1 5 10 15

Asn Pro Leu Phe Pro Gly Trp His Ser Asp Pro Ser Cys Ala Tyr Val

20 25 30

Ala Glu Gln Asp Thr Phe Phe Cys Val Thr Ser Thr Phe Ile Ala Phe

35 40 45

Pro Gly Leu Pro Leu Tyr Ala Ser Arg Asp Leu Gln Asn Trp Lys Leu

50 55 60

Ala Ser Asn Ile Phe Asn Arg Pro Ser Gln Ile Pro Asp Leu Arg Val

65 70 75 80

Thr Asp Gly Gln Gln Ser Gly Ile Tyr Ala Pro Thr Leu Arg Tyr His

85 90 95

Glu Gly Gln Phe Tyr Leu Ile Val Ser Tyr Leu Gly Pro Gln Thr Lys

100 105 110

Gly Leu Leu Phe Thr Ser Ser Asp Pro Tyr Asp Asp Ala Ala Trp Ser

115 120 125

Asp Pro Leu Glu Phe Ala Val His Gly Ile Asp Pro Asp Ile Phe Trp

130 135 140

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

145 150 155 160

Lys Gln Tyr Thr Leu Asp Leu Lys Thr Gly Ala Ile Gly Pro Val Asp

165 170 175
Tyr Leu Trp Asn Gly Thr Gly Gly Val Trp Pro Glu Gly Pro His Ile

180 185 190
Tyr Lys Arg Asp Gly Tyr Tyr Tyr Leu Met Ile Ala Glu Gly Gly Thr

195 200 205
Glu Leu Gly His Ser Glu Thr Met Ala Arg Ser Arg Thr Arg Thr Gly

210 215 220
Pro Trp Glu Pro Tyr Pro His Asn Pro Leu Leu Ser Asn Lys Gly Thr

225 230 235 240
Ser Glu Tyr Phe Gln Thr Val Gly His Ala Asp Leu Phe Gln Asp Gly

245 250 255
Asn Gly Asn Trp Trp Ala Val Ala Leu Ser Thr Arg Ser Gly Pro Ala

260 265 270
Trp Lys Asn Tyr Pro Met Gly Arg Glu Thr Val Leu Ala Pro Ala Ala

275 280 285
Trp Glu Lys Gly Glu Trp Pro Val Ile Gln Pro Val Arg Gly Gln Met

290 295 300
Gln Gly Pro Phe Pro Pro Pro Asn Lys Arg Val Pro Arg Gly Glu Gly

305 310 315 320
Gly Trp Ile Lys Gln Pro Asp Lys Val Asp Phe Arg Pro Gly Ser Lys

325 330 335
Ile Pro Ala His Phe Gln Tyr Trp Arg Tyr Pro Lys Thr Glu Asp Phe

340 345 350
Thr Val Ser Pro Arg Gly His Pro Asn Thr Leu Arg Leu Thr Pro Ser

355 360 365
Phe Tyr Asn Leu Thr Gly Thr Ala Asp Phe Lys Pro Asp Asp Gly Leu

370 375 380
Ser Leu Val Met Arg Lys Gln Thr Asp Thr Leu Phe Thr Tyr Thr Val

385 390 395 400
Asp Val Ser Phe Asp Pro Lys Val Ala Asp Glu Glu Ala Gly Val Thr

405 410 415

Val Phe Leu Thr Gln Gln Gln His Ile Asp Leu Gly Ile Val Leu Leu

420 425 430

Gln Thr Thr Glu Gly Leu Ser Leu Ser Phe Arg Phe Arg Val Glu Gly

435 440 445

Arg Gly Asn Tyr Glu Gly Pro Leu Pro Glu Ala Thr Val Pro Val Pro

450 455 460

Lys Glu Trp Cys Gly Gln Thr Ile Arg Leu Glu Ile Gln Ala Val Ser

465 470 475 480

Asp Thr Glu Tyr Val Phe Ala Ala Ala Pro Ala Arg His Pro Ala Gln

485 490 495

Arg Gln Ile Ile Ser Arg Ala Asn Ser Leu Ile Val Ser Gly Asp Thr

500 505 510

Gly Arg Phe Thr Gly Ser Leu Val Gly Val Tyr Ala Thr Ser Asn Gly

515 520 525

Gly Ala Gly Ser Thr Pro Ala Tyr Ile Ser Arg Trp Arg Tyr Glu Gly

530 535 540

Arg Gly Gln Met Ile Asp Phe Gly Arg Val Val Pro Ser Tyr

545 550 555

<210> 21

<211> 2320

<212> DNA

<213> Penicillium funiculosum

<400> 21

atgggaaaga tgtggcattc gatcttggtt gtgttgggct tattgtctgt cgggcatgcc 60

atcactatca acgtgtccca aagtggcggc aataagacca gtcctttgca atatggtctg 120

atgttcgagg taatccttct cttataccac atataaaagt tgcgtcattt ctaagacaag 180

tcaaggacat aaatcacggc ggtgatggcg gtctgtatgc agagcttggt cgaaaccgag 240

cattccaagg tagcacgctc tatccagcaa acctcgatgg atacgactcg gtcaatggag 300

caatcctagc gcttcagaat ttgacaaacc ctctatcacc ctccatgcct agctctctca 360

acgtcgccaa ggggtccaac aatggaagca tcggtttcgc aaatgaaggc tgggtggggga 420

tagaagtcaa gccgcaaaga tacgcgggct cattctacgt ccagggggac tatcaaggag 480

atttcgacat ctctcttcag tcgaaattga cacaagaagt cttcgcaacg gcaaaagtca	540
ggtcctcggg caaacacgag gactgggttc aatacaagta cgagttggtg cccaaaaagg	600
cagcatcaaa caccaataac actctgacca ttacttttga ctcaaaggtg tgttaaattt	660
tgggttttagt tcgatgtctg gcaattgtct tacgagaaac gtagggattg aaagacggat	720
ccttgaactt caacttgatc agcctatttc ccccaactta caacaatcgg cccaatggcc	780
taagaatcga cctggttgaa gctatggctg aactagaggg ggtaagctct tacaaatcaa	840
ctttatcttt acgaagacta atgtgaaaac ttagaaattt ctgcggtttc caggcggtag	900
cgatgtggaa ggtgtacaag ctccctactg gtataagtgg aatgaaacgg taggagatct	960
caaggaccgt tatagtaggc ccagtgcatg gacgtacgaa gaaagcaatg gaattggctt	1020
gattgagtac atgaattggt gtgatgacat ggggcttgag ccgagtgagt gtattccatt	1080
cagcgtcaaa tccagtgttc taatcatata catcagttct tgccgtatgg gatggacatt	1140
acctttcgaa cgaagtgata tcggaaaacg atttgcagcc atatatcgac gacaccctca	1200
accaactgga attcctgatg ggtgccccag atacgccata tggtagttgg cgtgcgtctc	1260
tgggctatcc gaagccgtgg acgattaact acgtcgagat tggaaacgaa gacaatctat	1320
acgggggact agaacatac atcgctacc ggtttcaggc atattacgac gctataacag	1380
ctaaatatcc ccatatgacg gtcatggaat ctttgacgga gatgcctggt ccggcgggccg	1440
ctgcaagcga ttaccatcaa tattctactc ctgatgggtt tgtttcccag ttcaactact	1500
ttgatcagat gccagtcact aatagaacac tgaacggtat gaaaaccccc ctttttttaa	1560
atatgctttt aatggtatta accatctttc ataggagaga ttgcaaccgt ttatccaaat	1620
aatcctagta attcgggtggc ctggggaagc ccattcccct tgtatccttg gtggattggg	1680
tccgttgcag aagctgtttt cctaattggt gaagagagga attcgccaaa gataatcgg	1740
gctagctacg tacggaattc tacttttcga gattttaaca ttggataaga aggactaacc	1800
tcaatacagg ctccaatgtt cagaaatac aacaattggc agtgggtctcc aacactcatc	1860
gcttttgacg ctgactcgtc gcgtacaagt cgttcaacaa gctggcatgt gatcaaggta	1920
tgctaatttt cctcctcatt caaacccgca gatgtgagct aactttccga agctttcttc	1980
gacaaacaaa atcacgcaaa atttaccac gacttggagt ggcggtgaca taggtccatt	2040
atactgggta gctggacgaa acgacaatac aggatcgaac atattcaagg ccgctgttta	2100
caacagcacc tcagacgtcc ctgtcacctg tcaatttgca ggatgcaacg caaagagcgc	2160
aaatttgacc atcttgtcat ccgacgatcc gaacgcatcg aactaccctg gggggcccga	2220
agtttgaag actgagatcc agtctgtcac tgcaaatgct catggagcat ttgagttcag	2280
tctcccgaac ctaagtgtgg ctgttctcaa aacggagtaa	2320

<210> 22

<211> 642

<212> PRT

<213> *Penicillium funiculosum*

<400> 22

Met Gly Lys Met Trp His Ser Ile Leu Val Val Leu Gly Leu Leu Ser
1 5 10 15

Val Gly His Ala Ile Thr Ile Asn Val Ser Gln Ser Gly Gly Asn Lys
20 25 30

Thr Ser Pro Leu Gln Tyr Gly Leu Met Phe Glu Asp Ile Asn His Gly
35 40 45

Gly Asp Gly Gly Leu Tyr Ala Glu Leu Val Arg Asn Arg Ala Phe Gln
50 55 60

Gly Ser Thr Val Tyr Pro Ala Asn Leu Asp Gly Tyr Asp Ser Val Asn
65 70 75 80

Gly Ala Ile Leu Ala Leu Gln Asn Leu Thr Asn Pro Leu Ser Pro Ser
85 90 95

Met Pro Ser Ser Leu Asn Val Ala Lys Gly Ser Asn Asn Gly Ser Ile
100 105 110

Gly Phe Ala Asn Glu Gly Trp Trp Gly Ile Glu Val Lys Pro Gln Arg
115 120 125

Tyr Ala Gly Ser Phe Tyr Val Gln Gly Asp Tyr Gln Gly Asp Phe Asp
130 135 140

Ile Ser Leu Gln Ser Lys Leu Thr Gln Glu Val Phe Ala Thr Ala Lys
145 150 155 160

Val Arg Ser Ser Gly Lys His Glu Asp Trp Val Gln Tyr Lys Tyr Glu
165 170 175

Leu Val Pro Lys Lys Ala Ala Ser Asn Thr Asn Asn Thr Leu Thr Ile
180 185 190

Thr Phe Asp Ser Lys Gly Leu Lys Asp Gly Ser Leu Asn Phe Asn Leu
195 200 205

Ile Ser Leu Phe Pro Pro Thr Tyr Asn Asn Arg Pro Asn Gly Leu Arg
210 215 220

Ile Asp Leu Val Glu Ala Met Ala Glu Leu Glu Gly Lys Phe Leu Arg
225 230 235 240

Phe Pro Gly Gly Ser Asp Val Glu Gly Val Gln Ala Pro Tyr Trp Tyr
245 250 255

Lys Trp Asn Glu Thr Val Gly Asp Leu Lys Asp Arg Tyr Ser Arg Pro
260 265 270

Ser Ala Trp Thr Tyr Glu Glu Ser Asn Gly Ile Gly Leu Ile Glu Tyr
275 280 285

Met Asn Trp Cys Asp Asp Met Gly Leu Glu Pro Ile Leu Ala Val Trp
290 295 300

Asp Gly His Tyr Leu Ser Asn Glu Val Ile Ser Glu Asn Asp Leu Gln
305 310 315 320

Pro Tyr Ile Asp Asp Thr Leu Asn Gln Leu Glu Phe Leu Met Gly Ala
325 330 335

Pro Asp Thr Pro Tyr Gly Ser Trp Arg Ala Ser Leu Gly Tyr Pro Lys
340 345 350

Pro Trp Thr Ile Asn Tyr Val Glu Ile Gly Asn Glu Asp Asn Leu Tyr
355 360 365

Gly Gly Leu Glu Thr Tyr Ile Ala Tyr Arg Phe Gln Ala Tyr Tyr Asp
370 375 380

Ala Ile Thr Ala Lys Tyr Pro His Met Thr Val Met Glu Ser Leu Thr
385 390 395 400

Glu Met Pro Gly Pro Ala Ala Ala Ala Ser Asp Tyr His Gln Tyr Ser
405 410 415

Thr Pro Asp Gly Phe Val Ser Gln Phe Asn Tyr Phe Asp Gln Met Pro
420 425 430

Val Thr Asn Arg Thr Leu Asn Gly Glu Ile Ala Thr Val Tyr Pro Asn
435 440 445

Asn Pro Ser Asn Ser Val Ala Trp Gly Ser Pro Phe Pro Leu Tyr Pro

450 Trp Trp Ile Gly Ser Val Ala Glu Ala Val Phe Leu Ile Gly Glu Glu
465 470 475 480
Arg Asn Ser Pro Lys Ile Ile Gly Ala Ser Tyr Ala Pro Met Phe Arg
485 490 495
Asn Ile Asn Asn Trp Gln Trp Ser Pro Thr Leu Ile Ala Phe Asp Ala
500 505 510
Asp Ser Ser Arg Thr Ser Arg Ser Thr Ser Trp His Val Ile Lys Leu
515 520 525

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

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

<210> 23

<211> 739

<212> DNA

<213> Aspergillus fumigates

<400> 23

atggttttctt tctcctacct gctgctggcg tgcctcgcca ttggagctct ggctgcccc

60

gtcgaacccg agaccacctc gttcaatgag actgctcttc atgagttcgc tgagcgcgcc 120
ggcaccccaa gctccaccgg ctggaacaac ggctactact actccttctg gactgatggc 180
ggcggcgacg tgacctacac caatggcgcc ggtggctcgt actccgtcaa ctggaggaac 240
gtgggcaact ttgtcggtag aaagggtgg aacctggaa gcgctaggta ccgagctttg 300
tcaacgtcgg atgtgcagac ctgtggctga cagaagtaga accatcaact acggaggcag 360
cttcaacccc agcggcaatg gctacctggc tgtctacggc tggaccacca accccttgat 420
tgagtactac gtgtttgagt cgtatggtac atacaacccc ggcagcggcg gtaccttcag 480

gggcactgtc aacaccgacg gtggcactta caacatctac acggccgttc gctacaatgc 540
tccctccatc gaaggcacca agaccttcac ccagtactgg tctgtgcgca cctccaagcg 600
taccggcggc actgtcacca tggccaacca cttcaacgcc tggagcagac tgggcatgaa 660
cctgggaact cacaactacc agattgtcgc cactgagggt taccagagca gcggatctgc 720
ttccatcact gtctactag 739

<210> 24

<211> 228

<212> PRT

<213> Aspergillus fumigates

<400> 24

Met Val Ser Phe Ser Tyr Leu Leu Leu Ala Cys Ser Ala Ile Gly Ala

1 5 10 15
Leu Ala Ala Pro Val Glu Pro Glu Thr Thr Ser Phe Asn Glu Thr Ala
20 25 30
Leu His Glu Phe Ala Glu Arg Ala Gly Thr Pro Ser Ser Thr Gly Trp
35 40 45
Asn Asn Gly Tyr Tyr Tyr Ser Phe Trp Thr Asp Gly Gly Gly Asp Val
50 55 60
Thr Tyr Thr Asn Gly Ala Gly Gly Ser Tyr Ser Val Asn Trp Arg Asn

65 70 75 80
Val Gly Asn Phe Val Gly Gly Lys Gly Trp Asn Pro Gly Ser Ala Arg
85 90 95
Thr Ile Asn Tyr Gly Gly Ser Phe Asn Pro Ser Gly Asn Gly Tyr Leu
100 105 110
Ala Val Tyr Gly Trp Thr Thr Asn Pro Leu Ile Glu Tyr Tyr Val Val

115 120 125
Glu Ser Tyr Gly Thr Tyr Asn Pro Gly Ser Gly Gly Thr Phe Arg Gly

130 135 140
Thr Val Asn Thr Asp Gly Gly Thr Tyr Asn Ile Tyr Thr Ala Val Arg
145 150 155 160
Tyr Asn Ala Pro Ser Ile Glu Gly Thr Lys Thr Phe Thr Gln Tyr Trp
165 170 175
Ser Val Arg Thr Ser Lys Arg Thr Gly Gly Thr Val Thr Met Ala Asn
180 185 190
His Phe Asn Ala Trp Ser Arg Leu Gly Met Asn Leu Gly Thr His Asn

195 200 205
Tyr Gln Ile Val Ala Thr Glu Gly Tyr Gln Ser Ser Gly Ser Ala Ser
210 215 220
Ile Thr Val Tyr

225

<210> 25

<211> 1002

<212> DNA

<213> Aspergillus fumigates

<400> 25

atgatctcca tttcctcgct cagctttgga ctgccgcta tcgccggcgc atatgctctt 60
ccgagtgaca aatccgtcag cttagcggaa cgtcagacga tcacgaccag ccagacaggc 120
acaaacaatg gctactacta ttccttctgg accaacggtg ccggatcagt gcaatataca 180

aatggtgctg gtggcgaata tagtgtgacg tgggcgaacc agaacggtgg tgactttacc 240
tgtgggaagg gctggaatcc agggagtgac cagtaggcaa cgcccagaaa ctatagaaga 300
ggacgcaaag aaagcactaa actctctact agtgacatta ccttctctgg cagcttcaat 360
ccttcggaa atgcttacct gtccgtgtat ggatggacta ccaacccct agtcgaatac 420
tacatcctcg agaactatgg cagttacaat cctggctcgg gcatgacgca caagggcacc 480
gtcaccagcg atggatccac ctacgacatc tatgagcacc aacagggtcaa ccagccttcg 540
atcgtcggca cgccacctt caaccaatac tggatccatc gccaaaacaa gcgatccagc 600

ggcacagtca ccaccgcgaa tcacttcaag gcctgggcta gtctggggat gaacctgggt 660

accataact atcagattgt ttccactgag ggatatgaga gcagcggtag ctcgaccatc 720
actgtctcgt ctggtgggtc ttcttctggt ggaagtgggt gcagctcgtc tactacttcc 780
tcaggcagct cccctactgg tggctccggc agtgaagtc ttcttccata tggttgtggc 840
tttatgtgta ttctgactgt gatagtgtc tgctttgtgg ggccagtgcg gtggaattgg 900
ctggtctggt cctacttctg gctcttcggg cacttgccag gtttcgaact cgtactactc 960
ccagtgttg tagtaccttc ttgcagggtt atatccaagt ga 1002

<210> 26

<211> 286

<212> PRT

<213> Aspergillus fumigates

<400> 26

Met Ile Ser Ile Ser Ser Leu Ser Phe Gly Leu Ala Ala Ile Ala Gly

1 5 10 15

Ala Tyr Ala Leu Pro Ser Asp Lys Ser Val Ser Leu Ala Glu Arg Gln

20 25 30

Thr Ile Thr Thr Ser Gln Thr Gly Thr Asn Asn Gly Tyr Tyr Tyr Ser

35 40 45

Phe Trp Thr Asn Gly Ala Gly Ser Val Gln Tyr Thr Asn Gly Ala Gly

50 55 60

Gly Glu Tyr Ser Val Thr Trp Ala Asn Gln Asn Gly Gly Asp Phe Thr

65 70 75 80

Cys Gly Lys Gly Trp Asn Pro Gly Ser Asp His Asp Ile Thr Phe Ser

85 90 95

Gly Ser Phe Asn Pro Ser Gly Asn Ala Tyr Leu Ser Val Tyr Gly Trp

100 105 110

Thr Thr Asn Pro Leu Val Glu Tyr Tyr Ile Leu Glu Asn Tyr Gly Ser

115 120 125

Tyr Asn Pro Gly Ser Gly Met Thr His Lys Gly Thr Val Thr Ser Asp

130 135 140

Gly Ser Thr Tyr Asp Ile Tyr Glu His Gln Gln Val Asn Gln Pro Ser

145 150 155 160

Ile Val Gly Thr Ala Thr Phe Asn Gln Tyr Trp Ser Ile Arg Gln Asn

165 170 175
Lys Arg Ser Ser Gly Thr Val Thr Thr Ala Asn His Phe Lys Ala Trp

180 185 190
Ala Ser Leu Gly Met Asn Leu Gly Thr His Asn Tyr Gln Ile Val Ser

195 200 205
Thr Glu Gly Tyr Glu Ser Ser Gly Thr Ser Thr Ile Thr Val Ser Ser

210 215 220
Gly Gly Ser Ser Ser Gly Gly Ser Gly Gly Ser Ser Ser Thr Thr Ser

225 230 235 240
Ser Gly Ser Ser Pro Thr Gly Gly Ser Gly Ser Cys Ser Ala Leu Trp

245 250 255
Gly Gln Cys Gly Gly Ile Gly Trp Ser Gly Pro Thr Cys Cys Ser Ser

260 265 270
Gly Thr Cys Gln Val Ser Asn Ser Tyr Tyr Ser Gln Cys Leu

275 280 285

<210> 27

<211> 1053

<212> DNA

<213> Fusarium verticilloides

<400> 27

atgcagctca agtttctgtc ttcagcattg ttgctgtctt tgaccggcaa ttgcgctgcg 60

caagacacta atgatatccc tctctgatac accgacctct ggtctgcgga tccctcggt 120

catgttttcg agggcaaact ctgggtttac ccatctcacg acatcgaagc caatgtcgtc 180

aacggcaccg gaggcgtca gtacccatg agagattatc acacctatc catgaagacc 240

atctatggaa aagatccgt tatcgacat ggcgtcgtc tgtcagtcga tgatgtccca 300

tggccaagc agcaaatgtg ggctcctgac gcagcttaca agaacggcaa atattatctc 360

tacttcccc ccaaggataa agatgagatc ttcagaattg gatttctgt ctccaacaag 420

cccagcggtc ctttaaggc cgacaagagc tggatcccc gtacttacag tatcgatcct 480

gctagctatg tcgacactaa tggcgaggca tacctcatct ggggcggtat ctggggcggc 540

cagcttcagg cctggcagga tcacaagacc tttaatgagt cgtggctcgg cgacaaagct 600

gctccaacg gcaccaacgc cctatctcct cagatcgcca agctaagcaa ggacatgcac 660

aagatcacgc agacaccccg cgatctcgtc atcctggccc ccgagacagg caagccctt 720
 caagcagagg acaataagcg acgatttttc gaggggcctt gggttcacaa gcgcggcaag 780
 ctgtactacc tcatgtactc taccggcgac acgcacttcc tcgtctacgc gacttccaag 840
 aacatctacg gtccttatac ctatcagggc aagattctcg accctgttga tgggtggact 900
 acgcatggaa gtattgttga gtacaaggga cagtgggtgt tgttctttgc ggatgcgcat 960

acttctggaa aggattatct gagacagggt aaggcgagga agatctggta tgacaaggat 1020
 ggcaagattt tgcttactcg tcctaagatt tag 1053

<210> 28

<211> 350

<212> PRT

<213> Fusarium verticilloides

<400> 28

Met Gln Leu Lys Phe Leu Ser Ser Ala Leu Leu Ser Leu Thr Gly

1 5 10 15

Asn Cys Ala Ala Gln Asp Thr Asn Asp Ile Pro Pro Leu Ile Thr Asp

20 25 30

Leu Trp Ser Ala Asp Pro Ser Ala His Val Phe Glu Gly Lys Leu Trp

35 40 45

Val Tyr Pro Ser His Asp Ile Glu Ala Asn Val Val Asn Gly Thr Gly

50 55 60

Gly Ala Gln Tyr Ala Met Arg Asp Tyr His Thr Tyr Ser Met Lys Thr

65 70 75 80

Ile Tyr Gly Lys Asp Pro Val Ile Asp His Gly Val Ala Leu Ser Val

85 90 95

Asp Asp Val Pro Trp Ala Lys Gln Gln Met Trp Ala Pro Asp Ala Ala

100 105 110

Tyr Lys Asn Gly Lys Tyr Tyr Leu Tyr Phe Pro Ala Lys Asp Lys Asp

115 120 125

Glu Ile Phe Arg Ile Gly Val Ala Val Ser Asn Lys Pro Ser Gly Pro

130 135 140

Phe Lys Ala Asp Lys Ser Trp Ile Pro Gly Thr Tyr Ser Ile Asp Pro

145 150 155 160

Ala Ser Tyr Val Asp Thr Asn Gly Glu Ala Tyr Leu Ile Trp Gly Gly

165 170 175
Ile Trp Gly Gly Gln Leu Gln Ala Trp Gln Asp His Lys Thr Phe Asn
180 185 190
Glu Ser Trp Leu Gly Asp Lys Ala Ala Pro Asn Gly Thr Asn Ala Leu
195 200 205
Ser Pro Gln Ile Ala Lys Leu Ser Lys Asp Met His Lys Ile Thr Glu
210 215 220
Thr Pro Arg Asp Leu Val Ile Leu Ala Pro Glu Thr Gly Lys Pro Leu

225 230 235 240
Gln Ala Glu Asp Asn Lys Arg Arg Phe Phe Glu Gly Pro Trp Val His
245 250 255
Lys Arg Gly Lys Leu Tyr Tyr Leu Met Tyr Ser Thr Gly Asp Thr His
260 265 270
Phe Leu Val Tyr Ala Thr Ser Lys Asn Ile Tyr Gly Pro Tyr Thr Tyr
275 280 285
Gln Gly Lys Ile Leu Asp Pro Val Asp Gly Trp Thr Thr His Gly Ser

290 295 300
Ile Val Glu Tyr Lys Gly Gln Trp Trp Leu Phe Phe Ala Asp Ala His
305 310 315 320
Thr Ser Gly Lys Asp Tyr Leu Arg Gln Val Lys Ala Arg Lys Ile Trp
325 330 335
Tyr Asp Lys Asp Gly Lys Ile Leu Leu Thr Arg Pro Lys Ile
340 345 350

<210> 29

<211> 1031

<212> DNA

<213> Penicillium funiculosum

<400> 29

atgagtcgca gcatccttcc gtacgcctct gttttcgccc tcttgggcgg ggctatcgcc 60
gaaccgtttt tggttctcaa tagcgatttt cccgatccca gtctcataga gacatccagc 120

ggatactatg cattcgggtac caccggaaac ggagtcaatg cgcaggttgc ttcttcacca 180
gactttaata cctggacttt gctttccggc acagatgccc tcccgggacc atttccgtca 240
tgggtagctt cgtctccaca aatctgggcg ccagatgttt tggttaaggt atgttcttat 300
ggaataacag ttttaggagt aggtcagcca ggatattgac aaaattataa taggccgatg 360
gtacctatgt catgtacttt tcggcatctg ctgcgagtga ctccgggcaaa cactgcgttg 420

gtgccgaac tgcgacctca ccggaaggac cttacacccc ggatcgatagc gctgttgctt 480
gtccattaga ccaggaggaga gctattgatg ccaatggatt tattgacacc gacggcacta 540
tatacgttgt atacaaaatt gatggaaaca gtctagacgg tgatggaacc acacatccta 600
cccccatcat gcttcaacaa atggaggcag acggaacaac cccaaccggc agcccaatcc 660
aactcattga ccgatccgac ctgcacggac ctttgatcga ggctcctagt ttgctcctct 720
ccaatggaat ctactacctc agtttctctt ccaactacta caacactaat tactacgaca 780
cttcatacgc ctatgcctcg tcgattactg gtccttggac caaacaatct gcgccttatg 840

cacccttggt gggtactgga accgagacta gcaatgacgg cgcattgagc gcccctggtg 900
gtgccgattt ctccgtcgat ggcaccaaga tgtgttcca cgcaaacctc aatggacaag 960
atatctcggg cggacgcgcc ttatttgctg cgtcaattac tgaggccagc gatgtggtta 1020
cattgcagta g 1031

<210> 30

<211> 321

<212> PRT

<213> Penicillium funiculosum

<400> 30

Met Ser Arg Ser Ile Leu Pro Tyr Ala Ser Val Phe Ala Leu Leu Gly
1 5 10 15

Gly Ala Ile Ala Glu Pro Phe Leu Val Leu Asn Ser Asp Phe Pro Asp
20 25 30

Pro Ser Leu Ile Glu Thr Ser Ser Gly Tyr Tyr Ala Phe Gly Thr Thr
35 40 45

Gly Asn Gly Val Asn Ala Gln Val Ala Ser Ser Pro Asp Phe Asn Thr
50 55 60

Trp Thr Leu Leu Ser Gly Thr Asp Ala Leu Pro Gly Pro Phe Pro Ser
65 70 75 80

Trp Val Ala Ser Ser Pro Gln Ile Trp Ala Pro Asp Val Leu Val Lys
 85 90 95
 Ala Asp Gly Thr Tyr Val Met Tyr Phe Ser Ala Ser Ala Ala Ser Asp
 100 105 110
 Ser Gly Lys His Cys Val Gly Ala Ala Thr Ala Thr Ser Pro Glu Gly
 115 120 125
 Pro Tyr Thr Pro Val Asp Ser Ala Val Ala Cys Pro Leu Asp Gln Gly
 130 135 140

 Gly Ala Ile Asp Ala Asn Gly Phe Ile Asp Thr Asp Gly Thr Ile Tyr
 145 150 155 160
 Val Val Tyr Lys Ile Asp Gly Asn Ser Leu Asp Gly Asp Gly Thr Thr
 165 170 175
 His Pro Thr Pro Ile Met Leu Gln Gln Met Glu Ala Asp Gly Thr Thr
 180 185 190
 Pro Thr Gly Ser Pro Ile Gln Leu Ile Asp Arg Ser Asp Leu Asp Gly
 195 200 205

 Pro Leu Ile Glu Ala Pro Ser Leu Leu Leu Ser Asn Gly Ile Tyr Tyr
 210 215 220
 Leu Ser Phe Ser Ser Asn Tyr Tyr Asn Thr Asn Tyr Tyr Asp Thr Ser
 225 230 235 240
 Tyr Ala Tyr Ala Ser Ser Ile Thr Gly Pro Trp Thr Lys Gln Ser Ala
 245 250 255
 Pro Tyr Ala Pro Leu Leu Val Thr Gly Thr Glu Thr Ser Asn Asp Gly
 260 265 270

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

<210> 31

<211> 2186

<212> DNA

<213> *Fusarium verticillioides*

<400> 31

atggttcgct tcagttcaat cctagcggct gcggcttgct tcgtggctgt tgagtcagtc 60

aacatcaagg tcgacagcaa gggcggaaac gctactagcg gtcaccaata tggttcctt 120

cacgaggttg gtattgacac accactggcg atgattggga tgctaacttg gagctaggat 180

atcaacaatt ccggtgatgg tggcatctac gctgagctca tccgcaatcg tgctttccag 240

tacagcaaga aataccctgt ttctctatct ggctggagac ccatcaacga tgctaagctc 300

tcctcaacc gtctcgacac tctctctctc gacgtctctc ccgtttccat gaacgtgaag 360

cctggaaagg gcaaggccaa ggagattggt ttctcaacg agggttactg gggaatggat 420

gtcaagaagc aaaagtacac tggctctttc tgggttaagg gcgcttaca ggccacttt 480

acagcttctt tgcgatctaa ccttaccgac gatgtctttg gcagcgtcaa ggtcaagtc 540

aaggccaaca agaagcagtg ggttagcat gagtttgtgc ttactcctaa caagaatgcc 600

cctaacagca acaacacttt tgctatcacc tacgatccca aggtgagtaa caatcaaac 660

tgggacgtga tgtatactga caattttagt ggcgctgatg gagctcttga cttcaacctc 720

attagcttgt tccctccacc ctacaaggcg cgcaagaacg gtcttcgagt tgatcttgcc 780

gaggctctcg aaggtctcca ccccgtaagg ttaccgtct cacgtgtatc gtgaacagtc 840

gtgacttgt agaaaagagc ctgctgcgct tccccggtgg taacatgctc gagggcaaca 900

ccaacaagac ctggtgggac tggaaggata ccctcggacc tctccgcaac cgtcctggtt 960

tcgaggggtg ctggaactac cagcagacc atggtcttgg aatcttggag tacctccagt 1020

gggctgagga catgaacctt gaaatcagta ggttctataa aattcagtga cggttatgtg 1080

catgctaaca gatttcagtt gtcggtgtct acgtggcct ctcctcgac ggctccgtca 1140

cccccaagga ccaactccag cccctcatcg acgacgcgt cgacgagatc gaattcatcc 1200

gaggtcccg cacttcaaag tggggaaaga agcgcgtga gctcggccac cccaagcctt 1260

tcagactctc ctacgttgaa gtcggaaacg aggactggct cgctggttat cccactggct 1320

ggaactctta caaggagtac cgcttcccca tgttctctga ggctatcaag aaagtcacc 1380

ccgatctcac cgtcatctcc tctggtgctt ctattgacct cggttggaag aaggatgctg 1440

gtttcgatat tctgtctctt ggaatcggtg actaccacct ttaccgcgag cctgatgttc 1500

ttgttgagga gttaacctg ttgtataaca ataagtatgg tcacatcatt ggtgaggttg 1560

cttctaccca cccaacggt ggaactggct ggagtggtaa ccttatgcct taccctggt 1620
ggatctctgg tgttggcgag gccgtcgctc tctgcggtta tgagcgcaac gccgatcgta 1680
ttcccgaac attctacgct cctatcctca agaacgagaa ccgttggcag tgggctatca 1740

ccatgatcca attcgccgcc gactccgcca tgaccaccgc ctccaccagc tggatatgtct 1800
ggtcactctt cgcaggccac cccatgacct atactctccc caccaccgcc gacttcgacc 1860
ccctctacta cgtcgctggt aagaacgagg acaagggaac tcttatctgg aagggtgctg 1920
cgtataacac caccaagggt gctgacgttc ccgtgtctct gtccttcaag ggtgtcaagc 1980
ccggtgctca agctgagctt actcttctga ccaacaagga gaaggatcct ttgctgttca 2040
atgatcctca caagggaac aatgttgttg atactaagaa gactgttctc aaggccgatg 2100
gaaagggtgc tttaacttc aagcttccta acctgagcgt cgctgttctt gagaccctca 2160

agaagggaag gccttactct agctag 2186

<210> 32

<211> 660

<212> PRT

<213> *Fusarium verticillioide*

<400> 32

Met Val Arg Phe Ser Ser Ile Leu Ala Ala Ala Ala Cys Phe Val Ala

1 5 10 15

Val Glu Ser Val Asn Ile Lys Val Asp Ser Lys Gly Gly Asn Ala Thr

20 25 30

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

35 40 45

Asp Gly Gly Ile Tyr Ala Glu Leu Ile Arg Asn Arg Ala Phe Gln Tyr

50 55 60

Ser Lys Lys Tyr Pro Val Ser Leu Ser Gly Trp Arg Pro Ile Asn Asp

65 70 75 80

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

85 90 95

Pro Val Ser Met Asn Val Lys Pro Gly Lys Gly Lys Ala Lys Glu Ile

100 105 110

Gly Phe Leu Asn Glu Gly Tyr Trp Gly Met Asp Val Lys Lys Gln Lys

115	120	125	
Tyr Thr Gly Ser Phe Trp Val Lys Gly Ala Tyr Lys Gly His Phe Thr			
130	135	140	
Ala Ser Leu Arg Ser Asn Leu Thr Asp Asp Val Phe Gly Ser Val Lys			
145	150	155	160
Val Lys Ser Lys Ala Asn Lys Lys Gln Trp Val Glu His Glu Phe Val			
165	170	175	
Leu Thr Pro Asn Lys Asn Ala Pro Asn Ser Asn Asn Thr Phe Ala Ile			
180	185	190	
Thr Tyr Asp Pro Lys Gly Ala Asp Gly Ala Leu Asp Phe Asn Leu Ile			
195	200	205	
Ser Leu Phe Pro Pro Thr Tyr Lys Gly Arg Lys Asn Gly Leu Arg Val			
210	215	220	
Asp Leu Ala Glu Ala Leu Glu Gly Leu His Pro Ser Leu Leu Arg Phe			
225	230	235	240
Pro Gly Gly Asn Met Leu Glu Gly Asn Thr Asn Lys Thr Trp Trp Asp			
245	250	255	
Trp Lys Asp Thr Leu Gly Pro Leu Arg Asn Arg Pro Gly Phe Glu Gly			
260	265	270	
Val Trp Asn Tyr Gln Gln Thr His Gly Leu Gly Ile Leu Glu Tyr Leu			
275	280	285	
Gln Trp Ala Glu Asp Met Asn Leu Glu Ile Ile Val Gly Val Tyr Ala			
290	295	300	
Gly Leu Ser Leu Asp Gly Ser Val Thr Pro Lys Asp Gln Leu Gln Pro			
305	310	315	320
Leu Ile Asp Asp Ala Leu Asp Glu Ile Glu Phe Ile Arg Gly Pro Val			
325	330	335	
Thr Ser Lys Trp Gly Lys Lys Arg Ala Glu Leu Gly His Pro Lys Pro			
340	345	350	
Phe Arg Leu Ser Tyr Val Glu Val Gly Asn Glu Asp Trp Leu Ala Gly			
355	360	365	

Tyr Pro Thr Gly Trp Asn Ser Tyr Lys Glu Tyr Arg Phe Pro Met Phe
 370 375 380
 Leu Glu Ala Ile Lys Lys Ala His Pro Asp Leu Thr Val Ile Ser Ser
 385 390 395 400
 Gly Ala Ser Ile Asp Pro Val Gly Lys Lys Asp Ala Gly Phe Asp Ile
 405 410 415
 Pro Ala Pro Gly Ile Gly Asp Tyr His Pro Tyr Arg Glu Pro Asp Val
 420 425 430

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

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

 Asn Glu Asp Lys Gly Thr Leu Ile Trp Lys Gly Ala Ala Tyr Asn Thr
 565 570 575
 Thr Lys Gly Ala Asp Val Pro Val Ser Leu Ser Phe Lys Gly Val Lys
 580 585 590
 Pro Gly Ala Gln Ala Glu Leu Thr Leu Leu Thr Asn Lys Glu Lys Asp
 595 600 605
 Pro Phe Ala Phe Asn Asp Pro His Lys Gly Asn Asn Val Val Asp Thr

610

615

620

Lys Lys Thr Val Leu Lys Ala Asp Gly Lys Gly Ala Phe Asn Phe Lys

625

630

635

640

Leu Pro Asn Leu Ser Val Ala Val Leu Glu Thr Leu Lys Lys Gly Lys

645

650

655

Pro Tyr Ser Ser

660

<210> 33

<400> 33

000

<210> 34

<400> 34

000

<210> 35

<400> 35

000

<210> 36

<400> 36

000

<210> 37

<400> 37

000

<210> 38

<400> 38

000

<210> 39

<400> 39

000

<210>

> 40

<400> 40

000

<210> 41

<211> 1352

<212> DNA

<213> *Trichoderma reesei*

<400> 41

atgaaagcaa acgtcatctt gtgcctcctg gccccctgg tcgccgtctt cccaccgaa	60
accatccacc tcgaccccgga gctcgccgct ctccgcgcca acctaccga gcgaacagcc	120
gacctctggg accgccaagc ctctcaaagc atcgaccagc tcatacaagag aaaaggcaag	180
ctctactttg gcaccgccac cgaccgcggc ctctccaac gggaaaagaa cgcggccatc	240
atccaggcag acctcgccca ggtgacgccg gagaacagca tgaagtggca gtcgctcgag	300
aacaaccaag gccagctgaa ctggggagac gccgactatc tcgtcaactt tgcccagcaa	360
aacggcaagt cgatacgcg ccacactctg atctggcact cgcagctgcc tgcgtgggtg	420
aacaatatca acaacgcgga tactctgcgg caagtcatcc gcacccatgt ctctactgtg	480
gttgggcggt acaagggcaa gattcgtgct tgggtgagtt ttgaacacca catgcccctt	540
ttcttagtcc gtcctcctc ctcttggaac ttctcacagt tatagccgta tacaacattc	600
gacaggaaat ttaggatgac aactactgac tgacttgtgt gtgtgatggc gataggacgt	660
ggtcaatgaa atcttcaacg aggatggaac gctgcgctct tcagtccttt ccaggctcct	720
cggcgaggag tttgtctcga ttgcctttcg tgcgtctcga gatgctgacc cttctgccc	780
tctttacatc aacgactaca atctcgaccg cgccaactat ggcaaggcca acgggttgaa	840
gacttacgtc tccaagtgga tctctcaagg agttccatt gacggtattg gtgagccacg	900
accctaaat gtccccatt agagtctctt tctagagcca aggcttgaag ccattcagg	960
actgacagca gacgcttctc tacaggaagc cagtccatc tcagcggcgg cggaggctct	1020
ggtacgtggt gtgcgtccca gcagctggca acggtaccg tcaccgagct ggccattacc	1080
gagctggaca ttcagggggc accgacgacg gattacaccc aagttgttca agcatgcctg	1140
agcgtctcca agtgcgtcgg catcaccgtg tggggcatca gtgacaaggt aagttgcttc	1200
ccctgtctgt gcttatcaac tgtaagcagc aacaactgat gctgtctgtc ttacctagg	1260
actcgtggcg tgccagcacc aacctcttc tgtttgacgc aaacttcaac cccaagccgg	1320
catataacag cattgttggc atcttacaat ag	1352

<210> 42

<211> 347

<212> PRT

<213> *Trichoderma reesei*

<400> 42

Met Lys Ala Asn Val Ile Leu Cys Leu Leu Ala Pro Leu Val Ala Ala

1 5 10 15
 Leu Pro Thr Glu Thr Ile His Leu Asp Pro Glu Leu Ala Ala Leu Arg
 20 25 30

 Ala Asn Leu Thr Glu Arg Thr Ala Asp Leu Trp Asp Arg Gln Ala Ser
 35 40 45
 Gln Ser Ile Asp Gln Leu Ile Lys Arg Lys Gly Lys Leu Tyr Phe Gly
 50 55 60
 Thr Ala Thr Asp Arg Gly Leu Leu Gln Arg Glu Lys Asn Ala Ala Ile
 65 70 75 80
 Ile Gln Ala Asp Leu Gly Gln Val Thr Pro Glu Asn Ser Met Lys Trp
 85 90 95

 Gln Ser Leu Glu Asn Asn Gln Gly Gln Leu Asn Trp Gly Asp Ala Asp
 100 105 110
 Tyr Leu Val Asn Phe Ala Gln Gln Asn Gly Lys Ser Ile Arg Gly His
 115 120 125
 Thr Leu Ile Trp His Ser Gln Leu Pro Ala Trp Val Asn Asn Ile Asn
 130 135 140
 Asn Ala Asp Thr Leu Arg Gln Val Ile Arg Thr His Val Ser Thr Val
 145 150 155 160

 Val Gly Arg Tyr Lys Gly Lys Ile Arg Ala Trp Asp Val Val Asn Glu
 165 170 175
 Ile Phe Asn Glu Asp Gly Thr Leu Arg Ser Ser Val Phe Ser Arg Leu
 180 185 190
 Leu Gly Glu Glu Phe Val Ser Ile Ala Phe Arg Ala Ala Arg Asp Ala
 195 200 205
 Asp Pro Ser Ala Arg Leu Tyr Ile Asn Asp Tyr Asn Leu Asp Arg Ala
 210 215 220

 Asn Tyr Gly Lys Val Asn Gly Leu Lys Thr Tyr Val Ser Lys Trp Ile
 225 230 235 240
 Ser Gln Gly Val Pro Ile Asp Gly Ile Gly Ser Gln Ser His Leu Ser
 245 250 255

Gly Gly Gly Gly Ser Gly Thr Leu Gly Ala Leu Gln Gln Leu Ala Thr
260 265 270

Val Pro Val Thr Glu Leu Ala Ile Thr Glu Leu Asp Ile Gln Gly Ala
275 280 285

Pro Thr Thr Asp Tyr Thr Gln Val Val Gln Ala Cys Leu Ser Val Ser
290 295 300

Lys Cys Val Gly Ile Thr Val Trp Gly Ile Ser Asp Lys Asp Ser Trp
305 310 315 320

Arg Ala Ser Thr Asn Pro Leu Leu Phe Asp Ala Asn Phe Asn Pro Lys
325 330 335

Pro Ala Tyr Asn Ser Ile Val Gly Ile Leu Gln
340 345

<210> 43

<211> 222

<212> PRT

<213> Trichoderma reesei

<400> 43

Met Val Ser Phe Thr Ser Leu Leu Ala Ala Ser Pro Pro Ser Arg Ala
1 5 10 15

Ser Cys Arg Pro Ala Ala Glu Val Glu Ser Val Ala Val Glu Lys Arg
20 25 30

Gln Thr Ile Gln Pro Gly Thr Gly Tyr Asn Asn Gly Tyr Phe Tyr Ser
35 40 45

Tyr Trp Asn Asp Gly His Gly Gly Val Thr Tyr Thr Asn Gly Pro Gly
50 55 60

Gly Gln Phe Ser Val Asn Trp Ser Asn Ser Gly Asn Phe Val Gly Gly
65 70 75 80

Lys Gly Trp Gln Pro Gly Thr Lys Asn Lys Val Ile Asn Phe Ser Gly
85 90 95

Ser Tyr Asn Pro Asn Gly Asn Ser Tyr Leu Ser Val Tyr Gly Trp Ser
100 105 110

Arg Asn Pro Leu Ile Glu Tyr Tyr Ile Val Glu Asn Phe Gly Thr Tyr

115 120 125
Asn Pro Ser Thr Gly Ala Thr Lys Leu Gly Glu Val Thr Ser Asp Gly
130 135 140
Ser Val Tyr Asp Ile Tyr Arg Thr Gln Arg Val Asn Gln Pro Ser Ile
145 150 155 160
Ile Gly Thr Ala Thr Phe Tyr Gln Tyr Trp Ser Val Arg Arg Asn His
165 170 175
Arg Ser Ser Gly Ser Val Asn Thr Ala Asn His Phe Asn Ala Trp Ala

180 185 190
Gln Gln Gly Leu Thr Leu Gly Thr Met Asp Tyr Gln Ile Val Ala Val
195 200 205
Glu Gly Tyr Phe Ser Ser Gly Ser Ala Ser Ile Thr Val Ser
210 215 220

<210> 44

<211> 797

<212> PRT

<213> Trichoderma reesei

<400> 44

Met Val Asn Asn Ala Ala Leu Leu Ala Ala Leu Ser Ala Leu Leu Pro
1 5 10 15

Thr Ala Leu Ala Gln Asn Asn Gln Thr Tyr Ala Asn Tyr Ser Ala Gln
20 25 30
Gly Gln Pro Asp Leu Tyr Pro Glu Thr Leu Ala Thr Leu Thr Leu Ser
35 40 45
Phe Pro Asp Cys Glu His Gly Pro Leu Lys Asn Asn Leu Val Cys Asp
50 55 60
Ser Ser Ala Gly Tyr Val Glu Arg Ala Gln Ala Leu Ile Ser Leu Phe
65 70 75 80

Thr Leu Glu Glu Leu Ile Leu Asn Thr Gln Asn Ser Gly Pro Gly Val
85 90 95

Pro Arg Leu Gly Leu Pro Asn Tyr Gln Val Trp Asn Glu Ala Leu His
100 105 110

Gly Leu Asp Arg Ala Asn Phe Ala Thr Lys Gly Gly Gln Phe Glu Trp
115 120 125

Ala Thr Ser Phe Pro Met Pro Ile Leu Thr Thr Ala Ala Leu Asn Arg
130 135 140

Thr Leu Ile His Gln Ile Ala Asp Ile Ile Ser Thr Gln Ala Arg Ala
145 150 155 160

Phe Ser Asn Ser Gly Arg Tyr Gly Leu Asp Val Tyr Ala Pro Asn Val
165 170 175

Asn Gly Phe Arg Ser Pro Leu Trp Gly Arg Gly Gln Glu Thr Pro Gly
180 185 190

Glu Asp Ala Phe Phe Leu Ser Ser Ala Tyr Thr Tyr Glu Tyr Ile Thr
195 200 205

Gly Ile Gln Gly Gly Val Asp Pro Glu His Leu Lys Val Ala Ala Thr
210 215 220

Val Lys His Phe Ala Gly Tyr Asp Leu Glu Asn Trp Asn Asn Gln Ser
225 230 235 240

Arg Leu Gly Phe Asp Ala Ile Ile Thr Gln Gln Asp Leu Ser Glu Tyr
245 250 255

Tyr Thr Pro Gln Phe Leu Ala Ala Ala Arg Tyr Ala Lys Ser Arg Ser
260 265 270

Leu Met Cys Ala Tyr Asn Ser Val Asn Gly Val Pro Ser Cys Ala Asn
275 280 285

Ser Phe Phe Leu Gln Thr Leu Leu Arg Glu Ser Trp Gly Phe Pro Glu
290 295 300

Trp Gly Tyr Val Ser Ser Asp Cys Asp Ala Val Tyr Asn Val Phe Asn
305 310 315 320

Pro His Asp Tyr Ala Ser Asn Gln Ser Ser Ala Ala Ala Ser Ser Leu
325 330 335

Arg Ala Gly Thr Asp Ile Asp Cys Gly Gln Thr Tyr Pro Trp His Leu

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

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

 Lys Ala Gly Tyr His Val Asn Phe Glu Leu Gly Thr Glu Ile Ala Gly
 465 470 475 480
 Asn Ser Thr Thr Gly Phe Ala Lys Ala Ile Ala Ala Ala Lys Lys Ser
 485 490 495
 Asp Ala Ile Ile Tyr Leu Gly Gly Ile Asp Asn Thr Ile Glu Gln Glu
 500 505 510
 Gly Ala Asp Arg Thr Asp Ile Ala Trp Pro Gly Asn Gln Leu Asp Leu
 515 520 525

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

Thr Thr Gln Tyr Pro Ala Glu Tyr Val His Gln Phe Pro Gln Asn Asp
595 600 605

Met Asn Leu Arg Pro Asp Gly Lys Ser Asn Pro Gly Gln Thr Tyr Ile
610 615 620

Trp Tyr Thr Gly Lys Pro Val Tyr Glu Phe Gly Ser Gly Leu Phe Tyr
625 630 635 640

Thr Thr Phe Lys Glu Thr Leu Ala Ser His Pro Lys Ser Leu Lys Phe
645 650 655

Asn Thr Ser Ser Ile Leu Ser Ala Pro His Pro Gly Tyr Thr Tyr Ser
660 665 670

Glu Gln Ile Pro Val Phe Thr Phe Glu Ala Asn Ile Lys Asn Ser Gly
675 680 685

Lys Thr Glu Ser Pro Tyr Thr Ala Met Leu Phe Val Arg Thr Ser Asn
690 695 700

Ala Gly Pro Ala Pro Tyr Pro Asn Lys Trp Leu Val Gly Phe Asp Arg
705 710 715 720

Leu Ala Asp Ile Lys Pro Gly His Ser Ser Lys Leu Ser Ile Pro Ile
725 730 735

Pro Val Ser Ala Leu Ala Arg Val Asp Ser His Gly Asn Arg Ile Val
740 745 750

Tyr Pro Gly Lys Tyr Glu Leu Ala Leu Asn Thr Asp Glu Ser Val Lys
755 760 765

Leu Glu Phe Glu Leu Val Gly Glu Glu Val Thr Ile Glu Asn Trp Pro
770 775 780

Leu Glu Glu Gln Gln Ile Lys Asp Ala Thr Pro Asp Ala
785 790 795

<210> 45

<211> 744

<212> PRT

<213> Trichoderma reesei

<400> 45

Met Arg Tyr Arg Thr Ala Ala Ala Leu Ala Leu Ala Thr Gly Pro Phe

1 5 10 15
 Ala Arg Ala Asp Ser His Ser Thr Ser Gly Ala Ser Ala Glu Ala Val
 20 25 30
 Val Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala Lys

 35 40 45
 Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser
 50 55 60
 Gly Val Gly Trp Asn Gly Gly Pro Cys Val Gly Asn Thr Ser Pro Ala
 65 70 75 80
 Ser Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu Gly
 85 90 95
 Val Arg Tyr Ser Thr Gly Ser Thr Ala Phe Thr Pro Gly Val Gln Ala

 100 105 110
 Ala Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe Ile
 115 120 125
 Gly Glu Glu Val Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro Val
 130 135 140
 Ala Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu Gly
 145 150 155 160
 Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr Ile

 165 170 175
 Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr Ile
 180 185 190
 Leu Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro Asp
 195 200 205
 Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala Val
 210 215 220
 Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn Thr

 225 230 235 240
 Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp
 245 250 255

Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His
260 265 270

Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly
275 280 285

Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn
290 295 300

Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val
305 310 315 320

Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly
325 330 335

Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr
340 345 350

Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp
355 360 365

Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly
370 375 380

Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn
385 390 395 400

Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly
405 410 415

Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr
420 425 430

Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn
435 440 445

Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val
450 455 460

Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn
465 470 475 480

Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu
485 490 495

Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His

500 505 510
 Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val
 515 520 525
 Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala
 530 535 540
 Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val

 545 550 555 560
 Tyr Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser
 565 570 575
 Gly Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His
 580 585 590
 Phe Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu
 595 600 605
 Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala

 610 615 620
 Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp
 625 630 635 640
 Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly
 645 650 655
 Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser
 660 665 670
 Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu

 675 680 685
 Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg
 690 695 700
 Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro
 705 710 715 720
 Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg
 725 730 735
 Leu Thr Ser Thr Leu Ser Val Ala
 740

<210> 46

<211> 2031

<212> DNA

<213> *Podospira anserina*

<400> 46

atgatccacc tcaagccagc cctcgcggcg ttgttggcgc tgtcgacgca atgtgtggct	60
attgatttgt ttgtcaagtc ttcggggggg aataagacga ctgatatcat gtatggtctt	120
atgcacgagg atatcaacaa ctccggcgac ggcggcacatc acgccgagct aatctccaac	180
cgcgcttcc aaggagtgga gaagttcccc tccaacctcg acaactggag ccccgtcggt	240
ggcgctacc ttaccttca gaagcttgcc aagccccctt cctctgcgtt gccttactcc	300
gtcaatgttg ccaaccccaa ggagggaag ggcaaggga aggacaccaa ggggaagaag	360
gttggcttgg ccaatgctgg gttttgggt atggatgtca agaggcagaa gtacactggt	420
agcttccacg ttactgttga gtacaagggt gactttgagg ttagcttgcg cagcgcgatt	480
accggggaga cttttggcaa gaaggtggtg aagggtggga gtaagaagg gaagtggacc	540
gagaaggagt ttgagttggt gcctttcaag gatgcgcca acagcaacaa cacctttgtt	600
gtgcagtggg atgccgagg cgcaaaggac ggatctttgg atctcaactt gatcagcttg	660
ttccctccga cattcaaggg aaggaagaat gggctgagaa ttgatcttgc gcagacgatg	720
gttgagctca agccgacctt cttgcgttc cccggtggca acatgctcga gggtaacacc	780
ttggacactt ggtggaagtg gtacgagacc attggccctc tgaaggatcg cccgggcatg	840
gttgggtgtc gggagtacca gcaaaccctt ggcttgggtc tggtcgagta catggagtgg	900
gccgatgaca tgaacttga gccattgtc ggtgtcttcg ctggtcttgc cctcgatggc	960
tcgttcgttc ccgaatccga gatgggatgg gtcatccaac aggcctctga cgaaatcgag	1020
ttcctcactg gcgatgctaa gaccaccaa tgggtgccc tccgcgcgaa gcttggtcac	1080
cccaagcctt ggaaggtcaa gtgggttgag atcggtaacg aggattggct tgccggacgc	1140
cctgctggct tcgagtcgta catcaactac cgttcccca tgatgatgaa ggccttcaac	1200
gaaaagtacc ccgacatcaa gatcatcgcc tcgccctcca tcttcgacaa catgacaatc	1260
cccgcgggtg ctgccgtga tcaccaccg tacctgactc ccgatgagtt cgttgagcga	1320
ttcgccaagt tcgataactt gagcaaggat aacgtgacgc tcacggcgga ggctgcgtcg	1380
acgcaccta acggtgggat cgcttgggag ggagatctca tgcccttgcc ttggtggggc	1440
ggcagtgtt ctgaggctat cttcttgatc agcactgaga gaaacggtga caagatcatc	1500
ggtgctactt acgcgcctgg tcttcgcagc ttggaccgct ggcaatggag catgacctgg	1560

gtgcagcatg ccgccgaccc ggccctcacc actcgctcga ccagttggta tgtctggaga 1620

atcctcgccc accacatcat ccgtgagacg ctcccggtcg atgccccggc cggcaagccc 1680

aactttgacc ctctgttcta cgttgccgga aagagcgaga gtggcaccgg tatcttcaag 1740

gtcgccgtct acaactcgac tgaatcgatc ccggtgtcgt tgaagtttga tggcttcaac 1800

gagggagcgg ttgccaactt gacggtgctt actgggcccgg aggatccgta tggatacaac 1860

gacccttca ctggtatcaa tgttgtcaag gagaagacca cttcatcaa ggccggaaag 1920

ggcggcaagt tcaccttcac cctgccgggc ttgagtgttg ctgtgttgga gacggccgac 1980

gcggtcaagg gtggcaaggg aaagggaag ggcaagggaagggttaactg a 2031

<210> 47

<211> 2031

<212> DNA

<213> Artificial Sequence

<220><223> synthetic codon optimized GH51 enzyme from Podospora anserina

<400> 47

atgatccacc tcaagccgc cctcgccgcc ctctcgccc tcagcaccca atgcgtcgcc 60

atcgacctct tcgtcaagag cagcggcggc aacaagacca ccgacatcat gtacggcctc 120

atgcacgagg acatcaacaa cagcggcgac ggccggcatct acgccgagct gatcagcaac 180

cgcgccttcc agggcagcga gaagttcccc agcaacctcg acaactggtc ccccgtcggc 240

ggcgccaccc tcacctcca gaagctcgcc aagccctgt cctctgccct cccctactcc 300

gtcaacgtcg ccaaccccaa ggagggttaag ggtaagggaaggacaccaa gggcaagaag 360

gtcggcctcg ccaacgccgg cttttggggc atggacgtca agcggccagaa atacaccggc 420

agcttcacag tcaccggcga gtacaagggc gacttcgagg tcagcctccg cagcgccatt 480

accggcgaga ccttcggcaa gaaggtcgtc aaggcgggca gcaagaaggg caagtggacc 540

gagaaggagt tcgagctggt ccccttcaag gacgccccca acagcaacaa caccttcgtc 600

gtccagtggg acgccgaggg cgccaaggac ggccagcctcg acctcaacct catcagcctc 660

ttcccgccca ccttcaaggg ccgcaagaac ggctccgca tcgacctcgc ccagaccatg 720

gtcgagctga agcccacctt cctcgcttt cccggcggca acatgctcga gggcaacacc 780

ctcgacacct ggttgaagtgt gtacgagacc atcgggcccc tgaaggaccg ccctggcatg 840

gccggcgtct gggagtacca gcagacgtg ggccctcggc tggtcgagta catggagtgg 900

gccgacgaca tgaacctcga gcccatcgtc ggctcttttg ctggcctggc cctggatggc 960

agctttgtcc ccgagagcga gatgggctgg gtcattccagc aggtctctcga tgagatcgag 1020

ttcctcaccg gcgacgcca gaccaccaag tggggcgccg tccgcgcca gctcggccac 1080
cctaagccct ggaaggtcaa atgggtcgag atcggcaacg aggactggct cgccggccga 1140

cctgcgggtc tcgagagcta catcaactac cgtttcccca tgatgatgaa ggccttcaac 1200
gagaaatacc ccgacatcaa gatcattgcc agccctcca tcttcgaaa catgaccatt 1260
ccagccggtg ctgccggtga ccaccacccc tacctacccc ccgacgaatt tctcgagcgc 1320
ttcgccaagt tcgacaacct cagcaaggac aacgtcacc tcattggcga ggccgccagc 1380
accaccccca acggcggcat tgctggggag ggcgacctca tgccctgcc ctggtggggc 1440
ggcagcgtcg ccgaggccat cttctcctc agcaccgagc gcaacggcga caagatcctc 1500
ggcgccacct acgcccctgg cctccgatct ctgcaccgct ggcatggag catgacctgg 1560

gtccagcagc ccgccgaccc tgccctcacc acccgagca ccagctggtc cgtctggcgc 1620
atcctcgccc accacatcat tcgcgagacc ctccccgtcg acgccccgc cggcaagccc 1680
aacttcgacc cctcttcta cgtcgtctggc aagtcggaga gcggcaccgg catcttcaag 1740
gccgccgtct acaacagcac cgagagcctc cccgtcagcc tcaagtcca cggcctcaac 1800
gagggcgccg tcgccaacct caccgtctc accggccccg aggacccta cggctacaac 1860
gacccttca ccggcatcaa cgtcgtcaag gaaaagacca ccttcatcaa ggccggcaag 1920
ggcggcaagt tcacctttac cctccccggc ctctctgtcg ccgtcctga gaccgccgac 1980

gccgtgaagg gtggcaaggg aaagggaag ggcaagggtc agggtaacta a 2031

<210> 48
<211> 1020
<212> DNA
<213> *Gibberella zeae*
<400> 48

atgtatcgga agttggccgt catctcgcc tttttggcca cagctcgtgc taccaacgac 60
gactgtctc tcactactag tagatggact gcggatcctt cggtcatgt cttaacgac 120
accttgtggc tctaccgtc tcatgacatc gatgtggat ttgagaatga tctgatgga 180
ggccagtacg ccatgagaga ttacatgtc tactctatcg acaagatcta cggttccctg 240
ccggtcgatc acggtacggc cctgtcagtg gaggatgtcc cctgggcctc tcgacagatg 300

tgggctcctg acgtgcacca caagaacggc aaatactacc tatacttccc tgccaaagac 360
aaggatgata tcttcagaat cggcgttgct gtctcaccia cccccggcg accattcgtc 420
cccgacaaga gttggatccc tcacactttc agcatcgacc ccgccagttt cgtcgatgat 480
gatgacagag cctacttggc atgggttgg atcatgggtg gccagcttca acgatggcag 540

gataagaaca agtacaacga atctggcact gagccaggaa acggcaccgc tgccttgagc 600
cctcagattg ccaagctgag caaggacatg cacactctgg cagagaagcc tcgcgacatg 660
ctcattcttg accccaagac tggcaagccg ctcccttctg aggatgaaga ccgacgcttc 720

ttcgaaggac cciggaattca caagcgcaac aagatttact acctcaccta ctctactggc 780
acaacccact atcttgtcta tgcgacttca aagacccctt atggctctta cacctaccag 840
ggcagaattc tggagccagt tgatggctgg actactcact ctagtatcgt caagtaccag 900
ggtcagtggg ggctatctta tcacgatgcc aagacatctg gcaaggacta tcttcgccag 960
gtaaaggcta agaagatttg gtacgatagc aaaggaaaga tcttgacaaa gaagccttga 1020

<210> 49
<211> 1038
<212> DNA
<213> *Fusarium oxysporum*
<400> 49

atgtatcgga agttggccgt catctcggcc ttcttggcca cagctcgtgc tcaagacact 60

aatgacattc ctccctgat caccgacctc tggctccgag atccctcggc tcatgttttc 120
gaaggcaagc tctgggttta cccatctcac gacatcgaag ccaatgttgt caacggcaca 180
ggaggcgctc aatagccat gagggattac catactact ccatgaagag catctatggg 240
aaagatcccc ttgtcgacca cggcgctcgt ctctcagtcg atgacgttcc ctgggcgaag 300
cagcaaatgt gggctcctga cgcagctcat aagaacggca aatattatct gtacttcccc 360
gccaaaggaca aggatgagat cttcagaatt ggagttgctg tctccaacaa gccagcggg 420
cctttcaagg ccgacaagag ctggatccct ggcacgtaca gtatcgatcc tgctagctac 480

gtcgacactg ataacgagge ctacctatc tggggcggta tctggggcgg ccagctccaa 540
gcctggcagg ataaaaagaa ctttaacgag tcgtggattg gagacaaggc tgctcctaac 600
ggcaccaatg ccctatctcc tcagatcgcc aagctaagca aggacatgca caagatcacc 660
gaaacacccc gcgatctcgt cattctcgcc cccgagacag gcaagcctct tcaggtgag 720
gacaacaagc gacgattctt cgagggccct tggatccaca agcgcggcaa gctttactac 780
ctcatgtact ccaccgtga taccacttc ctgtgtctag ctacttccaa gaacatctac 840
ggtccttata cctaccgggg caagattctt gatcctgttg atgggtggac tactcatgga 900

agtattgttg agtataaggg acagtgggtg cttttctttg ctgatgcga tacgtctggg 960
aaggattacc ttcgacaggt gaaggcgagg aagatctggg atgacaagaa cggcaagatc 1020
ttgcttcacc gtccttag 1038

<210> 50

<211> 1920

<212> DNA

<213> *Penicillium funiculosum*

<400> 50

atgtaccgga agctcgccgt gatcagcgcc ttcttgccga ctgctcgcg caccaccatc	60
aacgtcagcc agagcggcgg caacaagacc agcccgtcc agtacggcct catgttcgag	120
gacatcaacc acggcggcga cggcggcctc tacgccgagc tggtcggaa cggggccttc	180
cagggcagca ccgtctacc ggccaacctc gacggctacg actcggtgaa cggcgcgatt	240
ctcgcgctcc agaacctcac caaccgctc agcccagca tgcctcgtc gctgaacgtc	300
gccaagggtc cgaacaacgg cagcatcggc ttcgccaacg aggggtggtg gggcatcgag	360
gtcaagccgc agcggtagc cggcagcttc tacgtccagg gcgactacca gggcgacttc	420
gacatcagcc tccagagcaa gtcacccag gaggtcttcg cgacggcgaa ggtccggtcg	480
agcggcaagc acgaggactg ggtccagtac aagtacgagc tggtcggaa gaaggccgcc	540
agcaacacca acaacacct caccatcacc ttcgacagca agggcctcaa ggacggcagc	600
ctcaacttca acctcatcag cctcttcccg ccgacctaca acaaccggcc gaacggcctc	660
cggatcgacc tcgtcgaggc catggcggag ctggaggcca agttcctccg ctccccggc	720
ggctcggacg tggaggcgct ccagcccccg tactggtaca agtggaacga gaccgtcggc	780
gacctcaagg accgctactc gcgcccagc gcctggacct acgaggagag caacggcatc	840
ggcctcatcg agtacatgaa ctgggtcgac gacatgggcc tcgagccgat cctcgccgtc	900
tgggacggcc actacctcag caacagggtc atcagcgaga acgacctcca gccgtacatc	960
gacgacacc tcaaccagct cgagttctc atgggcgccc cggacactcc ctacgggtct	1020
tggagggcta gcctcggcta cccgaagccg tggacatca actacgtcga gatcggcaac	1080
gaggacaacc tctacggcgg cctcgagacc tacatgcct accggttcca ggcctactac	1140
gacgccatca ccgccaagta cccgcacatg accgtcatgg agagcctcac cgagatgccc	1200
ggccccgctg ccgcgcgctc ggactaccac cagtactcga cggccgacgg ctctgtcagc	1260
cagttcaact acttcgacca gatgccgtc accaaccgca cgtgaacgg cgagatcgcc	1320
accgtctacc ccaacaacc gagcaactcg gtggcggtgg gcagcccggt cccgtcttac	1380
ccgtggtgga tcgggtccgt ggctgaggcc gtcttcctca tcggcgagga gcggaacagc	1440
ccgaagatca tcggcgccag ctacgcccc atgttcgca acattaacaa ctggcagtg	1500
agcccgacct tgatcgctt cgacgccgac agcagccgga cgtcgcgctc tacttcttg	1560

cacgtcatca agctcctcag caccaacaag atcacccaga acctgcccac gacgtggtct 1620
 gggggggaca tcggcccgt ctactgggtc gccggccgga acgacaacac cggcagcaac 1680
 atcttcaagg ccgccgtcta caacagcacc agcgacgtcc cggtcaccgt ccagttcgcc 1740
 ggctgcaacg ccaagagcgc caacctcacc atcctctcgt cggacgaccc caacgccagc 1800
 aactaccgg gcggcccga ggtcgtcaag accgagatcc agagcgtcac cgccaacgcc 1860

cacggcgcct tcgagttcag cctcccgaac ctgtcgtgg ctgtgctgaa gacggagtag 1920

<210> 51

<211> 1044

<212> DNA

<213> Trichoderma reesei

<400> 51

atgatccaga agctttccaa cttttcttc accgcactag cggtggaac cggtgttgtt 60
 ggacacggac acatcaacaa cattgtcgtc aacggagtgt actaccaggg atatgatcct 120
 acatcgttcc catatgaatc tgaccgcgcc atagtgttg gctggacggc tgccgatctt 180
 gacaacggct tcgtctcacc cgacgcata cagagcccgg acatcatctg ccacaagaat 240
 gccaccaacg ccaaaggaca cgcgtccgtc aaggccggag acactattcc cctccagtgg 300

gtgccagttc ctigggcgca cccaggcccc atcgtcgact acctggccaa ctgcaacggc 360
 gactgcgaga ccgtggacaa gacgtccctt gagttcttca agattgacgg cgtcggcttc 420
 atcagcggcg gagatccggg caactgggcc tcggacgtgt tgattgcaa caacaacacc 480
 tgggttgtca agatccccga ggatctgcc ccgggcaact acgtgcttcg ccacgagatc 540
 atcgcttgc acagcgcggc gcaggcggac ggcgtcaga actaccctca gtgcttcaac 600
 ctgcgcgtcc caggctccgg atctctgcag ccgagcggcg tcaagggaac cgcgtctac 660
 cactccgatg accccggtgt cctcatcaac atctacacca gccctcttgc gtacaccatt 720

cctggacctt ccgtggatc aggcctcccc acgagtgtcg cccagggcag ctccgccg 780
 acggccactg ccagcgccac tgttcctggc gtagcggac cgggaaacc gaccagtaag 840
 actacgacga cggcgaggac gacacaggcc tcctctagca gggccagctc tactcctct 900
 gctactacgt cggcacctgg tggaggccca acccagactt tgtacggcca gtgtgtggc 960
 agcggctaca gtgtcctac tcgatgcgcg ccgccggcca cttgctctac cttgaacca 1020
 tactacgcc agtgccttaa ctag 1044

<210> 52

<211> 344

<212> PRT

<213> *Trichoderma reesei*

<400> 52

Met Ile Gln Lys Leu Ser Asn Leu Leu Val Thr Ala Leu Ala Val Ala

1 5 10 15

Thr Gly Val Val Gly His Gly His Ile Asn Asp Ile Val Ile Asn Gly

20 25 30

Val Trp Tyr Gln Ala Tyr Asp Pro Thr Thr Phe Pro Tyr Glu Ser Asn

35 40 45

Pro Pro Ile Val Val Gly Trp Thr Ala Ala Asp Leu Asp Asn Gly Phe

50 55 60

Val Ser Pro Asp Ala Tyr Gln Asn Pro Asp Ile Ile Cys His Lys Asn

65 70 75 80

Ala Thr Asn Ala Lys Gly His Ala Ser Val Lys Ala Gly Asp Thr Ile

85 90 95

Leu Phe Gln Trp Val Pro Val Pro Trp Pro His Pro Gly Pro Ile Val

100 105 110

Asp Tyr Leu Ala Asn Cys Asn Gly Asp Cys Glu Thr Val Asp Lys Thr

115 120 125

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

130 135 140

Asp Pro Gly Thr Trp Ala Ser Asp Val Leu Ile Ser Asn Asn Asn Thr

145 150 155 160

Trp Val Val Lys Ile Pro Asp Asn Leu Ala Pro Gly Asn Tyr Val Leu

165 170 175

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

180 185 190

Gln Asn Tyr Pro Gln Cys Phe Asn Ile Ala Val Ser Gly Ser Gly Ser

195 200 205

Leu Gln Pro Ser Gly Val Leu Gly Thr Asp Leu Tyr His Ala Thr Asp

210 215 220

Pro Gly Val Leu Ile Asn Ile Tyr Thr Ser Pro Leu Asn Tyr Ile Ile

225 230 235 240
 Pro Gly Pro Thr Val Val Ser Gly Leu Pro Thr Ser Val Ala Gln Gly
 245 250 255

Ser Ser Ala Ala Thr Ala Thr Ala Ser Ala Thr Val Pro Gly Gly Gly
 260 265 270
 Ser Gly Pro Thr Ser Arg Thr Thr Thr Thr Ala Arg Thr Thr Gln Ala
 275 280 285
 Ser Ser Arg Pro Ser Ser Thr Pro Pro Ala Thr Thr Ser Ala Pro Ala
 290 295 300
 Gly Gly Pro Thr Gln Thr Leu Tyr Gly Gln Cys Gly Gly Ser Gly Tyr
 305 310 315 320

Ser Gly Pro Thr Arg Cys Ala Pro Pro Ala Thr Cys Ser Thr Leu Asn
 325 330 335
 Pro Tyr Tyr Ala Gln Cys Leu Asn
 340

<210> 53

<211> 2260

<212> DNA

<213> Podospora anserina

<400> 53

atggctcttc aaaccttctt cctgctggcg gcagccatgc tggccaacgc agagacaaca	60
ggcgaaaagg tctctcgga agcaccgtct ggcgctcaag catgggccgc cgcccactcc	120
caggctgccg ccactctggc cagaatgtca cagcaagaca agatcaacat ggtcacgggc	180
attggctggg acagagggcc ttgcgtggga aacacagctg ccatcagctc catcaactat	240
cctcaaatct gtcttcagga tggaccattg ggcatctgct tcggcactgg taccacgcgc	300
ttcacacctg gcgtccaagc tgcttcgaca tgggacgttg atctgatccg gcagcgcggt	360
gcttacctgg gcgccgaagc caagggtgc ggcatcaca tccttttggg gcccggtgcc	420
ggtgccctgg gcaagattcc ccacggcggt cgcaactggg agggatttgg cgccgacccc	480
taccttgccg gtattgcat gaaggagacc atcgagggtt ttcagtcagc aggcgtccag	540
gccaacgcca agcactacat tgcaaacgaa caagagctca accgcgagac catgagcagc	600
aatgtggatg accgcactca gcacgagctc tacctctggc cctttgccga cgccgtgcac	660

gccaacgtcg ccagcgatcat gtgcagttac aacaagctca atggcacgtg ggcttgcgag 720
aatgacaagg ctctgaatca gatcttgaag aaggagctcg gattccaggg ctacgttctc 780
agcgactgga atgctcagca cagcactgct ctgtctgcta acagtggctt ggacatgact 840
atgcccggta ccgatttcaa cggccgcaat gtctactggg gccctcaact gaacaacgt 900
gtcaacgccg gccaggttca gagatccaga ctagacgaca tgtgcaagag aatcttggct 960
ggctggtaact tgcctggta gaaccagggc tatcccga tcaacatcag ggccaacgtt 1020
cagggcaacc ataaggagaa cgtacgtgct gttgccagag acggcatcgt cttgctgaag 1080

aacgatggaa ttctgccgt ttccaagccg agaaagattg ctgtcgtggg ctccactcc 1140
gtcaacaatc cccagggat caacgcctgt gttgacaagg gctgcaatgt tggcacctt 1200
ggcatgggct ggggttcagg cagcgtcaac taccctatc tcgtgtccc gtacgatgt 1260
ctccggactc gtctcagc cgatggcaca caaatcagcc tccacaacac tgacagcacc 1320
aacggtgtgt caaacgttgt gtctgacgt gatgctgtt ttgtgtcat cactgccgat 1380
tctggtgaag ggtacatcac tgcgagggc cagcgtggcg accgcagcca ccttgaccg 1440
tggcacaatg gcaaccaact tgttcaggct gccgcggctg ccaacaagaa cgtcatcgtt 1500

gttgtgcaca gtgttgcca gatcacctg gagactatcc tcaacaccaa tggagtccgc 1560
gcgattgtgt gggctggct tccgggcaa gagaatggca acgctcttgt tgatgttctc 1620
tacggcttgg ttctgccatc tggaaagctt cctacacca ttggcaagag ggagtccgac 1680
tatggcacag ccgttgttcg tgggatgat aacttcaggg agggcctttt tgttgactac 1740
cgtaactttg acaatgccag gatcgagccg cgtatgagt ttggctttgg tctttgtaag 1800
ttccagcggc ggagtgggt ttgatttcaa gctttcctaa cctgataaaa cagcttacac 1860
caatttcacc ttctccgaca tcaagattac ttccaatgtc aagccggggc ccgtactgg 1920

ccagaccatt cccggcggac ctgccgacct gtgggaggac gttgcgacag tcaactgcaac 1980
catcaccaac tcgggtgtcg tcgagggcgc tgaggttgcc cagctttaca tcggcctgcc 2040
gtcctcggct cctgcctctc cccgaagca gctgcgtgga tttccaagc tgaagctggc 2100
cccgggtgcc agcggcactg ccacattcaa cctcagacgc agagatctca gctattggga 2160
taccgcctc cagaactggg tcgtgccag cggcaacttt gtcgtcagcg tcggcgccag 2220
ctcgagagat atccgcttga cgggcacat cagggcgtag 2260

<210> 54

<211> 733

<212> PRT

<213> Podospira anserina

<400> 54

Met Ala Leu Gln Thr Phe Phe Leu Leu Ala Ala Ala Met Leu Ala Asn
1 5 10 15

Ala Glu Thr Thr Gly Glu Lys Val Ser Arg Gln Ala Pro Ser Gly Ala
20 25 30

Gln Ala Trp Ala Ala Ala His Ser Gln Ala Ala Ala Thr Leu Ala Arg
35 40 45

Met Ser Gln Gln Asp Lys Ile Asn Met Val Thr Gly Ile Gly Trp Asp
50 55 60

Arg Gly Pro Cys Val Gly Asn Thr Ala Ala Ile Ser Ser Ile Asn Tyr
65 70 75 80

Pro Gln Ile Cys Leu Gln Asp Gly Pro Leu Gly Ile Arg Phe Gly Thr
85 90 95

Gly Thr Thr Ala Phe Thr Pro Gly Val Gln Ala Ala Ser Thr Trp Asp
100 105 110

Val Asp Leu Ile Arg Gln Arg Gly Ala Tyr Leu Gly Ala Glu Ala Lys
115 120 125

Gly Cys Gly Ile His Ile Leu Leu Gly Pro Val Ala Gly Ala Leu Gly
130 135 140

Lys Ile Pro His Gly Gly Arg Asn Trp Glu Gly Phe Gly Ala Asp Pro
145 150 155 160

Tyr Leu Ala Gly Ile Ala Met Lys Glu Thr Ile Glu Gly Ile Gln Ser
165 170 175

Ala Gly Val Gln Ala Asn Ala Lys His Tyr Ile Ala Asn Glu Gln Glu
180 185 190

Leu Asn Arg Glu Thr Met Ser Ser Asn Val Asp Asp Arg Thr Gln His
195 200 205

Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val His Ala Asn Val Ala
210 215 220

Ser Val Met Cys Ser Tyr Asn Lys Leu Asn Gly Thr Trp Ala Cys Glu
225 230 235 240

Asn Asp Lys Ala Leu Asn Gln Ile Leu Lys Lys Glu Leu Gly Phe Gln

245												250												255											
Gly	Tyr	Val	Leu	Ser	Asp	Trp	Asn	Ala	Gln	His	Ser	Thr	Ala	Leu	Ser																				
260						265						270																							
Ala	Asn	Ser	Gly	Leu	Asp	Met	Thr	Met	Pro	Gly	Thr	Asp	Phe	Asn	Gly																				
275						280						285																							
Arg	Asn	Val	Tyr	Trp	Gly	Pro	Gln	Leu	Asn	Asn	Ala	Val	Asn	Ala	Gly																				
290						295						300																							
Gln	Val	Gln	Arg	Ser	Arg	Leu	Asp	Asp	Met	Cys	Lys	Arg	Ile	Leu	Ala																				
305						310						315						320																	
Gly	Trp	Tyr	Leu	Leu	Gly	Gln	Asn	Gln	Gly	Tyr	Pro	Ala	Ile	Asn	Ile																				
325						330						335																							
Arg	Ala	Asn	Val	Gln	Gly	Asn	His	Lys	Glu	Asn	Val	Arg	Ala	Val	Ala																				
340						345						350																							
Arg	Asp	Gly	Ile	Val	Leu	Leu	Lys	Asn	Asp	Gly	Ile	Leu	Pro	Leu	Ser																				
355						360						365																							
Lys	Pro	Arg	Lys	Ile	Ala	Val	Val	Gly	Ser	His	Ser	Val	Asn	Asn	Pro																				
370						375						380																							
Gln	Gly	Ile	Asn	Ala	Cys	Val	Asp	Lys	Gly	Cys	Asn	Val	Gly	Thr	Leu																				
385						390						395						400																	
Gly	Met	Gly	Trp	Gly	Ser	Gly	Ser	Val	Asn	Tyr	Pro	Tyr	Leu	Val	Ser																				
405						410						415																							
Pro	Tyr	Asp	Ala	Leu	Arg	Thr	Arg	Ala	Gln	Ala	Asp	Gly	Thr	Gln	Ile																				
420						425						430																							
Ser	Leu	His	Asn	Thr	Asp	Ser	Thr	Asn	Gly	Val	Ser	Asn	Val	Val	Ser																				
435						440						445																							
Asp	Ala	Asp	Ala	Val	Val	Val	Val	Ile	Thr	Ala	Asp	Ser	Gly	Glu	Gly																				
450						455						460																							
Tyr	Ile	Thr	Val	Glu	Gly	His	Ala	Gly	Asp	Arg	Ser	His	Leu	Asp	Pro																				
465						470						475						480																	
Trp	His	Asn	Gly	Asn	Gln	Leu	Val	Gln	Ala	Ala	Ala	Ala	Ala	Asn	Lys																				
485						490						495																							

Asn Val Ile Val Val Val His Ser Val Gly Gln Ile Thr Leu Glu Thr
500 505 510

Ile Leu Asn Thr Asn Gly Val Arg Ala Ile Val Trp Ala Gly Leu Pro
515 520 525

Gly Gln Glu Asn Gly Asn Ala Leu Val Asp Val Leu Tyr Gly Leu Val
530 535 540

Ser Pro Ser Gly Lys Leu Pro Tyr Thr Ile Gly Lys Arg Glu Ser Asp
545 550 555 560

Tyr Gly Thr Ala Val Val Arg Gly Asp Asp Asn Phe Arg Glu Gly Leu
565 570 575

Phe Val Asp Tyr Arg His Phe Asp Asn Ala Arg Ile Glu Pro Arg Tyr
580 585 590

Glu Phe Gly Phe Gly Leu Ser Tyr Thr Asn Phe Thr Phe Ser Asp Ile
595 600 605

Lys Ile Thr Ser Asn Val Lys Pro Gly Pro Ala Thr Gly Gln Thr Ile
610 615 620

Pro Gly Gly Pro Ala Asp Leu Trp Glu Asp Val Ala Thr Val Thr Ala
625 630 635 640

Thr Ile Thr Asn Ser Gly Ala Val Glu Gly Ala Glu Val Ala Gln Leu
645 650 655

Tyr Ile Gly Leu Pro Ser Ser Ala Pro Ala Ser Pro Pro Lys Gln Leu
660 665 670

Arg Gly Phe Ser Lys Leu Lys Leu Ala Pro Gly Ala Ser Gly Thr Ala
675 680 685

Thr Phe Asn Leu Arg Arg Arg Asp Leu Ser Tyr Trp Asp Thr Arg Leu
690 695 700

Gln Asn Trp Val Val Pro Ser Gly Asn Phe Val Val Ser Val Gly Ala
705 710 715 720

Ser Ser Arg Asp Ile Arg Leu Thr Gly Thr Ile Thr Ala
725 730

<210> 55

<211> 2551

<212> DNA

<213> *Fusarium verticillioides*

<400> 55

atgtttcctt cttccatata ttgtttggcg gccctgagtc tgatgagcca gggcttacta	60
gtcagagcc aaccggaaaa tgtcatcacc gatgatacct acttctacgg tcaatcgcca	120
ccagtgtatc ctacacgtaa gcactctctc tgatttccca acgaaagcaa tactgatctc	180
ttgaccagcg gaacaggtag acaccggctc atgggctgcc gctgtagcca aagccaagaa	240
cttgggtgcc cagttgactc ttgaagagaa agtcaacttg actacaggag gccagacgac	300
caccggctgc tctggcttca tccctggcat tccccgtga ggctttccag gactgtgttt	360
agcagacgct ggcaacggtg tccgcaacac agattatgtg agctcgtttc cctccgggat	420
tcatgtcggg gcaagctgga atccggagtt gacctacagc cggagctact acatgggtgc	480
tgaggccaaa gccaaaggcg ttaacatcct tctcggcca gtatttgac ctttgggccg	540
agtagttgaa ggtggacgca actgggaggg gttttcaat gatccctacc tggcgggtaa	600
attagggcat gaagctgtcg ccggtatcca agacgccgga gttgttgcac gcggaaaaca	660
tttcttgcct caagagcagg agaccatag acttgcggcg tctgtcactg gggctgatgc	720
aatctcatca aatctcgatg acaagacact ccatgaatta tatctctggt aagcacatca	780
tatcttggct gtagtagta acctactaa caccgaact gggcttttcg ctgatgcagt	840
ccacgccgga cttgccagtg tgatgtgcag ctacaacaga gcaacaatt cacacgcctg	900
ccaaaactcg aagcttctca atggccttct caaggcgag ttaggattcc agggttttgt	960
cgtctcggac tggggcgcac agcaatctgg tatggcttca gcattggctg gcctggatgt	1020
tgatcatccc agctcgatct tgtggggtgc caaccttacc cttggtgtga acaacggaac	1080
tattcccagc tcacagggtg acaatatggt tacacggtac gcgaagtctc agccttactt	1140
ctcaattctt ttgaactgac aatcgtgtag gctccttgca acttggatc agttgaacca	1200
ggaccaagac accgaagccc caggtcacgg actcgctgcc aagctttggg agcctcacc	1260
agtagtcgac gctcgcaacg caagctcaa gcctactatc tgggacgggtg cagtcgaggg	1320
ccatgttctt gttagaaca ccaacaacgc actgccattc aagcccaaca tgaaactcgt	1380
ttctttgttc ggatactctc acaaagctcc tgataagaac atcccagacc ccgccaagg	1440
catgtttctc gcttggctta tcggtgccca atccgccaac atcactgagc tgaacctcgg	1500
ctttctcgga aatttgagtc tcacatactc cgccatcgcg cccaacggaa ccatcatctc	1560
gggtggaggc tcgggtgcca gcgcttgac tctgttcagc tcacccttcg atgcattcgt	1620

ttctcgggcg aagaaagagg gtactgcgct ttctcgggat tttgagagct gggatcctta 1680
 tgtgaacctt acatctgaag cttgcatcgt tgctggtaat gcatgggcta gcaaggctg 1740
 ggatagacct gcaacctatg atgcctatac tgatgagctc atcaataacg tcgctgacaa 1800
 gtgcgctaac actattgttg ttcttcacaa tgctggaaca cgacttgtgg atggcttctt 1860

tggtcacccc aacgtcacgg ctattatcta cgctcatctc ccaggtcagg atagtggaga 1920
 tgctctggta tctttgctct atggcgatga gaacccatct ggctgcctcc cttaaccgt 1980
 tgcccgaac gagacggatt atggtcacct gctgaagcca gacttgactc tcgccccaa 2040
 ccagtaccaa cactttcccc agtcggactt ctccgagggt attttcattg actaccgaca 2100
 ttctgatgct aagaacatca cgcctcgctt cgagtttggg ttccggcttga gctacacaac 2160
 ctttgagtac gctagtctcc agatctcaaa gtcccaggcc cagacaccgg aatacccgagc 2220
 tgggtgctctt accgaggagg gccgttcaga ttgtggggac gtcgttgcta ctgtcacagc 2280

aagcgtcagg aacactgggt ctgtcgacgg caaggagggt gcacagctat acgttggtgt 2340
 tccaggtggg cctatgagac agctacgtgg ctttacgaaa ccagctatta aggctggaga 2400
 gacggctaca gtgacctttg agcttactcg ccgcgacttg agtgtctggg atgttaatgc 2460
 gcaggagtgg caacttcagc aaggcaacta tgctatctac gttggccgaa gtagtcgaga 2520
 ttgcctctg caaagtacct tgagcatcta g 2551

<210> 56

<211> 780

<212> PRT

<213> *Fusarium verticillioides*

<400> 56

Met Phe Pro Ser Ser Ile Ser Cys Leu Ala Ala Leu Ser Leu Met Ser

1	5	10	15
Gln Gly Leu Leu Ala Gln Ser Gln Pro Glu Asn Val Ile Thr Asp Asp			
20	25	30	
Thr Tyr Phe Tyr Gly Gln Ser Pro Pro Val Tyr Pro Thr His Thr Gly			
35	40	45	
Ser Trp Ala Ala Ala Val Ala Lys Ala Lys Asn Leu Val Ser Gln Leu			
50	55	60	
Thr Leu Glu Glu Lys Val Asn Leu Thr Thr Gly Gly Gln Thr Thr Thr			

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

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

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

 385 390 395 400
 Pro Asp Lys Asn Ile Pro Asp Pro Ala Gln Gly Met Phe Ser Ala Trp
 405 410 415
 Ser Ile Gly Ala Gln Ser Ala Asn Ile Thr Glu Leu Asn Leu Gly Phe
 420 425 430
 Leu Gly Asn Leu Ser Leu Thr Tyr Ser Ala Ile Ala Pro Asn Gly Thr
 435 440 445
 Ile Ile Ser Gly Gly Gly Ser Gly Ala Ser Ala Trp Thr Leu Phe Ser

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

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

Leu Tyr Gly Asp Glu Asn Pro Ser Gly Arg Leu Pro Tyr Thr Val Ala

580 585 590

Arg Asn Glu Thr Asp Tyr Gly His Leu Leu Lys Pro Asp Leu Thr Leu

595 600 605

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

610 615 620

Ile Phe Ile Asp Tyr Arg His Phe Asp Ala Lys Asn Ile Thr Pro Arg

625 630 635 640

Phe Glu Phe Gly Phe Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ala Ser

645 650 655

Leu Gln Ile Ser Lys Ser Gln Ala Gln Thr Pro Glu Tyr Pro Ala Gly

660 665 670

Ala Leu Thr Glu Gly Gly Arg Ser Asp Leu Trp Asp Val Val Ala Thr

675 680 685

Val Thr Ala Ser Val Arg Asn Thr Gly Ser Val Asp Gly Lys Glu Val

690 695 700

Ala Gln Leu Tyr Val Gly Val Pro Gly Gly Pro Met Arg Gln Leu Arg

705 710 715 720

Gly Phe Thr Lys Pro Ala Ile Lys Ala Gly Glu Thr Ala Thr Val Thr

725 730 735

Phe Glu Leu Thr Arg Arg Asp Leu Ser Val Trp Asp Val Asn Ala Gln

740 745 750

Glu Trp Gln Leu Gln Gln Gly Asn Tyr Ala Ile Tyr Val Gly Arg Ser

755 760 765

Ser Arg Asp Leu Pro Leu Gln Ser Thr Leu Ser Ile

770 775 780

<210> 57

<211> 2487

<212> DNA

<213> *Fusarium verticillioides*

<400> 57

atggctagca ttcgatctgt gttggtctcg ggtcttttgg ccgcgggtgt caatgcccaa	60
gcctacgatg cgagtgatcg cgctgaagat gctttcagct gggtcagacc caagaacacc	120
actattcttg gacagtacgg ccattcgctt cattaccctg ccagtatgtt caccaactac	180
accaagtgc acigaggctg tactgacatt ctagacaatg ctactggcaa gggctgggaa	240
gatgccttcg ccaaggctca aaactttgtc tccaactaa ccctcgagga aaaggccgac	300
atggtcacag gaactccagg tccttgcgtc ggcaacatcg tcgccattcc ccgtctcaac	360
ttcaacggtc tcgtcttca cgacggcccc ctgccatcc gagtagcaga ctacgccagt	420
gttttccccg ctggtgtatc agccgcttca tcgtgggaca aggacctcct ctaccagcgc	480
ggtctcgcca tgggtcaaga gttcaaggcc aagggtgctc acatcctcct cggccccgtc	540
gccggtctc ttggccgctc ggcatactct ggtcgttaact gggagggttt ctgccggac	600
ccttacctca ctggtattgc gatggaggag actatcatgg gacatcaaga tgctggtgtt	660
caggctactg cgaagcactt tatcggtaat gaggaggagg tcatgcgaaa ccctactttt	720
gtcaaggatg ggtatatgg tgaggttgac aaggaggctc tttcgtctaa catggatgat	780
cgaacctatg acgagcttta cctctggccc ttgccaatg ctgttcatgc caaggttcc	840
agcatgatgt gctcgtacca gcgtctcaac ggctcctacg cctgccagaa ctcaaaggtc	900
ctcaacggaa ttctgcgtga tgagcttggg ttccagggtt acgtcatgtc agattggggg	960
gccaccacg ccggtgttgc tgccatcaac agcgggtctc acatggacat gcccgttgg	1020
atcgggtcct acggaacata ctttaccag tccttcttcg gcggcaacct caccgcgcc	1080
gtaccaacg gcacctcga cgagaccgc gtcaacgaca tgatcacccg catcatgact	1140
ccctacttct ggctcgcca ggacaaggac tatccctccg tcgaccctc cagcgtgat	1200
ctcaaacct tcagcccaa gagtccttg ttccgcgagt tcaacctcac cggcgagcgc	1260
agccgtgacg tccggttaa ccacggcgac ttgatcgcga agcacggcgc cgagtctacc	1320
gtctttctca agaacgagaa gaacgccctt ccctcaaga agcccaagtc catcgtgtc	1380
tttggcaacg atgtgtga tatcactgag ggtttctaca accagaatga ctacgaattt	1440
ggcactcttg ttgtgttg ttggtctgga actggtcgtt tgacatact tgtttgcct	1500
ctagccgcca tcaatgctcg tgctaagcag gacgtactc ttgttcagca gtggatgaac	1560
aacactctta ttgtaccac caacgtcact gatctctgga tcctgtctac tcccgatgtc	1620
tgctctgtt tctgaagac ttgggtgag gaggtgctg atcgtgagca cctctccgtt	1680
gactgggacg gtaatgatg tgttgagtct gttgccaagt actgcaataa cactgtcgtc	1740
gtcactcact cttctggat caacactctt ccttgggtg accacccaa cgtcacccgt	1800
attctcgtg cccacttccc cggtcaggag tctggcaact ccctcgttga cctcctctac	1860

ggcgaatgca acccctctgg tcgtcttccc tacacatcg ccttcaacgg caccgactac 1920
aacgctcccc ccaccactgc cgtcaacacc accggcaagg aggactggca gtcttggttc 1980

gacgagaagc tcgagattga ctaccgctac ttcgacgcgc acaacatctc cgtccgctac 2040
gaattcggct tcggtctctc ctactccacc ttcgaaatct ccgacatctc cgctgagcca 2100
ctcgcatccg acattacctc ccagcccag gatctccccg tgcagcccgg cggcaacccc 2160
gccctctggg agaccgtcta caacgtgacc gtctccgtct ccaacacggg caaggtcgac 2220
ggcgccactg tccccagct atacgtgaca tccccgaca gcgcgcctgc cggtacacca 2280
cccaagcagc tcggtgggtt cgacaaggtc ttccttgagg ctggcgagag caagagtgtc 2340
agctttgagc tgatgcgccg tgatctgagc tactgggata tcatttctca gaagtggctc 2400

atccctgagg gagagtttac tattcgtgtt ggattcagca gtcgggactt gaaggaggag 2460
acaaaggtta ctgttgttga ggcgtaa 2487

<210> 58
<211> 811
<212> PRT
<213> Fusarium verticillioides
<400> 58

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

115 120 125
 Val Ser Ala Ala Ser Ser Trp Asp Lys Asp Leu Leu Tyr Gln Arg Gly
 130 135 140
 Leu Ala Met Gly Gln Glu Phe Lys Ala Lys Gly Ala His Ile Leu Leu
 145 150 155 160
 Gly Pro Val Ala Gly Pro Leu Gly Arg Ser Ala Tyr Ser Gly Arg Asn

 165 170 175
 Trp Glu Gly Phe Ser Pro Asp Pro Tyr Leu Thr Gly Ile Ala Met Glu
 180 185 190
 Glu Thr Ile Met Gly His Gln Asp Ala Gly Val Gln Ala Thr Ala Lys
 195 200 205
 His Phe Ile Gly Asn Glu Gln Glu Val Met Arg Asn Pro Thr Phe Val
 210 215 220
 Lys Asp Gly Tyr Ile Gly Glu Val Asp Lys Glu Ala Leu Ser Ser Asn

 225 230 235 240
 Met Asp Asp Arg Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asn
 245 250 255
 Ala Val His Ala Lys Ala Ser Ser Met Met Cys Ser Tyr Gln Arg Leu
 260 265 270
 Asn Gly Ser Tyr Ala Cys Gln Asn Ser Lys Val Leu Asn Gly Ile Leu
 275 280 285
 Arg Asp Glu Leu Gly Phe Gln Gly Tyr Val Met Ser Asp Trp Gly Ala

 290 295 300
 Thr His Ala Gly Val Ala Ala Ile Asn Ser Gly Leu Asp Met Asp Met
 305 310 315 320
 Pro Gly Gly Ile Gly Ala Tyr Gly Thr Tyr Phe Thr Lys Ser Phe Phe
 325 330 335
 Gly Gly Asn Leu Thr Arg Ala Val Thr Asn Gly Thr Leu Asp Glu Thr
 340 345 350
 Arg Val Asn Asp Met Ile Thr Arg Ile Met Thr Pro Tyr Phe Trp Leu

 355 360 365

Gly Gln Asp Lys Asp Tyr Pro Ser Val Asp Pro Ser Ser Gly Asp Leu
 370 375 380
 Asn Thr Phe Ser Pro Lys Ser Ser Trp Phe Arg Glu Phe Asn Leu Thr
 385 390 395 400
 Gly Glu Arg Ser Arg Asp Val Arg Gly Asn His Gly Asp Leu Ile Arg
 405 410 415
 Lys His Gly Ala Glu Ser Thr Val Leu Leu Lys Asn Glu Lys Asn Ala
 420 425 430
 Leu Pro Leu Lys Lys Pro Lys Ser Ile Ala Val Phe Gly Asn Asp Ala
 435 440 445
 Gly Asp Ile Thr Glu Gly Phe Tyr Asn Gln Asn Asp Tyr Glu Phe Gly
 450 455 460
 Thr Leu Val Ala Gly Gly Gly Ser Gly Thr Gly Arg Leu Thr Tyr Leu
 465 470 475 480
 Val Ser Pro Leu Ala Ala Ile Asn Ala Arg Ala Lys Gln Asp Gly Thr
 485 490 495
 Leu Val Gln Gln Trp Met Asn Asn Thr Leu Ile Ala Thr Thr Asn Val
 500 505 510
 Thr Asp Leu Trp Ile Pro Ala Thr Pro Asp Val Cys Leu Val Phe Leu
 515 520 525
 Lys Thr Trp Ala Glu Glu Ala Ala Asp Arg Glu His Leu Ser Val Asp
 530 535 540
 Trp Asp Gly Asn Asp Val Val Glu Ser Val Ala Lys Tyr Cys Asn Asn
 545 550 555 560
 Thr Val Val Val Thr His Ser Ser Gly Ile Asn Thr Leu Pro Trp Ala
 565 570 575
 Asp His Pro Asn Val Thr Ala Ile Leu Ala Ala His Phe Pro Gly Gln
 580 585 590
 Glu Ser Gly Asn Ser Leu Val Asp Leu Leu Tyr Gly Asp Val Asn Pro
 595 600 605
 Ser Gly Arg Leu Pro Tyr Thr Ile Ala Phe Asn Gly Thr Asp Tyr Asn

610 615 620
Ala Pro Pro Thr Thr Ala Val Asn Thr Thr Gly Lys Glu Asp Trp Gln
625 630 635 640
Ser Trp Phe Asp Glu Lys Leu Glu Ile Asp Tyr Arg Tyr Phe Asp Ala
 645 650 655
His Asn Ile Ser Val Arg Tyr Glu Phe Gly Phe Gly Leu Ser Tyr Ser
 660 665 670
Thr Phe Glu Ile Ser Asp Ile Ser Ala Glu Pro Leu Ala Ser Asp Ile

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

 740 745 750
Val Phe Leu Glu Ala Gly Glu Ser Lys Ser Val Ser Phe Glu Leu Met
 755 760 765
Arg Arg Asp Leu Ser Tyr Trp Asp Ile Ile Ser Gln Lys Trp Leu Ile
 770 775 780
Pro Glu Gly Glu Phe Thr Ile Arg Val Gly Phe Ser Ser Arg Asp Leu
785 790 795 800
Lys Glu Glu Thr Lys Val Thr Val Val Glu Ala
 805 810

<210> 59

<211> 3269

<212> DNA

<213> Fusarium verticillioides

<400> 59

atgaagctga attgggtcgc cgcagccctg tctataggtg ctgctggcac tgacagcgca 60

gttgctcttg cttctgcagt tccagacact ttggctggtg taaaggtcag ttttttttca 120

ccatttcctc gtctaattctc agccttgttg ccatatcgcc cttgttcgct cggacgccac 180

gcaccagatc gcgatcattt cctcccttgc agccttgggt cctcttacga tcttccctcc 240

gcaattatca gcgcccttag tctacacaaa aacccccgag acagtctttc attgagtttg 300

tcgacatcaa gttagcttctc aactgtgcat ttgcgtggct gtctacttct gcctctagac 360

aaccaaattct gggcgcaatt gaccgctcaa accttgttca aataaccttt ttatttcgag 420

acgcacattt ataaatatgc gcctttcaat aataccgact ttatgcgcgg cggtctctgt 480

ggcgggtgat cagaaagctg acgctcaaaa ggttgtcacg agagatacac tcgcatactc 540

gccgcctcat tatccttcac catggatgga ccctaagtct gttggctggg aggaagctta 600

cgccaaagcc aagagctttg tgtcccaact cactctcatg gaaaaggta acttgaccac 660

tgggtgttgg taagcagctc cttgcaaaca gggtatctca atccctcag ctaacaactt 720

ctcagatggc aaggcgaacg ctgtgttagga aacgtgggat caattctctg tctcggtatg 780

cgaggctctc gtctccagga tggctctctt ggaattcgtc tgtccgacta caacagcgct 840

tttcccgtg gcaccacagc tgggtcttct tggagcaagt ctctctggtg tgagagaggt 900

ctctgatgg gcactgagtt caaggagaag ggtatcgata tcgctcttgg tcctgtact 960

ggacctcttg gtcgcactgc tgcgtgtgga cgaaactggg aaggcttcac cgttgatcct 1020

tatatggctg gccacgcat ggccgaggcc gtcaagggtg ttcaagacgc aggtgtcatt 1080

gcttgtgcta agcattacat cgcaaacgag cagggtgaagc cacttggacg atttgaggaa 1140

ttgacagaga actgaccctc ttgtagagca ctccgacag agtggcgagg tccagtcccg 1200

caagtacaac atctccgagt ctctctctc caacctggat gacaagacta tgcacgagct 1260

ctacgcctgg cccttcgctg acgccgtccg cgccggcgtc ggttccgtca tgtgctcgta 1320

caaccagatc aacaactcgt acggttgcca gaactccaag ctctcaacg gtatctcaa 1380

ggacgagatg ggcttccagg gtttcgtcat gagcgattgg gcggcccagc ataccggtgc 1440

cgcttctgcc gtcgctggtc tcgatatgag catgcctggg gacactgcct tcgacagcgg 1500

atacagcttc tggggcgga acttgactct ggctgtcatc aacggaactg ttcccgcctg 1560

gcgagttgat gacatggctc tcggaatcat gtctgccttc ttcaaggttg gaaagacgat 1620

agaggatctt cccgacatca acttctctc ctggaccgcg gacaccttcg gcttcgtgca 1680

tacatttgcg caagagaacc gcgagcaggt caactttgga gtcaacgtcc agcacgacca 1740

caagagccac atccgtgagg ccgctgccaa gggaagcgtc gtgctcaaga acaccgggtc 1800

ccttccctc aagaaccaa agttctctgc tgcattggg gaggacgccg gtcccaaccc 1860

tgttggaacc aatggttgtg gtgaccgtgg ttgcgataat ggtaccctgg ctatggcttg 1920

gggctcggga acttcccaat tcccttactt gatcaccccc gatcaagggc tctctaactg 1980

agctactcaa gacggaactc gatatgagag catcttgacc aacaacgaat gggcttcagt 2040

acaagctctt gtcagccagc ctaacgtgac cgctatcggtt ttcgccaatg ccgactctgg 2100

tgagggatac attgaagtcg acggaaactt tggatgatgc aagaacctca ccctctggca 2160

gcagggagac gagctcatca agaactgttc gtccatgatc cccaacacca ttgtagttct 2220

gcacaccgtc ggcctgttc tactcgccga ctacgagaag aacccaaca tcaactgcat 2280

cgtctgggct ggtcttcccg gccaaagatc aggcgaatgc atcgctgatc tcctctacgg 2340

caaggtcagc cctggccgat ctcccttcac ttggggccgc acccgcgaga gctacggtac 2400

tgaggttctt tatgaggcga acaacggccg tggcgctcct caggatgact tctctgaggg 2460

tgctttcatc gactaccgtc acttcgaccg acgatctcca agcaccgatg gaaagagctc 2520

tccaacaac accgtgtctc ctctctacga gttcggtcac ggtctatctt ggtccacctt 2580

tgagtactct gacctcaaca tccagaagaa cgtcgagaac cctactctc ctcccgtgg 2640

ccagaccatc cccgcccac cttttggcaa ctccagcaag aacctcaacg actacgtgtt 2700

cccaagggc gtccgataca tctacaagt ctctacccc ttctcaaca cctctcatc 2760

cgccagcgag gcatccaacg atggtggcca gtttgtaag actgccgaag agttctccc 2820

tccaacgcc ctcaacggct cagcccagcc tcgtcttccc gcctctggtg cccaggtgg 2880

taacctcaa ttgtgggaca tttgttacac cgtcacagcc acaatcacca acacaggcaa 2940

cgccacctcc gacgagattc ccagctgia tgcagcctc ggtggcgaga acgagcccat 3000

ccgtgttctc cgcggtttcg accgtatcga gaacattgct cccggccaga gcgccatctt 3060

caacgtcaa ttgaccgtc gcgatctgag taactgggat acaaatgccc agaactgggt 3120

catcactgac catccaaga ctgtctgggt tggaagcagc tctcgcaagc tgcctctcag 3180

cgccaaagtg gactaagaaa gccaaacaag ggttgtttt ttgactgcaa ttttttggga 3240

ggacatagta gccgcgcgcc agttacgtc 3269

<210> 60

<211> 899

<212> PRT

<213> *Fusarium verticillioides*

<400> 60

Met Lys Leu Asn Trp Val Ala Ala Ala Leu Ser Ile Gly Ala Ala Gly

1 5 10 15

Thr Asp Ser Ala Val Ala Leu Ala Ser Ala Val Pro Asp Thr Leu Ala

20 25 30

Gly Val Lys Lys Ala Asp Ala Gln Lys Val Val Thr Arg Asp Thr Leu
 35 40 45
 Ala Tyr Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp Pro Asn Ala
 50 55 60
 Val Gly Trp Glu Glu Ala Tyr Ala Lys Ala Lys Ser Phe Val Ser Gln
 65 70 75 80
 Leu Thr Leu Met Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gln
 85 90 95
 Gly Glu Arg Cys Val Gly Asn Val Gly Ser Ile Pro Arg Leu Gly Met
 100 105 110
 Arg Gly Leu Cys Leu Gln Asp Gly Pro Leu Gly Ile Arg Leu Ser Asp
 115 120 125
 Tyr Asn Ser Ala Phe Pro Ala Gly Thr Thr Ala Gly Ala Ser Trp Ser
 130 135 140
 Lys Ser Leu Trp Tyr Glu Arg Gly Leu Leu Met Gly Thr Glu Phe Lys
 145 150 155 160
 Glu Lys Gly Ile Asp Ile Ala Leu Gly Pro Ala Thr Gly Pro Leu Gly
 165 170 175
 Arg Thr Ala Ala Gly Gly Arg Asn Trp Glu Gly Phe Thr Val Asp Pro
 180 185 190
 Tyr Met Ala Gly His Ala Met Ala Glu Ala Val Lys Gly Ile Gln Asp
 195 200 205
 Ala Gly Val Ile Ala Cys Ala Lys His Tyr Ile Ala Asn Glu Gln Glu
 210 215 220
 His Phe Arg Gln Ser Gly Glu Val Gln Ser Arg Lys Tyr Asn Ile Ser
 225 230 235 240
 Glu Ser Leu Ser Ser Asn Leu Asp Asp Lys Thr Met His Glu Leu Tyr
 245 250 255
 Ala Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met
 260 265 270
 Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Gln Asn Ser Lys

275 280 285
 Leu Leu Asn Gly Ile Leu Lys Asp Glu Met Gly Phe Gln Gly Phe Val
 290 295 300
 Met Ser Asp Trp Ala Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala

 305 310 315 320
 Gly Leu Asp Met Ser Met Pro Gly Asp Thr Ala Phe Asp Ser Gly Tyr
 325 330 335
 Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Ile Asn Gly Thr Val
 340 345 350
 Pro Ala Trp Arg Val Asp Asp Met Ala Leu Arg Ile Met Ser Ala Phe
 355 360 365
 Phe Lys Val Gly Lys Thr Ile Glu Asp Leu Pro Asp Ile Asn Phe Ser

 370 375 380
 Ser Trp Thr Arg Asp Thr Phe Gly Phe Val His Thr Phe Ala Gln Glu
 385 390 395 400
 Asn Arg Glu Gln Val Asn Phe Gly Val Asn Val Gln His Asp His Lys
 405 410 415
 Ser His Ile Arg Glu Ala Ala Ala Lys Gly Ser Val Val Leu Lys Asn
 420 425 430
 Thr Gly Ser Leu Pro Leu Lys Asn Pro Lys Phe Leu Ala Val Ile Gly

 435 440 445
 Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys Gly Asp Arg
 450 455 460
 Gly Cys Asp Asn Gly Thr Leu Ala Met Ala Trp Gly Ser Gly Thr Ser
 465 470 475 480
 Gln Phe Pro Tyr Leu Ile Thr Pro Asp Gln Gly Leu Ser Asn Arg Ala
 485 490 495
 Thr Gln Asp Gly Thr Arg Tyr Glu Ser Ile Leu Thr Asn Asn Glu Trp

 500 505 510
 Ala Ser Val Gln Ala Leu Val Ser Gln Pro Asn Val Thr Ala Ile Val
 515 520 525

Phe Ala Asn Ala Asp Ser Gly Glu Gly Tyr Ile Glu Val Asp Gly Asn
 530 535 540
 Phe Gly Asp Arg Lys Asn Leu Thr Leu Trp Gln Gln Gly Asp Glu Leu
 545 550 555 560
 Ile Lys Asn Val Ser Ser Ile Cys Pro Asn Thr Ile Val Val Leu His
 565 570 575
 Thr Val Gly Pro Val Leu Leu Ala Asp Tyr Glu Lys Asn Pro Asn Ile
 580 585 590
 Thr Ala Ile Val Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ala
 595 600 605
 Ile Ala Asp Leu Leu Tyr Gly Lys Val Ser Pro Gly Arg Ser Pro Phe
 610 615 620
 Thr Trp Gly Arg Thr Arg Glu Ser Tyr Gly Thr Glu Val Leu Tyr Glu
 625 630 635 640
 Ala Asn Asn Gly Arg Gly Ala Pro Gln Asp Asp Phe Ser Glu Gly Val
 645 650 655
 Phe Ile Asp Tyr Arg His Phe Asp Arg Arg Ser Pro Ser Thr Asp Gly
 660 665 670
 Lys Ser Ser Pro Asn Asn Thr Ala Ala Pro Leu Tyr Glu Phe Gly His
 675 680 685
 Gly Leu Ser Trp Ser Thr Phe Glu Tyr Ser Asp Leu Asn Ile Gln Lys
 690 695 700
 Asn Val Glu Asn Pro Tyr Ser Pro Pro Ala Gly Gln Thr Ile Pro Ala
 705 710 715 720
 Pro Thr Phe Gly Asn Phe Ser Lys Asn Leu Asn Asp Tyr Val Phe Pro
 725 730 735
 Lys Gly Val Arg Tyr Ile Tyr Lys Phe Ile Tyr Pro Phe Leu Asn Thr
 740 745 750
 Ser Ser Ser Ala Ser Glu Ala Ser Asn Asp Gly Gly Gln Phe Gly Lys
 755 760 765
 Thr Ala Glu Glu Phe Leu Pro Pro Asn Ala Leu Asn Gly Ser Ala Gln

770 775 780
 Pro Arg Leu Pro Ala Ser Gly Ala Pro Gly Gly Asn Pro Gln Leu Trp
 785 790 795 800
 Asp Ile Leu Tyr Thr Val Thr Ala Thr Ile Thr Asn Thr Gly Asn Ala
 805 810 815
 Thr Ser Asp Glu Ile Pro Gln Leu Tyr Val Ser Leu Gly Gly Glu Asn

 820 825 830
 Glu Pro Ile Arg Val Leu Arg Gly Phe Asp Arg Ile Glu Asn Ile Ala
 835 840 845
 Pro Gly Gln Ser Ala Ile Phe Asn Ala Gln Leu Thr Arg Arg Asp Leu
 850 855 860
 Ser Asn Trp Asp Thr Asn Ala Gln Asn Trp Val Ile Thr Asp His Pro
 865 870 875 880
 Lys Thr Val Trp Val Gly Ser Ser Ser Arg Lys Leu Pro Leu Ser Ala

 885 890 895
 Lys Leu Glu

<210> 61

<211> 2370

<212> DNA

<213> Trichoderma reesei

<400> 61

atgcgttacc gaacagcagc tgcgctggca cttgccactg ggccctttgc tagggcagac	60
agtcagtata gctgggtccca tactgggatg tgatatgtat cctggagaca ccatgctgac	120
tcttgaatca aggtagctca acatcggggg cctcggctga ggagtttga cctcctgcag	180
ggactccatg gggaaccgcg tacgacaagg cgaaggccgc attggcaaag ctcaatctcc	240
aagataaggt cggcatcgtg agcgggtgtcg gctggaacgg cggtccttgc gttggaaca	300
catctccggc ctccaagatc agctatccat cgctatgcct tcaagacgga cccctcgggtg	360
ttcgatactc gacaggcagc acagccttta cgccgggcgt tcaagcggcc tcgacgtggg	420
atgtcaattt gatccgcgaa cgtggacagt tcacggtga ggaggtgaag gcctcgggga	480
ttcatgtcat acttggctcct gtggctgggc cgctgggaaa gactccgcag ggcggtcgca	540
actgggaggg cttcgggtgc gatccatatc tcacgggcat tgccatgggt caaacatca	600

acggcatcca gtcggtaggc gtgcaggcga cagcgaagca ctatatactc aacgagcagg	660
agctcaatcg agaaaccatt tcgagcaacc cagatgaccg aactctccat gagctgtata	720
cttggccatt tgccgacgcg gttcaggcca atgtcgettc tgtcatgtgc tcgtacaaca	780
aggtcaatac cacctgggcc tgcgaggatc agtacacgct gcagactgtg ctgaaagacc	840
agctggggtt cccaggctat gtcatgacgg actggaacgc acagcacacg actgtccaaa	900
gcgcgaattc tgggcttgac atgtcaatgc ctggcacaga cttcaacggt aacaatcggc	960
tctgggggcc agctctcacc aatgcggtaa atagcaatca ggtccccacg agcagagtcg	1020
acgatatggt gactcgtatc ctgcgccgat ggtacttgac aggccaggac caggcaggct	1080
atccgtcgtt caacatcagc agaaatgttc aaggaaacca caagaccaat gtcagggcaa	1140
ttgccaggga cgcatcgtt ctgctcaaga atgacgcaa catcctgccg ctcaagaagc	1200
ccgctagcat tgccgtcgtt ggatctgccg caatcattgg taaccacgcc agaaactcgc	1260
cctcgtgcaa cgacaaaggc tgcgacgacg gggccttggg catgggttgg ggttcggcg	1320
ccgtcaacta tccgtacttc gtcgcgccct acgatgccat caataccaga gcgtcttcgc	1380
agggcaccca ggttaccttg agcaacaccg acaacacgtc ctgaggcgca tctgcagcaa	1440
gaggaaagga cgtcgccatc gtcttcacat ccgccgactc gggtaaggc tacatcacgc	1500
tggagggcaa cgcgggcgat cgcaacaacc tggatccgtg gcacaacggc aatgccctgg	1560
tccaggcggg ggccgggtgcc aacagcaacg tcattgttgt tgtccactcc gttggcgcca	1620
tcattctgga gcagattctt gctcttcgc aggtcaaggc cgttgtctgg gcgggtcttc	1680
cttctcagga gagcggcaat gcgctcgtcg acgtgctgtg gggagatgtc agcccttctg	1740
gcaagctggt gtacaccatt gcgaagagcc ccaatgacta taacactcgc atcgtttccg	1800
gcggcagtga cagcttcagc gagggactgt tcatcgacta taagcacttc gacgacgcca	1860
atatcacgcc gcggtacgag ttccgctatg gactgtgtaa gtttgctaac ctgaacaatc	1920
tattagacag gttgactgac ggatgactgt ggaatgatag cttacaccaa gttcaactac	1980
tcacgcctct ccgtcttgtc gaccgccaag tctggctctg cgactggggc cgttgtgccg	2040
ggaggccccg gtgatctgtt ccagaatgtc gcgacagtca ccgttgacat cgaaaactct	2100
ggccaagtga ctggtgccga ggtagcccag ctgtacatca cctacccatc ttcagcacc	2160
aggaccctc cgaagcagct gcgaggcttt gccaaagtga acctcacgcc tggtcagagc	2220
ggaacagcaa cgttcaacat ccgacgacga gatctcagct actgggacac ggcttcgcag	2280
aaatgggtgg tgccgtcggg gtcgtttggc atcagcgtgg gagcgagcag ccgggatatc	2340

aggctgacga gcactctgtc ggtagcgtag

2370

<210> 62

<211> 744

<212> PRT

<213> Trichoderma reesei

<400> 62

Met Arg Tyr Arg Thr Ala Ala Ala Leu Ala Leu Ala Thr Gly Pro Phe

1 5 10 15

Ala Arg Ala Asp Ser His Ser Thr Ser Gly Ala Ser Ala Glu Ala Val

20 25 30

Val Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala Lys

35 40 45

Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser

50 55 60

Gly Val Gly Trp Asn Gly Gly Pro Cys Val Gly Asn Thr Ser Pro Ala

65 70 75 80

Ser Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu Gly

85 90 95

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

100 105 110

Ala Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe Ile

115 120 125

Gly Glu Glu Val Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro Val

130 135 140

Ala Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu Gly

145 150 155 160

Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr Ile

165 170 175

Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr Ile

180 185 190

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

195 200 205
 Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala Val
 210 215 220
 Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn Thr
 225 230 235 240
 Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp

 245 250 255
 Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His
 260 265 270
 Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly
 275 280 285
 Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn
 290 295 300
 Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val

 305 310 315 320
 Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly
 325 330 335
 Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr
 340 345 350
 Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp
 355 360 365
 Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly

 370 375 380
 Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn
 385 390 395 400
 Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly
 405 410 415
 Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr
 420 425 430
 Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn

 435 440 445

Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val
 450 455 460
 Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn
 465 470 475 480
 Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu
 485 490 495
 Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His
 500 505 510
 Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val
 515 520 525
 Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala
 530 535 540
 Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val
 545 550 555 560
 Tyr Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser
 565 570 575
 Gly Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His
 580 585 590
 Phe Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu
 595 600 605
 Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala
 610 615 620
 Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp
 625 630 635 640
 Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly
 645 650 655
 Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser
 660 665 670
 Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu
 675 680 685
 Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg

690	695	700	
Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro			
705	710	715	720
Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg			
	725	730	735
Leu Thr Ser Thr Leu Ser Val Ala			
	740		
<210> 63			
<211> 2625			
<212> DNA			
<213> Trichoderma reesei			
<400> 63			
atgaagacgt tgtcagtgtt tgctgccgcc cttttggcgg ccgtagctga ggccaatccc			60
taccgcctc ctactccaa ccaggcgtac tcgcctcctt tctacccttc gccatggatg			120
gaccccagtg ctccaggctg ggagcaagcc tatgcccaag ctaaggagtt cgtctcgggc			180
ttgactctct tggagaaggt caacctcacc accggtgttg gctggatggg tgagaagtgc			240
gttggaacg ttggtacctg gcctcgcttg ggcatgcgaa gtctttgcat gcaggacggc			300
cccctgggtc tccgattcaa cacgtacaac agcgctttca gcgttggctt gacggccgcc			360
gccagctgga gccgacacct ttgggttgac cgcggtaccg ctctgggctc cgaggcaaag			420
ggcaagggtg tcgatgttct tctcggaccc gtggctggcc ctctcggtcg caaceccaac			480
ggaggccgta acgtcgaggg ttctggctcg gatccctatc tggcgggttt ggctctggcc			540
gataccgtga ccggaatcca gaacgcgggc accatcgctt gtccaagca ctctctctc			600
aacgagcagg agcatttccg ccaggctggc gaagctaacg gttacggata cccatcacc			660
gaggctctgt ctccaacgt tgatgacaag acgattcacg aggtgtacgg ctggcccttc			720
caggatgctg tcaaggctgg tgctgggtcc ttcattgtgt cgtacaacca ggtcaacaac			780
tcgtacgctt gccaaaactc caagctcatc aacggcttgc tcaaggagga gtacggtttc			840
caaggctttg tcatgagcga ctggcaggcc cagcacacgg gtgtcgcgtc tgctgttgcc			900
ggtctcgata tgaccatgcc tggtagacac gccttcaaca ccggcgcata ctactttgga			960
agcaacctga cgcttgctgt tctcaacggc accgtccccg agtggcgcac tgacgacatg			1020
gtgatgcgta tcatggctcc cttcttcaag gtgggcaaga cggttgacag cctcattgac			1080
accaactttg attcttggac caatggcgag tacggctacg ttcaggccgc cgtcaatgag			1140
aactgggaga aggtcaacta cggcgtcgat gtccgcgcca accatgcgaa ccacatccgc			1200

gaggttggcg ccaagggaac tgtcatcttc aagaacaacg gcatcctgcc ccttaagaag 1260
 cccaagtcc tgaccgtcat tggtaggat gctggcggca accctgccgg cccaacggc 1320

tgcggtgacc gcggctgtga cgacggcact cttgccatgg agtggggatc tggactacc 1380
 aacttccct acctcgtcac ccccgacgcg gccctgcaga gccaggctct ccaggacggc 1440
 acccgctacg agagcatcct gtccaactac gccatctcgc agaccaggc gctcgtcagc 1500
 cagccgatg ccattgccat tgtctttgcc aactcggata gcggcgaggg ctacatcaac 1560
 gtcgatggca acgagggcga ccgcaagaac ctgacgctgt ggaagaacgg cgacgatctg 1620
 atcaagactg ttgctgctgt caacccaag acgattgtcg tcattccactc gaccggcccc 1680
 gtgattctca aggactacgc caaccacccc aacatctctg ccattctgtg ggccggtgct 1740

cctggccagg agtctggcaa ctgctggtc gacattctgt acggcaagca gagcccgggc 1800
 cgactccct tcacctgggg ccgctcgtg gagagctacg gagtagtgt tatgaccacg 1860
 cccaacaacg gcaacggcgc tcccaggat aacttcaacg agggcgctt catcgactac 1920
 cgctactttg acaaggtggc tcccggcaag cctcgcagct cggacaaggc tcccacgtac 1980
 gagtttggct tcggactgtc gtggtcgacg ttcaagttct ccaacctcca catccagaag 2040
 aacaatgtcg gcccctagc cccgccaac ggcaagacga ttgcggctcc ctctctgggc 2100
 agcttcagca agaaccttaa ggactatggc ttcccaaga acgttcgccg catcaaggag 2160

tttatctacc cctacctgag caccactacc tctggcaagg aggcgtcggg tgacgtcac 2220
 tacggccaga ctgcgaagga gtctctccc gccggtgccc tggacggcag ccctcagcct 2280
 cgctctgcgg cctctggcga acccgcggc aaccgccagc tgtacgacat tctctacacc 2340
 gtgacggcca ccattacaa cacgggctcg gtcatggacg acgccgttcc ccagctgtac 2400
 ctgagccacg gcggtccaa cgagccgcc aaggtgtctg gtggttcga ccgcatcgag 2460
 cgcatgtctc ccggccagag cgtcacgttc aaggcagacc tgacgcgccg tgacctgtcc 2520
 aactgggaca cgaagaagca gcagtgggtc attaccgact accccaagac tgtgtacgtg 2580

ggcagctcct cgcgcgacct gccgtgagc gccgcctgc catga 2625

<210> 64

<211> 874

<212> PRT

<213> Trichoderma reesei

<400> 64

Met Lys Thr Leu Ser Val Phe Ala Ala Ala Leu Leu Ala Ala Val Ala

1 5 10 15

Glu Ala Asn Pro Tyr Pro Pro Pro His Ser Asn Gln Ala Tyr Ser Pro
 20 25 30
 Pro Phe Tyr Pro Ser Pro Trp Met Asp Pro Ser Ala Pro Gly Trp Glu
 35 40 45

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

 Phe Ser Val Gly Leu Thr Ala Ala Ala Ser Trp Ser Arg His Leu Trp
 115 120 125
 Val Asp Arg Gly Thr Ala Leu Gly Ser Glu Ala Lys Gly Lys Gly Val
 130 135 140
 Asp Val Leu Leu Gly Pro Val Ala Gly Pro Leu Gly Arg Asn Pro Asn
 145 150 155 160
 Gly Gly Arg Asn Val Glu Gly Phe Gly Ser Asp Pro Tyr Leu Ala Gly
 165 170 175

 Leu Ala Leu Ala Asp Thr Val Thr Gly Ile Gln Asn Ala Gly Thr Ile
 180 185 190
 Ala Cys Ala Lys His Phe Leu Leu Asn Glu Gln Glu His Phe Arg Gln
 195 200 205
 Val Gly Glu Ala Asn Gly Tyr Gly Tyr Pro Ile Thr Glu Ala Leu Ser
 210 215 220
 Ser Asn Val Asp Asp Lys Thr Ile His Glu Val Tyr Gly Trp Pro Phe
 225 230 235 240

 Gln Asp Ala Val Lys Ala Gly Val Gly Ser Phe Met Cys Ser Tyr Asn
 245 250 255
 Gln Val Asn Asn Ser Tyr Ala Cys Gln Asn Ser Lys Leu Ile Asn Gly

260 265 270
 Leu Leu Lys Glu Glu Tyr Gly Phe Gln Gly Phe Val Met Ser Asp Trp
 275 280 285
 Gln Ala Gln His Thr Gly Val Ala Ser Ala Val Ala Gly Leu Asp Met
 290 295 300

 Thr Met Pro Gly Asp Thr Ala Phe Asn Thr Gly Ala Ser Tyr Phe Gly
 305 310 315 320
 Ser Asn Leu Thr Leu Ala Val Leu Asn Gly Thr Val Pro Glu Trp Arg
 325 330 335
 Ile Asp Asp Met Val Met Arg Ile Met Ala Pro Phe Phe Lys Val Gly
 340 345 350
 Lys Thr Val Asp Ser Leu Ile Asp Thr Asn Phe Asp Ser Trp Thr Asn
 355 360 365

 Gly Glu Tyr Gly Tyr Val Gln Ala Ala Val Asn Glu Asn Trp Glu Lys
 370 375 380
 Val Asn Tyr Gly Val Asp Val Arg Ala Asn His Ala Asn His Ile Arg
 385 390 395 400
 Glu Val Gly Ala Lys Gly Thr Val Ile Phe Lys Asn Asn Gly Ile Leu
 405 410 415
 Pro Leu Lys Lys Pro Lys Phe Leu Thr Val Ile Gly Glu Asp Ala Gly
 420 425 430

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

 Ala Leu Val Ser Gln Pro Asp Ala Ile Ala Ile Val Phe Ala Asn Ser
 500 505 510

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

Val Ile Leu Lys Asp Tyr Ala Asn His Pro Asn Ile Ser Ala Ile Leu
565 570 575
Trp Ala Gly Ala Pro Gly Gln Glu Ser Gly Asn Ser Leu Val Asp Ile
580 585 590
Leu Tyr Gly Lys Gln Ser Pro Gly Arg Thr Pro Phe Thr Trp Gly Pro
595 600 605
Ser Leu Glu Ser Tyr Gly Val Ser Val Met Thr Thr Pro Asn Asn Gly
610 615 620

Asn Gly Ala Pro Gln Asp Asn Phe Asn Glu Gly Ala Phe Ile Asp Tyr
625 630 635 640
Arg Tyr Phe Asp Lys Val Ala Pro Gly Lys Pro Arg Ser Ser Asp Lys
645 650 655
Ala Pro Thr Tyr Glu Phe Gly Phe Gly Leu Ser Trp Ser Thr Phe Lys
660 665 670
Phe Ser Asn Leu His Ile Gln Lys Asn Asn Val Gly Pro Met Ser Pro
675 680 685

Pro Asn Gly Lys Thr Ile Ala Ala Pro Ser Leu Gly Ser Phe Ser Lys
690 695 700
Asn Leu Lys Asp Tyr Gly Phe Pro Lys Asn Val Arg Arg Ile Lys Glu
705 710 715 720
Phe Ile Tyr Pro Tyr Leu Ser Thr Thr Thr Ser Gly Lys Glu Ala Ser
725 730 735
Gly Asp Ala His Tyr Gly Gln Thr Ala Lys Glu Phe Leu Pro Ala Gly
740 745 750

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

755 760 765
 Gly Gly Asn Arg Gln Leu Tyr Asp Ile Leu Tyr Thr Val Thr Ala Thr
 770 775 780
 Ile Thr Asn Thr Gly Ser Val Met Asp Asp Ala Val Pro Gln Leu Tyr
 785 790 795 800
 Leu Ser His Gly Gly Pro Asn Glu Pro Pro Lys Val Leu Arg Gly Phe
 805 810 815

Asp Arg Ile Glu Arg Ile Ala Pro Gly Gln Ser Val Thr Phe Lys Ala
 820 825 830
 Asp Leu Thr Arg Arg Asp Leu Ser Asn Trp Asp Thr Lys Lys Gln Gln
 835 840 845
 Trp Val Ile Thr Asp Tyr Pro Lys Thr Val Tyr Val Gly Ser Ser Ser
 850 855 860
 Arg Asp Leu Pro Leu Ser Ala Arg Leu Pro
 865 870

<210> 65

<211> 2577

<212>

> DNA

<213> Artificial Sequence

<220><223> synthetic codon optimized GH3 family beta-glucosidase from
 Talaromyces emersonii

<400> 65

atgcgcaacg gcctcctcaa ggtcgccgcc ttagccgtg ccagcgccgt caacggcgag	60
aacctgcct acagcccc cttctaccc agccctggg ccaacggcca gggcgactgg	120
gccgaggcct accagaagge cgtccagttc gtcagccagc tcacctcgc cgagaaggte	180
aacctacca ccggcaccgg ctgggagcag gaccgtgcg tcggccaggt cggcagcatc	240
ccccgcttag gcttccccg cctctgcatg caggacagcc ccctcggcgt ccgcgacacc	300
gactacaaca gcgccttccc tgccggcggt aacgtcgccg ccacctggga ccgcaactta	360
gcctaccgca gaggcgtcgc catgggcgag gaacaccgcg gcaagggcgt cgacgtccag	420
ttaggccccg tcgcccgcct cttaggccgc tctctgatg ccggccgcaa ctgggagggc	480
ttcgcccccg acccgtcct caccggcaac atgatggcca gcaccatcca ggcatccag	540
gatgctggcg tcattgcctg cgccaagcac ttcacctct acgagcagga acacttcgc	600

caggcgcccc aggcaggcta cgacatcagc gacagcatca gcgccaacgc cgacgacaag	660
accatgcacg agttatacct ctggcccttc gccgatgccg tccgcgccgg tgtcggcagc	720
gtcatgtgca gctacaacca ggtcaacaac agctacgcct gcagcaacag ctacaccatg	780
aacaagctcc tcaagagcga gttaggcttc cagggccttcg tcatgaccga ctggggcggc	840
caccacagcg gcgtcggctc tgccctcgcc ggccctcgaca tgagcatgcc cggcgacatt	900
gccttcgaca ggggcacgtc tttctggggc accaacctca ccgttgccgt cctcaacggc	960
tccatccccg agtggcgcggt cgacgacatg gccgtccgca tcatgagcgc ctactacaag	1020
gtcggccgcg accgctacag cgtcccatc aacttcgaca gctggaccct cgacacctac	1080
ggccccgagc actacgccgt cggccagggc cagaccaaga tcaacgagca cgtcgacgtc	1140
cgcggaacc acgccagat catccacgag atcggcgccg cctccgccgt cctcctcaag	1200
aacaagggcg gcctccccct cactggcacc gagcgcttcg tcggtgtctt tggcaaggat	1260
gctggcagca acccctgggg cgtcaacggc tgcagcgacc gcggctgcga caacggcacc	1320
ctcgccatgg gctggggcag cggcaccgcc aactttccct acctcgtcac ccccagcag	1380
gccatccagc gcgaggtcct cagccgcaac ggcacctca ccggcatcac cgacaacggc	1440
gccttagccg agatggccgc tgccgcctct caggccgaca cctgcctcgt ctttgccaac	1500
gccgactccg gcgagggcta catcaccgtc gatggcaacg agggcgaccg caagaacctc	1560
accctctggc agggcgccga ccaggctatc cacaacgtca gcgccaactg caacaacacc	1620
gtcgtcgtct tacacaccgt cggccccgtc ctcatcgacg actggtacga ccacccaac	1680
gtcacgcca tcctctgggc cggtttacc ggtcaggaaa gcggcaacag cctcgtcgac	1740
gtcctctacg gccgcgtcaa ccccggaag acccccttca cctggggcag agcccgcgac	1800
gactatggcg cccctctcat cgtcaagcct aacaacggca agggcgcccc ccagcaggac	1860
ttcacggagg gcattctcat cgactaccgc cgttcgaca agtacaacat caccceccatc	1920
tacgagttcg gcttcggcct cagctacacc accttcgagt tcagccagtt aaacgtccag	1980
cccatcaacg cccctcccta cacccecgcc agcggttita cgaaggccgc ccagagcttc	2040
ggccagccct ccaatgccag cgacaacctc taccctagcg acatcgagcg cgtccccctc	2100
tacatctacc cctggctcaa cagcaccgac ctcaaggcca gcgccaacga ccccgactac	2160
ggcctcccca ccgagaagta cgtccccccc aacgccacca acggcgaccc ccagccatt	2220
gaccctgccg gcggtgcccc tggcggcaac cccagcctct acgagcccgt cggccgcgtc	2280
accaccatca tcaccaacac cggcaaggtc accggcgacg aggtcccca gctctatgtc	2340

agcttaggcg gccctgacga cgcccccaag gtccctccgcg gcttcgaccg catcacccctc 2400

gcccctggcc agcagtacct ctggaccacc accctcactc gccgcgacat cagcaactgg 2460

gaccccgta cccagaactg ggctcgtcacc aactacacca agaccatcta cgtcggcaac 2520

agcagccgca acctccccct ccaggccccc ctcaagccct accccggcat ctgatga 2577

<210> 66

<211> 857

<212> PRT

<213> Talaromyces emersonii

<400> 66

Met Arg Asn Gly Leu Leu Lys Val Ala Ala Leu Ala Ala Ala Ser Ala

1 5 10 15

Val Asn Gly Glu Asn Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser Pro

20 25 30

Trp Ala Asn Gly Gln Gly Asp Trp Ala Glu Ala Tyr Gln Lys Ala Val

35 40 45

Gln Phe Val Ser Gln Leu Thr Leu Ala Glu Lys Val Asn Leu Thr Thr

50 55 60

Gly Thr Gly Trp Glu Gln Asp Arg Cys Val Gly Gln Val Gly Ser Ile

65 70 75 80

Pro Arg Leu Gly Phe Pro Gly Leu Cys Met Gln Asp Ser Pro Leu Gly

85 90 95

Val Arg Asp Thr Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn Val

100 105 110

Ala Ala Thr Trp Asp Arg Asn Leu Ala Tyr Arg Arg Gly Val Ala Met

115 120 125

Gly Glu Glu His Arg Gly Lys Gly Val Asp Val Gln Leu Gly Pro Val

130 135 140

Ala Gly Pro Leu Gly Arg Ser Pro Asp Ala Gly Arg Asn Trp Glu Gly

145 150 155 160

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

165 170 175

Gln Gly Ile Gln Asp Ala Gly Val Ile Ala Cys Ala Lys His Phe Ile
180 185 190

Leu Tyr Glu Gln Glu His Phe Arg Gln Gly Ala Gln Asp Gly Tyr Asp
195 200 205

Ile Ser Asp Ser Ile Ser Ala Asn Ala Asp Asp Lys Thr Met His Glu
210 215 220

Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser
225 230 235 240

Val Met Cys Ser Tyr Asn Gln Val Asn Asn Ser Tyr Ala Cys Ser Asn
245 250 255

Ser Tyr Thr Met Asn Lys Leu Leu Lys Ser Glu Leu Gly Phe Gln Gly
260 265 270

Phe Val Met Thr Asp Trp Gly Gly His His Ser Gly Val Gly Ser Ala
275 280 285

Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Ile Ala Phe Asp Ser
290 295 300

Gly Thr Ser Phe Trp Gly Thr Asn Leu Thr Val Ala Val Leu Asn Gly
305 310 315 320

Ser Ile Pro Glu Trp Arg Val Asp Asp Met Ala Val Arg Ile Met Ser
325 330 335

Ala Tyr Tyr Lys Val Gly Arg Asp Arg Tyr Ser Val Pro Ile Asn Phe
340 345 350

Asp Ser Trp Thr Leu Asp Thr Tyr Gly Pro Glu His Tyr Ala Val Gly
355 360 365

Gln Gly Gln Thr Lys Ile Asn Glu His Val Asp Val Arg Gly Asn His
370 375 380

Ala Glu Ile Ile His Glu Ile Gly Ala Ala Ser Ala Val Leu Leu Lys
385 390 395 400

Asn Lys Gly Gly Leu Pro Leu Thr Gly Thr Glu Arg Phe Val Gly Val
405 410 415

Phe Gly Lys Asp Ala Gly Ser Asn Pro Trp Gly Val Asn Gly Cys Ser

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

 Ala Leu Ala Glu Met Ala Ala Ala Ala Ser Gln Ala Asp Thr Cys Leu
 485 490 495
 Val Phe Ala Asn Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Asp Gly
 500 505 510
 Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp Gln Gly Ala Asp Gln
 515 520 525
 Val Ile His Asn Val Ser Ala Asn Cys Asn Asn Thr Val Val Val Leu
 530 535 540

 His Thr Val Gly Pro Val Leu Ile Asp Asp Trp Tyr Asp His Pro Asn
 545 550 555 560
 Val Thr Ala Ile Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn
 565 570 575
 Ser Leu Val Asp Val Leu Tyr Gly Arg Val Asn Pro Gly Lys Thr Pro
 580 585 590
 Phe Thr Trp Gly Arg Ala Arg Asp Asp Tyr Gly Ala Pro Leu Ile Val
 595 600 605

 Lys Pro Asn Asn Gly Lys Gly Ala Pro Gln Gln Asp Phe Thr Glu Gly
 610 615 620
 Ile Phe Ile Asp Tyr Arg Arg Phe Asp Lys Tyr Asn Ile Thr Pro Ile
 625 630 635 640
 Tyr Glu Phe Gly Phe Gly Leu Ser Tyr Thr Thr Phe Glu Phe Ser Gln
 645 650 655
 Leu Asn Val Gln Pro Ile Asn Ala Pro Pro Tyr Thr Pro Ala Ser Gly
 660 665 670

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

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

Ala Pro Gly Gln Gln Tyr Leu Trp Thr Thr Thr Leu Thr Arg Arg Asp
805 810 815
Ile Ser Asn Trp Asp Pro Val Thr Gln Asn Trp Val Val Thr Asn Tyr
820 825 830
Thr Lys Thr Ile Tyr Val Gly Asn Ser Ser Arg Asn Leu Pro Leu Gln
835 840 845
Ala Pro Leu Lys Pro Tyr Pro Gly Ile
850 855

<210> 67

<211> 2586

<212> DNA

<213> Aspergillus niger

<400> 67

atgcgttca ccagcatcga ggccgtcgcc ctaccgccg tcagcctcgc cagcgccgac	60
gagttagcct acagccccc ctactacccc agcccctggg ccaacggcca gggcgactgg	120
gccgaggcct accagcgcgc cgtcgacatc gtcagccaga tgaccctcgc cgagaaggtc	180

aacctcacca ccggcaccgg ctgggagtta gagttatgcg tcggccagac tggtagcgctc	240
ccccgcctcg gcatccccgg catgtgcgcc caggacagcc ccctcggcgt ccgcgacagc	300
gactacaaca gcgccttccc tggcggcgctc aacgtcgccg ccacctggga caagaacctc	360
gcctacctcc gcggccaggc catgggccag gaattcagcg acaagggcgc cgacatccag	420
ttaggccccg ctgccggccc ttagggccgc tctcccgacg gcggcagaaa ctgggagggc	480
ttcagccccg accccgctct cagcggcgctc ctcttcgccg agactatcaa gggcatccag	540
gatgctggcg tcgtcgccac cgccaagcac tacattgcct acgagcagga acacttcgc	600
caggcccccg aggcccaggc ctacggcttc aacatcacgc agagcggcag cgccaacctc	660
gacgacaaga ccatgcagga gttatacttc tggcccttcg ccgacgccat tagagctggc	720
gctggtgctg tcatgtgcag ctacaaccag atcaacaaca gctacggctg ccagaacagc	780
tacacctca acaagctcct caaggccgag ttaggcttcc agggcttcgt catgtccgac	840
tgggcgccc accacgccgg cgtcagcggc gccttagccg gcctcgacat gagcatgccc	900
ggcgacgtcg actacgacag cggcaccagc tactggggca ccaacctcac catcagcgtc	960
ctcaacggca ccgtccccc gtggcgcgctc gacgacatgg ccgtccgcat catggccgcc	1020
tactacaagg tcggcccgca ccgcctctgg accccccca acttcagcag ctggaccgc	1080
gacgagtacg gcttcaagta ctactacgtc agcggaggcc cctatgagaa ggtcaaccag	1140
ttcgtcaacg tcacgcgcaa ccacagcgag ttaatccgc gcctcggcgc cgacagcacc	1200
gtcctcctca agaacgacgg cgccctcccc ctaccggca aggaacgcct cgtcgccctc	1260
atcggcgagg acgccggcag caacccttac ggcgccaacg gctgcagcga ccgcggctgc	1320
gacaacggca ccctcgccat gggctggggc agcggcaccg ccaacttccc ttacctgctc	1380
acccccgagc aggccatcag caacgaggtc ctcaagaaca agaacggcgt ctttaccgcc	1440
accgacaact ggcccatcga ccagatcgag gccttagcca agaccgcctc tgtcagcctc	1500
gtctttgtca acgccgacag cggcgagggc tacatcaacg tcgacggcaa cctcggcgac	1560
cggcgcaacc tcacctctg gcgcaacggc gacaacgtca tcaaggccgc cgccagcaac	1620
tgcaacaaca ccatgctcat catccacagc gtcggccccg tctctgtcaa cgagtggtag	1680
gacaaccccc acgtcacgc catcctctgg ggcggttac ccggccagga aagcggaac	1740
agcctcgccg acgtcctcta ggcccgctc aacctggcg ccaagagccc ctacacctgg	1800
ggcaagacct gcgaggccta tcaggactac ctctacaccg agcccaaca cggcaacggc	1860
gccccccagg aagatttcgt cgagggcgctc ttatcgact accgcggctt tgacaagcgc	1920
aacgagactc ccatctacga gttcggctac ggccctcagct acaccacctt caactacagc	1980
aacctccagg tcgaggtcct cagcgccct gcctacgagc ccgccagcgg cgagactgag	2040

gccgccccca ccttcggcga ggtcggcaac gccagcgact attataccc cgacggcctc 2100

cagcgcatca ccaagttcat ctacccttgg ctcaacagca ccgacctga ggccagcagc 2160

ggcgagcctt cttacggcca ggacgcctcc gactacctcc ccgagggtgc caccgacggc 2220

agcgctcagc ccatcttacc tgccggtggc ggtgctggcg gcaaccccag actctacgac 2280

gagctgatcc gcgtcagcgt caccatcaag aacaccggca aggtcgctgg tgacgaggtc 2340

ccccagctct acgtcagctt aggcgggcct aacagagcca agatcgctct ccgccagttc 2400

gagcgcatca cctccagcc cagcaaggaa actcagtga gcaccacct cactcgccgc 2460

gacctcgcca actggaacgt cgagactcag gactgggaga tcaccagcta cccaagatg 2520

gtctttgccg gcagcagcag ccgcaagctc cccctccgcg ccagcctccc caccgtccac 2580

tgatga 2586

<210> 68

<211> 860

<212> PRT

<213> *Aspergillus niger*

<400> 68

Met Arg Phe Thr Ser Ile Glu Ala Val Ala Leu Thr Ala Val Ser Leu

1 5 10 15

Ala Ser Ala Asp Glu Leu Ala Tyr Ser Pro Pro Tyr Tyr Pro Ser Pro

20 25 30

Trp Ala Asn Gly Gln Gly Asp Trp Ala Glu Ala Tyr Gln Arg Ala Val

35 40 45

Asp Ile Val Ser Gln Met Thr Leu Ala Glu Lys Val Asn Leu Thr Thr

50 55 60

Gly Thr Gly Trp Glu Leu Glu Leu Cys Val Gly Gln Thr Gly Gly Val

65 70 75 80

Pro Arg Leu Gly Ile Pro Gly Met Cys Ala Gln Asp Ser Pro Leu Gly

85 90 95

Val Arg Asp Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn Val

100 105 110

Ala Ala Thr Trp Asp Lys Asn Leu Ala Tyr Leu Arg Gly Gln Ala Met

115 120 125

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

370 375 380
 Gln Arg Asn His Ser Glu Leu Ile Arg Arg Ile Gly Ala Asp Ser Thr
 385 390 395 400
 Val Leu Leu Lys Asn Asp Gly Ala Leu Pro Leu Thr Gly Lys Glu Arg
 405 410 415
 Leu Val Ala Leu Ile Gly Glu Asp Ala Gly Ser Asn Pro Tyr Gly Ala

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

 485 490 495
 Ser Val Ser Leu Val Phe Val Asn Ala Asp Ser Gly Glu Gly Tyr Ile
 500 505 510
 Asn Val Asp Gly Asn Leu Gly Asp Arg Arg Asn Leu Thr Leu Trp Arg
 515 520 525
 Asn Gly Asp Asn Val Ile Lys Ala Ala Ala Ser Asn Cys Asn Asn Thr
 530 535 540
 Ile Val Ile Ile His Ser Val Gly Pro Val Leu Val Asn Glu Trp Tyr

 545 550 555 560
 Asp Asn Pro Asn Val Thr Ala Ile Leu Trp Gly Gly Leu Pro Gly Gln
 565 570 575
 Glu Ser Gly Asn Ser Leu Ala Asp Val Leu Tyr Gly Arg Val Asn Pro
 580 585 590
 Gly Ala Lys Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ala Tyr Gln
 595 600 605
 Asp Tyr Leu Tyr Thr Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln Glu

 610 615 620

Asp Phe Val Glu Gly Val Phe Ile Asp Tyr Arg Gly Phe Asp Lys Arg
 625 630 635 640
 Asn Glu Thr Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr
 645 650 655
 Phe Asn Tyr Ser Asn Leu Gln Val Glu Val Leu Ser Ala Pro Ala Tyr
 660 665 670
 Glu Pro Ala Ser Gly Glu Thr Glu Ala Ala Pro Thr Phe Gly Glu Val

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

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

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

<210> 69

<211> 3203

<212> DNA

<213> *Fusarium oxysporum*

<400> 69

atgaagctga actgggtcgc cgcagccctc tctatagggtg ctgctggcac tgatgggtgca	60
gttgctcttg cttctgaagt tccaggcact ttggctgggtg taaaggctcg tttttttacc	120
atctctcac ctaatctcag ccttgttgcc atategccct tattegctcg gacgctacgc	180
accaaactgc gatcatttcc tcccttgtag ccttgttttc ttttttcgat ctccctccg	240
caatgcagg cacccttagc ctacacaaaa acccccagaga cagtctcatt gaggttgtcg	300
acatcaagtt gcttctcaag tgtgcatctg cgtggctgtc tacttctgcc tctagaccac	360
caaatctggg cgcaattgat cgctcaaacc ttgttcgaat aagcctttta ttcgagacgt	420
ccaattttta cagagaatgt acctttcaat aataccgacg ttatgcgcgg cgggtggctgc	480
tgtgatgggt gtgatcaga atactgacgc tcaaaagggt gtcacgagag atacactcgc	540
acactcacct cctcactatc cttcaccatg gatggatcct aatgccattg gctgggagga	600
agcttacgcc aaagcaaaga actttgtgtc ccagctcact ctctcgaaa aggtcaactt	660
gaccactggt gtgggtaag tagctccttg cgaacagtgc atctcggtct ccttgactaa	720
cgactctctc aggtggcaag gcgaacgtg ttaggaaac gtgggatcaa ttcctcgtct	780
tggatgcga ggctttgtc ttcaggatgg tctctttgga attcgtctgt ccgattacaa	840
cagtgccttt cccgctggca ccacagctgg tgcttcttgg agcaagtctc tctgggatga	900
gaggggtctt ctgatgggaa ctgagttcaa gggaagggt atcgatatcg ctcttggccc	960
tgctactggt cctcttggcc gcactgctgc tggtagcga aactgggagg gctttaccgt	1020
tgatccttat atggctggcc atgcatggc cgaggccgtc aaggccatcc aagacgcagg	1080
tgtcattgct tgtgctaagc attacatcgc aaacgagcaa ggtaagccaa ttggacggtt	1140
tgggaaatcg acagagaact gaccccttg tagagcactt ccgacagagt ggccaggtcc	1200
agtcccga gtaacatc tccagctctc tctctccaa cctggacgac aagactttgc	1260
acgagctcta cgctggccc tttgtgatg ccgtccgcgc tggcgtcggg tcagtcattg	1320
gctcttaca tcagatcaac aactcgtac gtggcagaa ctccaagctc ctcaacggtg	1380
tcctcaagga cgagatgggt ttccagggtc tcgtcatgag cgattgggag gccagcaca	1440
ccggtgctgc ttctgccgtc gctggtcttg atatgagcat gcctgggtgac accgcgttcg	1500
acagtggata tagcttctgg ggtggaaacc tgactcttgc tgtcatcaac ggaactgttc	1560
ccgctggcg agttgatgac atggctctgc gaatcatgac ggcttcttc aaggttggaa	1620
agacggtaga ggacctcccc gacatcaact tctctcctg gaccgcgac accttcgggt	1680

tcgtccaaac atttgcitcaa gagaaccgcg aacaagtcaa ctttggagtt aacgtccagc 1740

acgaccacaa gaaccacatc cgtgagtctg ccgccaaggg aagcgtcatc ctcaagaaca 1800

ccggctccct tcccctcaac aatcccaagt tctcgtctgt cattgggtgag gacgccggtc 1860

ccaacctgc tggaccaat ggttgcggcg accgtggttg cgacaatggt accctggcta 1920

tggcttgggg ctcggaact tctcaattcc ctacttgat cacacccgac caaggtctcc 1980

agaaccgagc tgccaagac ggaactcgat atgagagcat cttgaccaac aacgaatggg 2040

cccagacaca ggctcttgtc agccaacca acgtgaccgc tatcgttttt gccaacgccg 2100

actctggtga gggttacatt gaagtcgacg gaaacttcgg tgatcgcaag aacctcacc 2160

tctggcaaca gggagacgag ctcatcaaga acgtctctgc catctgcccc aacaccattg 2220

tcgttctgca taccgtcggc cctgtcctgc tcgccgacta cgagaagaac cccaacatca 2280

ccgccatcgt ctgggctggt ctcccgccg aagagtctgg caatgccatc gctgatctcc 2340

tctacggcaa ggtaagccct ggccgatctc ccttacttg gggccgcacc cgtgagagct 2400

acggtaccga ggttctttat gaggcgaaca acggccgtgg cgctcctcag gatgacttct 2460

cggagggtgt ctcatcgac taccgtcact ttgatcgacg atctcccagc accgatggca 2520

agagcgctcc caacaacacc gctgtcctc tctacgagtt cggatcatggt ctgtcttga 2580

ctacctttga gtattcagac ctcaacatcc agaagaacgt taactccacc tactctctc 2640

ctgttggtca gaccattcct gccccaacct ttggcaactt cagcaagaac ctcaacgact 2700

acgtgttccc taagggtgtc cgatacatct acaagttcat ctacccttc ctgaacactt 2760

cctcatccgc cagcgaggca tctaacgacg gcggccagtt tggtaagact gccgaagagt 2820

tctacctcc aaacgcctc aacggctcag ccagcctcg tcttccctct tctggtgccc 2880

caggcggtaa cctcaattg tgggatatcc tgtacaccgt cacagccaca atcaccaaca 2940

caggcaacgc cacctccgac gagattcccc agctgtatgt cagcctcggg ggcgagaacg 3000

aaccggttcg tgtcctccgc ggtttcgacc gtatcgagaa cattgctccc ggccagagcg 3060

ccatcttcaa cgtcaattg acccgtcgcg atctgagcaa ctgggatgtg gatgcccgaga 3120

actgggttat caccgacat ccaaagacgg tgtgggttgg aagtagttct cgcaagctgc 3180

ctctcagcgc caagttggaa taa 3203

<210> 70

<211> 899

<212> PRT

<213> Fusarium oxysporum

<400> 70

Met Lys Leu Asn Trp Val Ala Ala Ala Leu Ser Ile Gly Ala Ala Gly
1 5 10 15

Thr Asp Gly Ala Val Ala Leu Ala Ser Glu Val Pro Gly Thr Leu Ala
20 25 30

Gly Val Lys Asn Thr Asp Ala Gln Lys Val Val Thr Arg Asp Thr Leu
35 40 45

Ala His Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp Pro Asn Ala
50 55 60

Ile Gly Trp Glu Glu Ala Tyr Ala Lys Ala Lys Asn Phe Val Ser Gln
65 70 75 80

Leu Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gln
85 90 95

Gly Glu Arg Cys Val Gly Asn Val Gly Ser Ile Pro Arg Leu Gly Met
100 105 110

Arg Gly Leu Cys Leu Gln Asp Gly Pro Leu Gly Ile Arg Leu Ser Asp
115 120 125

Tyr Asn Ser Ala Phe Pro Ala Gly Thr Thr Ala Gly Ala Ser Trp Ser
130 135 140

Lys Ser Leu Trp Tyr Glu Arg Gly Leu Leu Met Gly Thr Glu Phe Lys
145 150 155 160

Gly Lys Gly Ile Asp Ile Ala Leu Gly Pro Ala Thr Gly Pro Leu Gly
165 170 175

Arg Thr Ala Ala Gly Gly Arg Asn Trp Glu Gly Phe Thr Val Asp Pro
180 185 190

Tyr Met Ala Gly His Ala Met Ala Glu Ala Val Lys Gly Ile Gln Asp
195 200 205

Ala Gly Val Ile Ala Cys Ala Lys His Tyr Ile Ala Asn Glu Gln Glu
210 215 220

His Phe Arg Gln Ser Gly Glu Val Gln Ser Arg Lys Tyr Asn Ile Ser
225 230 235 240

Glu Ser Leu Ser Ser Asn Leu Asp Asp Lys Thr Leu His Glu Leu Tyr

245 250 255
Ala Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met
260 265 270

Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Gln Asn Ser Lys
275 280 285

Leu Leu Asn Gly Ile Leu Lys Asp Glu Met Gly Phe Gln Gly Phe Val
290 295 300

Met Ser Asp Trp Ala Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala
305 310 315 320

Gly Leu Asp Met Ser Met Pro Gly Asp Thr Ala Phe Asp Ser Gly Tyr
325 330 335

Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Ile Asn Gly Thr Val
340 345 350

Pro Ala Trp Arg Val Asp Asp Met Ala Leu Arg Ile Met Ser Ala Phe
355 360 365

Phe Lys Val Gly Lys Thr Val Glu Asp Leu Pro Asp Ile Asn Phe Ser
370 375 380

Ser Trp Thr Arg Asp Thr Phe Gly Phe Val Gln Thr Phe Ala Gln Glu
385 390 395 400

Asn Arg Glu Gln Val Asn Phe Gly Val Asn Val Gln His Asp His Lys
405 410 415

Asn His Ile Arg Glu Ser Ala Ala Lys Gly Ser Val Ile Leu Lys Asn
420 425 430

Thr Gly Ser Leu Pro Leu Asn Asn Pro Lys Phe Leu Ala Val Ile Gly
435 440 445

Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys Gly Asp Arg
450 455 460

Gly Cys Asp Asn Gly Thr Leu Ala Met Ala Trp Gly Ser Gly Thr Ser
465 470 475 480

Gln Phe Pro Tyr Leu Ile Thr Pro Asp Gln Gly Leu Gln Asn Arg Ala
485 490 495

Ala Gln Asp Gly Thr Arg Tyr Glu Ser Ile Leu Thr Asn Asn Glu Trp
500 505 510

Ala Gln Thr Gln Ala Leu Val Ser Gln Pro Asn Val Thr Ala Ile Val
515 520 525

Phe Ala Asn Ala Asp Ser Gly Glu Gly Tyr Ile Glu Val Asp Gly Asn
530 535 540

Phe Gly Asp Arg Lys Asn Leu Thr Leu Trp Gln Gln Gly Asp Glu Leu
545 550 555 560

Ile Lys Asn Val Ser Ser Ile Cys Pro Asn Thr Ile Val Val Leu His
565 570 575

Thr Val Gly Pro Val Leu Leu Ala Asp Tyr Glu Lys Asn Pro Asn Ile
580 585 590

Thr Ala Ile Val Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ala
595 600 605

Ile Ala Asp Leu Leu Tyr Gly Lys Val Ser Pro Gly Arg Ser Pro Phe
610 615 620

Thr Trp Gly Arg Thr Arg Glu Ser Tyr Gly Thr Glu Val Leu Tyr Glu
625 630 635 640

Ala Asn Asn Gly Arg Gly Ala Pro Gln Asp Asp Phe Ser Glu Gly Val
645 650 655

Phe Ile Asp Tyr Arg His Phe Asp Arg Arg Ser Pro Ser Thr Asp Gly
660 665 670

Lys Ser Ala Pro Asn Asn Thr Ala Ala Pro Leu Tyr Glu Phe Gly His
675 680 685

Gly Leu Ser Trp Thr Thr Phe Glu Tyr Ser Asp Leu Asn Ile Gln Lys
690 695 700

Asn Val Asn Ser Thr Tyr Ser Pro Pro Ala Gly Gln Thr Ile Pro Ala
705 710 715 720

Pro Thr Phe Gly Asn Phe Ser Lys Asn Leu Asn Asp Tyr Val Phe Pro
725 730 735

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

740 745 750
Ser Ser Ser Ala Ser Glu Ala Ser Asn Asp Gly Gly Gln Phe Gly Lys

755 760 765
Thr Ala Glu Glu Phe Leu Pro Pro Asn Ala Leu Asn Gly Ser Ala Gln
770 775 780

Pro Arg Leu Pro Ser Ser Gly Ala Pro Gly Gly Asn Pro Gln Leu Trp
785 790 795 800

Asp Ile Leu Tyr Thr Val Thr Ala Thr Ile Thr Asn Thr Gly Asn Ala
805 810 815

Thr Ser Asp Glu Ile Pro Gln Leu Tyr Val Ser Leu Gly Gly Glu Asn
820 825 830

Glu Pro Val Arg Val Leu Arg Gly Phe Asp Arg Ile Glu Asn Ile Ala
835 840 845

Pro Gly Gln Ser Ala Ile Phe Asn Ala Gln Leu Thr Arg Arg Asp Leu
850 855 860

Ser Asn Trp Asp Val Asp Ala Gln Asn Trp Val Ile Thr Asp His Pro
865 870 875 880

Lys Thr Val Trp Val Gly Ser Ser Ser Arg Lys Leu Pro Leu Ser Ala
885 890 895

Lys Leu Glu

<210> 71

<211> 3134

<212> DNA

<213> Gibberella zeae

<400> 71

atgaaggcca attggcttgc cgcggccgtt tatttgctg ctggcaccga tgctgcagtc 60

cctgacactt tggcaggagt caatgtaagc tactcttcaa tttcatctca tctcaacttt 120

gccaggccac aacaactttt cttcactcac gatcttttca ccataaacgc aacagtttca 180

caaaaaataa agcccaaatc atgtctctga tcgttgaact cgccatcttc gtttacatcg 240

cggttgtctt tttcttcttg tacttctcat tcgttgttgt tctctacatt ttcgactggc 300

tgtttagcct ttagattctt ctcactcccc gtgatgccta gatcactctc tgaggcgttt 360

aatctacttg tagagatgcg cctctcattt gttgtgtcgc tagtcgcgat agttgctgga 420
attgcagtcc ttgatcttcc tactgacact caaaagctcg ttgcgcggga cacactcgct 480

cactctcttc ctactatcc ctgccatgg atggacccta acgctgtcgg ctgggaggac 540
gcctacgcca aggccaagga ctttgtctcc cagatgactc tcctagaaaa ggtcaacttg 600
accactgggt ttgggtaagt aacgagcgac aagacgtcta caatccacta acacgatctc 660
tagatggcag ggcgaacgtt gtgttgaaa cgtgggatct atccctcgtc tcggtatgcg 720
aggcctctgt ctccaggatg gtcctctcgg aattcgcttc tccgactaca acagcgcttt 780
ccctactggt gtcaccgctg gtgcttcttg gagtaaggcc ctttggtacg agcgaggacg 840
attgatgggt accgagttta aggagaaggg tatcgatatt gctctcggcc ctgcaactgg 900

tcctctcggg cgccacgctg ctggtggacg aaactgggaa ggcttactg tcgacccta 960
cgccgctggc catgctatgg ctgagactgt caagggtatc caagattctg gagtcattgc 1020
ttgtgctaag cattacatcg caaacgagca aggtatgtac aggccattc aatggcttca 1080
ggaacgaaaa ctaactctta atagaacact tccgtcaacg aggcgatgtc atgtctcaaa 1140
agttcaacat ttccgagctt ctgtcttcca accttgacga taagactatg cacgagctct 1200
acaactggcc ttccgccgac gccgtccgcg ccggtgttgg ctccattatg tgctcttaca 1260
accaggtcaa caactcatat gcttgccaga actccaagct cctcaacggc atcctcaagg 1320

acgagatggg tticcagggt ttcgtcatga gcgattggca ggctcagcac accggtgccg 1380
cctccgctgt tgccggtctt gacatgacca tgcttggtga caccgagttc aacactggct 1440
tcagcttctg ggggtgaaac ctgaccctcg ctgttatcaa cggctactgt cccgctgga 1500
gaatcgacga catggctacc cgaattatgg ctgctttctt caaggttggc cgatctgttg 1560
aggaggaacc cgacatcaac ttctcagctt ggactcgtga tgagtatgac ttcgtccaga 1620
cctacgcccc agagaaccga gaaaagggtca actttgctgt taatgtccag cacgaccaca 1680
agcgcccat tcgcgagggt ggcgcaaagg gatccgtcgt cctcaagaac actggctcac 1740

ttctctttaa gaagccccag ttctcgtg tcattggaga ggacgtggt tccaaccctg 1800
ccggacccaa cggttgcgt gaccgtggat gcgacaacgg tactcttgcc atggcatggg 1860
gttccggaac ctctcaattc ccctaccttg tcacccccga ccaaggcatc tcgtccagg 1920
ctattcagga cggctactgt tatgagagca tctcaacaa caaccagtgg cccagacac 1980
aagctcttgt cagccagccc aacgtcaccg ccattgtctt tgccaatgcc gattctggtg 2040
agggtacat cgaggttgac ggcaactacg gcgaccgcaa gaacctcact ctgtggaagc 2100

aaggcgatga gctcatcaag aacgtctctg ctatctgccc caacaccatt gtggtccttc 2160

acaccgttgg ccccgctcct ctaaccgagt ggacacaaca cccaacatc accgccattg 2220

tttgggctgg tgtgcctgga caggagtcg gtaacgccat cgccgacatc ctctacggca 2280

agaccagccc tggacgttct ccttcacct ggggtcgac ttatgacagc tatggcacca 2340

aggtttctta caaggccaac aatggagagg gtgcccctca agaggacttt gtcgagggca 2400

acttcacga ctaccgccac ttgaccgac aatccccag caccaacgga aagagtcca 2460

ccaacgactc ttctgctcct ctctacgagt tcggtttcgg tctgtcctgg actacctttg 2520

agtactctga tctcaaagtc gagtctgtca gcaacgcctc ttacagcccc tctgtcggaa 2580

acaccattcc tgcccctacc tacggcaact tcagcaagaa cctggacgat tacacattcc 2640

cctcaggtgt ccgatacctc tacaagtcca tctacccta cctcaacacc tcttcctcgg 2700

ctgagaaggc ttccggcgat gtcaagggca gatttggtga gaccggcgac gatttcctcc 2760

ctcccaacgc tctcaacggt tcatcgagc ctgctcttcc ttccagtggg gctcccggcg 2820

gtaaccctca gctctgggac attatgtaca ccgtcactgc caccatcacc aacactggtg 2880

acgttacctc ggatgaggtt cccagctgt acgtcagcct cgggtggtgag ggcgagcctg 2940

tccgtgtcct ccgtggcttc gagcgtcttg aaaacattgc tcctggtgag agtgccacat 3000

tcaccgctca gcttactcgc cgtgacctga gcaactggga cgtcaacgtc cagaactggg 3060

tcatcaccca tcacccaag aagatctggg tcggcagcag ctctcgcaat ctgcccctca 3120

gcgccgacct gtag 3134

<210> 72

<211> 886

<212> PRT

<213> *Gibberella zeae*

<400> 72

Met Lys Ala Asn Trp Leu Ala Ala Val Tyr Leu Ala Ala Gly Thr

1 5 10 15

Asp Ala Ala Val Pro Asp Thr Leu Ala Gly Val Asn Leu Val Ala Arg

20 25 30

Asp Thr Leu Ala His Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp

35 40 45

Pro Asn Ala Val Gly Trp Glu Asp Ala Tyr Ala Lys Ala Lys Asp Phe

50 55 60

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

305 310 315 320
 Thr Gly Phe Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Ile Asn
 325 330 335
 Gly Thr Val Pro Ala Trp Arg Ile Asp Asp Met Ala Thr Arg Ile Met
 340 345 350

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

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

 Leu Gln Ala Ile Gln Asp Gly Thr Arg Tyr Glu Ser Ile Leu Asn Asn
 485 490 495
 Asn Gln Trp Pro Gln Thr Gln Ala Leu Val Ser Gln Pro Asn Val Thr
 500 505 510
 Ala Ile Val Phe Ala Asn Ala Asp Ser Gly Glu Gly Tyr Ile Glu Val
 515 520 525
 Asp Gly Asn Tyr Gly Asp Arg Lys Asn Leu Thr Leu Trp Lys Gln Gly
 530 535 540

 Asp Glu Leu Ile Lys Asn Val Ser Ala Ile Cys Pro Asn Thr Ile Val
 545 550 555 560

Val Leu His Thr Val Gly Pro Val Leu Leu Thr Glu Trp His Asn Asn
565 570 575

Pro Asn Ile Thr Ala Ile Val Trp Ala Gly Val Pro Gly Gln Glu Ser
580 585 590

Gly Asn Ala Ile Ala Asp Ile Leu Tyr Gly Lys Thr Ser Pro Gly Arg
595 600 605

Ser Pro Phe Thr Trp Gly Arg Thr Tyr Asp Ser Tyr Gly Thr Lys Val
610 615 620

Leu Tyr Lys Ala Asn Asn Gly Glu Gly Ala Pro Gln Glu Asp Phe Val
625 630 635 640

Glu Gly Asn Phe Ile Asp Tyr Arg His Phe Asp Arg Gln Ser Pro Ser
645 650 655

Thr Asn Gly Lys Ser Ala Thr Asn Asp Ser Ser Ala Pro Leu Tyr Glu
660 665 670

Phe Gly Phe Gly Leu Ser Trp Thr Thr Phe Glu Tyr Ser Asp Leu Lys
675 680 685

Val Glu Ser Val Ser Asn Ala Ser Tyr Ser Pro Ser Val Gly Asn Thr
690 695 700

Ile Pro Ala Pro Thr Tyr Gly Asn Phe Ser Lys Asn Leu Asp Asp Tyr
705 710 715 720

Thr Phe Pro Ser Gly Val Arg Tyr Leu Tyr Lys Phe Ile Tyr Pro Tyr
725 730 735

Leu Asn Thr Ser Ser Ser Ala Glu Lys Ala Ser Gly Asp Val Lys Gly
740 745 750

Arg Phe Gly Glu Thr Gly Asp Glu Phe Leu Pro Pro Asn Ala Leu Asn
755 760 765

Gly Ser Ser Gln Pro Arg Leu Pro Ser Ser Gly Ala Pro Gly Gly Asn
770 775 780

Pro Gln Leu Trp Asp Ile Met Tyr Thr Val Thr Ala Thr Ile Thr Asn
785 790 795 800

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

805 810 815
 Gly Gly Glu Gly Glu Pro Val Arg Val Leu Arg Gly Phe Glu Arg Leu
 820 825 830
 Glu Asn Ile Ala Pro Gly Glu Ser Ala Thr Phe Thr Ala Gln Leu Thr
 835 840 845
 Arg Arg Asp Leu Ser Asn Trp Asp Val Asn Val Gln Asn Trp Val Ile
 850 855 860

Thr Asp His Ala Lys Lys Ile Trp Val Gly Ser Ser Ser Arg Asn Leu
 865 870 875 880
 Pro Leu Ser Ala Asp Leu
 885

<210> 73

<211> 2796

<212> DNA

<213> Nectria haematococca

<400> 73

atgcggttca ccgtccttct cgcggcattt tcggggcttg tcccatggt tggttcgcaa 60
 gctgaccaga aaccactaca gctcgggtgtg aacaataaca ctctggcgca ttcacctcct 120
 cactatcctt cgccatggat ggatcctgct gctcctggct gggaggaagc ctatctcaag 180
 gcgaaagatt ttgtttcaca gcttaccctt ctgaaaagg tcaacttgac cactggtgtt 240

gggtgagtca ctgttttcc tctctcctga cgtgacactt tgctttggcc tgcttcctat 300
 atcgtctact agcattgcta aactcgagg cagatggatg ggcgaaagtt gcgtcggcaa 360
 cgtgggttca ctccctcgtt ttggaatgcg tggctctctgc atgcaggatg gccccctcgg 420
 catccgcttg tctgactata actctgcctt tctactgggt attacagctg gtgcctcttg 480
 gagccgtgcc ctttggtacc aacgtggcct cctgatgggc accgagcatc gtgaaaaagg 540
 catcgacgtt gcaactgggc ctgctactgg tctcttgggt cgtactccta ctggcggccg 600
 caactgggag ggtttctcgg ttgatcccta cgttgctggc gttgccatgg ccgagactgt 660

tagcggcatt caagatggtg gtactatcgc ctgtgctaag cactacatcg gcaacgaaca 720
 aggtatgcct cttcacttct cctcgtgat aaatctgctc acaacaacct agagcaccat 780
 cgccaagccc ccgaatccat tggccgcggc tacaacatca ccgagtcctt gtcgtcgaac 840
 gttagtgaca agacctcca cgagctctat ctctggccgt tcgcagatgc cgtcaaggct 900
 ggtgttggtg ctatcatgtg ttctaccag cagctgaaca actcttacgg ttgccaaaac 960

tctaagcttc tcaacggaat tctcaaggac gagctaggat tccagggtt cgtcatgagt	1020
gactggcaag cccaacatgc tggagctgct accgctgttg caggccttga catgaccatg	1080
cccggtgaca ctttgttcaa caccggatac agcttctggg gtggtaacct gaccctcgt	1140
glagtcaatg gcactgttcc cgactggcgt attgacgaca tggctatgag aatcatggca	1200
gctttcttca aggttggcaa gactgttgag gaccttctg acatcaactt ttcttcttgg	1260
tctcgagaca cttttggcta cgttcaagcc gctgccaag agaactggga acagatcaac	1320
ttcggagtgt atgttcgtca cgaccacagc gaacacattc gactctcggc cgccaagggc	1380
accgtctcc ttaagaactc tggctcattg cctctgaaga agcccaagtt cttgcccgc	1440
gttggcgagg acgccggccc gaacctgct ggccccaacg gctgtaacga ccgcggatgt	1500
aacaacggca ctctggccat gtcctggggc tcaggaaacag ccagttccc ttacctcgtt	1560
actcccgact cagcgttaca gaaccagct gtcctcgacg gcactcgta cgagagtgtc	1620
ttgcggaaca accagtggga acagacacgc agtctcatta gccaacctaa cgtgacggct	1680
attgtgtttg ccaatgcaa ttccggagag ggatatatcg atgttgacgg caacgaaggc	1740
gatcggaaga atttgacctt gtggaacgag ggtgatgacc taattaagaa cgtctcctca	1800
atctgcccc acaccattgt tgttctgcac actgttggcc ctgtcatcct gacggaatgg	1860
tatgacaacc cgaacattac cgccatagtg tgggctgggtg tacctggaca ggagtccggc	1920
aatgctcttg tggacatcct ttatggcaaa acaagccctg gtcgctctcc cttcacatgg	1980
ggtcgaccc gaaagagtta cggcactgat gtctatacg agcccaaca tggtcagggt	2040
gtcctcaag atgatttcac ggaggaggctc ttatcgact atcgtcattt tgaccagggt	2100
tctcctagca ccgacggcag caagtcta at gatgagcca gtcccatcta cgagtttggc	2160
catggtctgt cctggaccac gtttgagtac tctgaactca acattcaagc tcacaacaag	2220
attcccttcg atcctcctat tggcgagacg attgccgctc cggtccttgg caactacagt	2280
accgaccttg ccgattacac gttccccgat ggaattcgt acatctacca gttcatctat	2340
ccctggttga atacttcttc ttccggaaga gaggtctctg gcgatcccga ctacggaaag	2400
acggccgaag agttcctgcc ccccgagct ctcgacgggt cagctcagcc gcgacctcca	2460
tcctctggtg ctccagggtg aaacctcat ctttgggatg tgttgtacac tgtagtgct	2520
atcatcacca aacttgcaa cgccacctcg gacgagatcc cgcagctcta cgttagtctc	2580
ggtggcgaga acgagccgt ccgcgtcctt cgcgggttcg accgaattga gaacattgcg	2640
cctggccaga gtgtcagatt cacaactgac atcactcgcc gcgacctgag caactgggac	2700

gtcgtctctc agaactgggt cattacagac tacgagaaga ccgtatatgt cgggagcagc 2760

tcccgaacc tgcctctcaa ggcaaccctg aagtaa 2796

<210> 74

<211> 880

<212> PRT

<213> *Nectria haematococca*

<400> 74

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

1 5 10 15

Val Gly Ser Gln Ala Asp Gln Lys Pro Leu Gln Leu Gly Val Asn Asn

20 25 30

Asn Thr Leu Ala His Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp

35 40 45

Pro Ala Ala Pro Gly Trp Glu Glu Ala Tyr Leu Lys Ala Lys Asp Phe

50 55 60

Val Ser Gln Leu Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val

65 70 75 80

Gly Trp Met Gly Glu Arg Cys Val Gly Asn Val Gly Ser Leu Pro Arg

85 90 95

Phe Gly Met Arg Gly Leu Cys Met Gln Asp Gly Pro Leu Gly Ile Arg

100 105 110

Leu Ser Asp Tyr Asn Ser Ala Phe Pro Thr Gly Ile Thr Ala Gly Ala

115 120 125

Ser Trp Ser Arg Ala Leu Trp Tyr Gln Arg Gly Leu Leu Met Gly Thr

130 135 140

Glu His Arg Glu Lys Gly Ile Asp Val Ala Leu Gly Pro Ala Thr Gly

145 150 155 160

Pro Leu Gly Arg Thr Pro Thr Gly Gly Arg Asn Trp Glu Gly Phe Ser

165 170 175

Val Asp Pro Tyr Val Ala Gly Val Ala Met Ala Glu Thr Val Ser Gly

180 185 190

Ile Gln Asp Gly Gly Thr Ile Ala Cys Ala Lys His Tyr Ile Gly Asn
 195 200 205
 Glu Gln Glu His His Arg Gln Ala Pro Glu Ser Ile Gly Arg Gly Tyr
 210 215 220
 Asn Ile Thr Glu Ser Leu Ser Ser Asn Val Asp Asp Lys Thr Leu His
 225 230 235 240

 Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Lys Ala Gly Val Gly
 245 250 255
 Ala Ile Met Cys Ser Tyr Gln Gln Leu Asn Asn Ser Tyr Gly Cys Gln
 260 265 270
 Asn Ser Lys Leu Leu Asn Gly Ile Leu Lys Asp Glu Leu Gly Phe Gln
 275 280 285
 Gly Phe Val Met Ser Asp Trp Gln Ala Gln His Ala Gly Ala Ala Thr
 290 295 300

 Ala Val Ala Gly Leu Asp Met Thr Met Pro Gly Asp Thr Leu Phe Asn
 305 310 315 320
 Thr Gly Tyr Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Val Asn
 325 330 335
 Gly Thr Val Pro Asp Trp Arg Ile Asp Asp Met Ala Met Arg Ile Met
 340 345 350
 Ala Ala Phe Phe Lys Val Gly Lys Thr Val Glu Asp Leu Pro Asp Ile
 355 360 365

 Asn Phe Ser Ser Trp Ser Arg Asp Thr Phe Gly Tyr Val Gln Ala Ala
 370 375 380
 Ala Gln Glu Asn Trp Glu Gln Ile Asn Phe Gly Val Asp Val Arg His
 385 390 395 400
 Asp His Ser Glu His Ile Arg Leu Ser Ala Ala Lys Gly Thr Val Leu
 405 410 415
 Leu Lys Asn Ser Gly Ser Leu Pro Leu Lys Lys Pro Lys Phe Leu Ala
 420 425 430

 Val Val Gly Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys

435 440 445
 Asn Asp Arg Gly Cys Asn Asn Gly Thr Leu Ala Met Ser Trp Gly Ser
 450 455 460
 Gly Thr Ala Gln Phe Pro Tyr Leu Val Thr Pro Asp Ser Ala Leu Gln
 465 470 475 480
 Asn Gln Ala Val Leu Asp Gly Thr Arg Tyr Glu Ser Val Leu Arg Asn
 485 490 495

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

 Val Leu His Thr Val Gly Pro Val Ile Leu Thr Glu Trp Tyr Asp Asn
 565 570 575
 Pro Asn Ile Thr Ala Ile Val Trp Ala Gly Val Pro Gly Gln Glu Ser
 580 585 590
 Gly Asn Ala Leu Val Asp Ile Leu Tyr Gly Lys Thr Ser Pro Gly Arg
 595 600 605
 Ser Pro Phe Thr Trp Gly Arg Thr Arg Lys Ser Tyr Gly Thr Asp Val
 610 615 620

 Leu Tyr Glu Pro Asn Asn Gly Gln Gly Ala Pro Gln Asp Asp Phe Thr
 625 630 635 640
 Glu Gly Val Phe Ile Asp Tyr Arg His Phe Asp Gln Val Ser Pro Ser
 645 650 655
 Thr Asp Gly Ser Lys Ser Asn Asp Glu Ser Ser Pro Ile Tyr Glu Phe
 660 665 670
 Gly His Gly Leu Ser Trp Thr Thr Phe Glu Tyr Ser Glu Leu Asn Ile
 675 680 685

Gln Ala His Asn Lys Ile Pro Phe Asp Pro Pro Ile Gly Glu Thr Ile
690 695 700
Ala Ala Pro Val Leu Gly Asn Tyr Ser Thr Asp Leu Ala Asp Tyr Thr
705 710 715 720
Phe Pro Asp Gly Ile Arg Tyr Ile Tyr Gln Phe Ile Tyr Pro Trp Leu
725 730 735
Asn Thr Ser Ser Ser Gly Arg Glu Ala Ser Gly Asp Pro Asp Tyr Gly
740 745 750

Lys Thr Ala Glu Glu Phe Leu Pro Pro Gly Ala Leu Asp Gly Ser Ala
755 760 765
Gln Pro Arg Pro Pro Ser Ser Gly Ala Pro Gly Gly Asn Pro His Leu
770 775 780
Trp Asp Val Leu Tyr Thr Val Ser Ala Ile Ile Thr Asn Thr Gly Asn
785 790 795 800
Ala Thr Ser Asp Glu Ile Pro Gln Leu Tyr Val Ser Leu Gly Gly Glu
805 810 815

Asn Glu Pro Val Arg Val Leu Arg Gly Phe Asp Arg Ile Glu Asn Ile
820 825 830
Ala Pro Gly Gln Ser Val Arg Phe Thr Thr Asp Ile Thr Arg Arg Asp
835 840 845
Leu Ser Asn Trp Asp Val Val Ser Gln Asn Trp Val Ile Thr Asp Tyr
850 855 860
Glu Lys Thr Val Tyr Val Gly Ser Ser Ser Arg Asn Leu Pro Leu Lys
865 870 875 880

<210> 75

<211> 3169

<212> DNA

<213> *Verticillium dahliae*

<400> 75

atgaagctga ccctcgctac tgccttactg gcagccagcg ggtgtgtctc tgcgggacaa	60
cccaagctca aggtacgtac ttgcctcttt ttcacaagga aaccaaacc gcaccataat	120
ggtgattgag cagtcgtgct ttctcaacc cgaatcaaac ccatgccgtg ttcgcgcatg	180

ccctttcgat cgictgttgt gtgtgaaccc acgctcttca agcatcgac atagcaccac	240
tccatcttca ttttcgagca atttcgggcc gcagagagcg gtccttact tcaccacaat	300
cgttcatgcc tcgtgcccc ctgccatgtt tcttcccagt attctacttc tgagagcctt	360
gaccaccgtt gtcgacatct cgtcgccaag gctcgttgac acggactctg tttcccttgg	420
aattaatatt cgaacaatg ctgaccagca tcttcagcgc cagactaaca gctctagcga	480
gctcgccctt tcccctccgc actacccttc tccatggatg aacccccaa cgactgggtg	540
ggaggacgcc tacgcccgtg ccagagaggt gtagagcag atgactctgc tcgaaaaggt	600
caactgacg acaggtgtcg ggtaagcttc acagaccccg tcttgccatc caaagtcac	660
tgacagaatc ctagctggag cggatgcttc tgcgtcggaa acgtcggctc gatccccga	720
atcggctgga gggggctttg tttgcaggat ggcccacagg gtatccgttt cgcggactac	780
gtctcgtact tcaattcgag ccagacagcc ggcgctacct gggaccgagg gcttctgtac	840
cagcgcgctc acgccattgg cgccaagga gtagccaagg gcgtcgacgt cgtcctcggg	900
cccgccattg gccctctagg tcgcttccc gccggaggtc gtaactggga gggtttcgcc	960
gtggaccctt acctcagtg cgttgctgtc gccgaatccg tcaggggcat ccaggatgct	1020
ggtgctattg ccaacgtcaa gcactacatc gtcaatgagc aggaacattt ccgccaggct	1080
ggcgaggctc aaggttacgg ctacgatgtc gacgaggcat tatcgtcgaa cgttgacgac	1140
aagaccatgc atgagcttta cctttggcca tttgcagacg ctgtccgtgc tggagccggc	1200
agtgatcatgt gttcttatca acaggtgggg gcaataccat tctctcctct ttccttgcag	1260
acagtgcact gaccgacctt ttttcccaa gatcaacaac agttacggct gtcaaaactc	1320
acatcttctg aatgggctcc tcaaggacga actcggcttt caggggttcg tcctcagcga	1380
ttggcaagcg cagcatgctg gtgctgccac tgccttgct ggacttgaca tggccatgcc	1440
cggatgacact cgttcaaca ccggagtcgc cttctggggc gctaacctta ccaatgcat	1500
tttgaacggc accgttcccg aatatcggct cgatgacatg gccatgcgta ttatggcggc	1560
ctttttcaaa gttggaaaga ccctggacga tgttctgac atcaacttct cgtcttggac	1620
aaaagacacc atcgccccg tgcactgggc ggcccaggac aatgtgcagg tcatcaacca	1680
acacgttgat gtccgtcaag accacggcgc cctcattcgc accatcgctg cccgcggtac	1740
tgtcttacta aaaaatgagg gatcactgcc tctgaacaag ccgaaatttg ttgctgtcat	1800
tggatgaagat gctggccctc gtccgttttg tcccaatggc tgccctgac aggggttgcaa	1860
taacggcact ctggctgctg gatggggatc tggcaccgcc agtttccctt atctcatcac	1920
tcctgatagt gctcttcagt ttcaagccgt ttcgatggc tcgcgatac aaagcatcct	1980

cagcaactgg gattatgagc gcacagaggc cttggtttcc caggcggatg ctactgctct 2040

ggttttcgtc aatgcaaact ctggcgaagg atatatcagc gttgatggaa acgaaggtga 2100

tcgcaagaac ctactctct ggaatggagg agacgagctt attcaacgag tcgctcgggc 2160

caacaacaac accatcgta tcatccattc ggttgggtccc gttctagtca ctgactggta 2220

cgagaatccc aatatcacgg ctatcatctg ggccggctta cccggacagg agtctggcaa 2280

ctctatcgcc gatattcttt acggccgctg gaaccttggg ggcaagacac ctttcacctg 2340

gggtccaact gttgagagct acggcgttga cgtcctgaga gagcccaaca atggcaatgg 2400

tgctccccag agcgatttcg acgagggagt cttcatcgat taccgttggg ttgaccggca 2460

gtcgggtgtt gataacaatg catcagcgcc gaggaacagc agcagcagcc acgccccaat 2520

cttcgagttt ggctatggcc ttctgtacac aacctttgaa ttctccaatc ttcagattga 2580

gaggcatgac gttcacgatt acgtccctac cactgggcag acgagccctg cgccgagatt 2640

tgggtgtaac tacagtacga actacgacga ctacgtcttt cccgagggcg aaatccgtta 2700

catctatcaa cacatctacc catacctcaa ttctcagac ccaaaggagg cattggctga 2760

tcctaaatac ggccaaactg cagaagagtt cctcccagag ggcgctcttg atgcctcacc 2820

gcagcctagg ctcccagctt ctggagggcc cggaggcaac ccaatgcttt gggacgtcat 2880

attcacggtc accgcgaccg tgaccaacac gggttaagggt gctggggacg aagtggcaca 2940

gcttttacgtt tcctttgggt gacctgacga tccgattcga gtcctccgtg ggttcgaccg 3000

cattcacatc gcgcctggag cctcgcaaac ctccgtgctg gaactcacgc gccgggacct 3060

cagcaactgg gatgttgta cgcaaaattg gttcatcagc cagtacgaaa agacggtctt 3120

tgtcgggagc tcatcccga accctcctct cagcactcgc ctggaatag 3169

<210> 76

<211> 890

<212> PRT

<213> *Verticillium dahliae*

<400> 76

Met Lys Leu Thr Leu Ala Thr Ala Leu Leu Ala Ala Ser Gly Cys Val

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

Asn Pro Gln Ala Thr Gly Trp Glu Asp Ala Tyr Ala Arg Ala Arg Glu
 50 55 60
 Val Val Glu Gln Met Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly

 65 70 75 80
 Val Gly Trp Ser Gly Asp Leu Cys Val Gly Asn Val Gly Ser Ile Pro
 85 90 95
 Arg Ile Gly Trp Arg Gly Leu Cys Leu Gln Asp Gly Pro Gln Gly Ile
 100 105 110
 Arg Phe Ala Asp Tyr Val Ser Tyr Phe Thr Ser Ser Gln Thr Ala Gly
 115 120 125
 Ala Thr Trp Asp Arg Gly Leu Leu Tyr Gln Arg Ala His Ala Ile Gly

 130 135 140
 Ala Glu Gly Val Ala Lys Gly Val Asp Val Val Leu Gly Pro Ala Ile
 145 150 155 160
 Gly Pro Leu Gly Arg Leu Pro Ala Gly Gly Arg Asn Trp Glu Gly Phe
 165 170 175
 Ala Val Asp Pro Tyr Leu Ser Gly Val Ala Val Ala Glu Ser Val Arg
 180 185 190
 Gly Ile Gln Asp Ala Gly Ala Ile Ala Asn Val Lys His Tyr Ile Val

 195 200 205
 Asn Glu Gln Glu His Phe Arg Gln Ala Gly Glu Ala Gln Gly Tyr Gly
 210 215 220
 Tyr Asp Val Asp Glu Ala Leu Ser Ser Asn Val Asp Asp Lys Thr Met
 225 230 235 240
 His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Ala
 245 250 255
 Gly Ser Val Met Cys Ser Tyr Gln Gln Ile Asn Asn Ser Tyr Gly Cys

 260 265 270
 Gln Asn Ser His Leu Leu Asn Gly Leu Leu Lys Asp Glu Leu Gly Phe
 275 280 285
 Gln Gly Phe Val Leu Ser Asp Trp Gln Ala Gln His Ala Gly Ala Ala

290 295 300
 Thr Ala Val Ala Gly Leu Asp Met Ala Met Pro Gly Asp Thr Arg Phe
 305 310 315 320
 Asn Thr Gly Val Ala Phe Trp Gly Ala Asn Leu Thr Asn Ala Ile Leu

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

 385 390 395 400
 Gln Asp His Gly Ala Leu Ile Arg Thr Ile Ala Ala Arg Gly Thr Val
 405 410 415
 Leu Leu Lys Asn Glu Gly Ser Leu Pro Leu Asn Lys Pro Lys Phe Val
 420 425 430
 Ala Val Ile Gly Glu Asp Ala Gly Pro Arg Pro Val Gly Pro Asn Gly
 435 440 445
 Cys Pro Asp Gln Gly Cys Asn Asn Gly Thr Leu Ala Ala Gly Trp Gly

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

 515 520 525
 Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp Asn Gly
 530 535 540

Gly Asp Glu Leu Ile Gln Arg Val Ala Ala Ala Asn Asn Asn Thr Ile
 545 550 555 560
 Val Ile Ile His Ser Val Gly Pro Val Leu Val Thr Asp Trp Tyr Glu
 565 570 575
 Asn Pro Asn Ile Thr Ala Ile Ile Trp Ala Gly Leu Pro Gly Gln Glu
 580 585 590
 Ser Gly Asn Ser Ile Ala Asp Ile Leu Tyr Gly Arg Val Asn Pro Gly
 595 600 605
 Gly Lys Thr Pro Phe Thr Trp Gly Pro Thr Val Glu Ser Tyr Gly Val
 610 615 620
 Asp Val Leu Arg Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln Ser Asp
 625 630 635 640
 Phe Asp Glu Gly Val Phe Ile Asp Tyr Arg Trp Phe Asp Arg Gln Ser
 645 650 655
 Gly Val Asp Asn Asn Ala Ser Ala Pro Arg Asn Ser Ser Ser Ser His
 660 665 670
 Ala Pro Ile Phe Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu
 675 680 685
 Phe Ser Asn Leu Gln Ile Glu Arg His Asp Val His Asp Tyr Val Pro
 690 695 700
 Thr Thr Gly Gln Thr Ser Pro Ala Pro Arg Phe Gly Ala Asn Tyr Ser
 705 710 715 720
 Thr Asn Tyr Asp Asp Tyr Val Phe Pro Glu Gly Glu Ile Arg Tyr Ile
 725 730 735
 Tyr Gln His Ile Tyr Pro Tyr Leu Asn Ser Ser Asp Pro Lys Glu Ala
 740 745 750
 Leu Ala Asp Pro Lys Tyr Gly Gln Thr Ala Glu Glu Phe Leu Pro Glu
 755 760 765
 Gly Ala Leu Asp Ala Ser Pro Gln Pro Arg Leu Pro Ala Ser Gly Gly
 770 775 780
 Pro Gly Gly Asn Pro Met Leu Trp Asp Val Ile Phe Thr Val Thr Ala

785 790 795 800
 Thr Val Thr Asn Thr Gly Lys Val Ala Gly Asp Glu Val Ala Gln Leu
 805 810 815
 Tyr Val Ser Leu Gly Gly Pro Asp Asp Pro Ile Arg Val Leu Arg Gly
 820 825 830
 Phe Asp Arg Ile His Ile Ala Pro Gly Ala Ser Gln Thr Phe Arg Ala

 835 840 845
 Glu Leu Thr Arg Arg Asp Leu Ser Asn Trp Asp Val Val Thr Gln Asn
 850 855 860
 Trp Phe Ile Ser Gln Tyr Glu Lys Thr Val Phe Val Gly Ser Ser Ser
 865 870 875 880
 Arg Asn Leu Pro Leu Ser Thr Arg Leu Glu
 885 890

<210> 77

<211> 2418

<212> DNA

<213> Podospora anserina

<400> 77

atgaaactca ataagccatt cctggccatt tatttgctt tcaacttggc cgaggcttcg 60

aaaactccgg attgcatcag tggtcgctg gcaaagacct tggcatgtga tacaacggcg 120

tcacctcctg cgcgagcagc tgctcttggt caggctttaa atatcacgga aaagcttggtg 180

aatctagtgg agtatgtcaa gtcaagagaa gtccttttag ggatttcaat tcagctaatac 240

actcctcata gcatgagcct cgggtgcagaa aggatcggcc ttccagctta tgcttggtgg 300

aacgaagctc ttcatggtgt tgccgcgtcg cctgggggtct ctttcaatca ggccggacaa 360

gaattctcac acgctacttc atttgcaat actattacgc tagcagccgc ctttgacaat 420

gacctggttt acgaggtggc ggataccatc agcactgaag cgcgagcggt cagcaatgcc 480

gagctcgtg gactggatta ctggacgcct aacatcaacc cgtacaaaga tccgagatgg 540

gggagggggc atgaggtttg ttaccttagc cttcttttcc gtgccgtgca gttgctgaga 600

actcaaaga caccgggaga agatccggtg cacatcaaag gctacgtcca agcacttctc 660

gagggtctag aaggagagaga caagatcaga aaggtgattg ccacttgtaa acactttgca 720

gcctatgatt tggagagatg gcaaggggct cttagataga ggttcaatgc tgttgtagacc 780

tcgcaggatc tttcgagata ctacctcaa ccgtttcaac aatgcgctcg agacagcaag 840

gtcgggtctt tcatgtgtc atataatgcg ctcaacggaa caccggcatg tgcaagcacg 900

tatttgatgg acgacatcct tcgaaaacac tggaattgga ccgagcacia caactatata 960

acgagcgact gtaatgtat tcaggacttc ctccccaact ttcacaactt cagccaaact 1020

ccagctcaag ccgccgtga tgcttataac gccggtacag acaccgtctg tgagggtcct 1080

ggatacccc cactcacaga tgtaatcgga gcatacaatc agtctctgct gtcagaggaa 1140

attatcgacc gagcacttcg cagattatac gaaggcctca tccgagctgg ctatctcgac 1200

tcagcctccc cacatccata caccaaaatc tcatggtccc aagtaaacac ccccaaagcc 1260

caagccctgg ctctccagtc cgccaccgac gggatagtc tctcaaaaa caacggcctc 1320

cttcccctag acctcacaa caaaaccata gccctcatag gccactgggc caatgcaacc 1380

cgccaaatgc taggcggcta cagcggatc ccccttact acgccaaccc aatctatgca 1440

gccaccagc tcaacgtcac ttttcatcac gcccaggac cggtaacca gtcctctccc 1500

tccacaaatg acacctggac ctccccgcc ctctccgagg cttccaaatc ggataatc 1560

ctctacctcg gcggcaccga cctctccatc gcagccgaag accgagacag agactccatc 1620

gcctggccat ccgtcaact ttcttgtta acctccctcg cccagatggg aaaaccaca 1680

atcgtagcaa gactaggcga ccaagtagac gacaccccc tgctctcaa cccaaacatc 1740

tcctccatc tatgggtagg ctaccaggc caatcaggcg gaacagccct ctgaaacatc 1800

atcacggag tcagctcccc cgccgtcga ctgcccgtca cagctaccc agaaacttac 1860

acctccctca tccccctgac agccatgtcc ctccgccc aa cctccgccg cccaggccgg 1920

acttacaggt ggtaccctc ccccgctc ccttcggcc acggcctca ctacacaacc 1980

tttacgcca aatcggcgt ctttgagtcc ctacccatca acattgccga actcgtttcc 2040

aactgtaac aacgatacct cgacctctgc cggttccgc aggtgtccgt ctgggtgtcg 2100

aatacgggag aactcaaatc tgactatgtc gcccttgttt ttgtcagggg tgagtacgga 2160

ccggagccgt acccgatcaa gacgtggtg gggtaacaagc ggataaggga tatcgagccg 2220

gggactacgg gggcgccgc ggtgggggtg gtggtggggg atttggttag ggtggatttg 2280

ggggggaata gggttttgtt tccggggaag tatgagtctc tgctggatgt ggaggggggg 2340

agggataggg ttgatcgga gttggttggg gaggaggtgg tgttgagaa gttccctcag 2400

ccgctgagg cgggttga 2418

<210> 78

<211> 805

<212> PRT

<213> Podospira anserina

<400> 78

Met Lys Leu Asn Lys Pro Phe Leu Ala Ile Tyr Leu Ala Phe Asn Leu

1 5 10 15
Ala Glu Ala Ser Lys Thr Pro Asp Cys Ile Ser Gly Pro Leu Ala Lys
20 25 30
Thr Leu Ala Cys Asp Thr Thr Ala Ser Pro Pro Ala Arg Ala Ala Ala
35 40 45
Leu Val Gln Ala Leu Asn Ile Thr Glu Lys Leu Val Asn Leu Val Glu
50 55 60
Tyr Val Lys Ser Arg Glu Ala Pro Leu Gly Ile Ser Ile Gln Leu Ile

65 70 75 80
Thr Pro His Ser Met Ser Leu Gly Ala Glu Arg Ile Gly Leu Pro Ala
85 90 95
Tyr Ala Trp Trp Asn Glu Ala Leu His Gly Val Ala Ala Ser Pro Gly
100 105 110
Val Ser Phe Asn Gln Ala Gly Gln Glu Phe Ser His Ala Thr Ser Phe
115 120 125
Ala Asn Thr Ile Thr Leu Ala Ala Ala Phe Asp Asn Asp Leu Val Tyr

130 135 140
Glu Val Ala Asp Thr Ile Ser Thr Glu Ala Arg Ala Phe Ser Asn Ala
145 150 155 160
Glu Leu Ala Gly Leu Asp Tyr Trp Thr Pro Asn Ile Asn Pro Tyr Lys
165 170 175
Asp Pro Arg Trp Gly Arg Gly His Glu Val Cys Tyr Leu Ser Leu Leu
180 185 190
Phe Arg Ala Val Gln Leu Leu Arg Thr Gln Lys Thr Pro Gly Glu Asp

195 200 205
Pro Val His Ile Lys Gly Tyr Val Gln Ala Leu Leu Glu Gly Leu Glu
210 215 220
Gly Arg Asp Lys Ile Arg Lys Val Ile Ala Thr Cys Lys His Phe Ala
225 230 235 240

Ala Tyr Asp Leu Glu Arg Trp Gln Gly Ala Leu Arg Tyr Arg Phe Asn
245 250 255

Ala Val Val Thr Ser Gln Asp Leu Ser Glu Tyr Tyr Leu Gln Pro Phe
260 265 270

Gln Gln Cys Ala Arg Asp Ser Lys Val Gly Ser Phe Met Cys Ser Tyr
275 280 285

Asn Ala Leu Asn Gly Thr Pro Ala Cys Ala Ser Thr Tyr Leu Met Asp
290 295 300

Asp Ile Leu Arg Lys His Trp Asn Trp Thr Glu His Asn Asn Tyr Ile
305 310 315 320

Thr Ser Asp Cys Asn Ala Ile Gln Asp Phe Leu Pro Asn Phe His Asn
325 330 335

Phe Ser Gln Thr Pro Ala Gln Ala Ala Ala Asp Ala Tyr Asn Ala Gly
340 345 350

Thr Asp Thr Val Cys Glu Val Pro Gly Tyr Pro Pro Leu Thr Asp Val
355 360 365

Ile Gly Ala Tyr Asn Gln Ser Leu Leu Ser Glu Glu Ile Ile Asp Arg
370 375 380

Ala Leu Arg Arg Leu Tyr Glu Gly Leu Ile Arg Ala Gly Tyr Leu Asp
385 390 395 400

Ser Ala Ser Pro His Pro Tyr Thr Lys Ile Ser Trp Ser Gln Val Asn
405 410 415

Thr Pro Lys Ala Gln Ala Leu Ala Leu Gln Ser Ala Thr Asp Gly Ile
420 425 430

Val Leu Leu Lys Asn Asn Gly Leu Leu Pro Leu Asp Leu Thr Asn Lys
435 440 445

Thr Ile Ala Leu Ile Gly His Trp Ala Asn Ala Thr Arg Gln Met Leu
450 455 460

Gly Gly Tyr Ser Gly Ile Pro Pro Tyr Tyr Ala Asn Pro Ile Tyr Ala
465 470 475 480

Ala Thr Gln Leu Asn Val Thr Phe His His Ala Pro Gly Pro Val Asn

485 490 495
 Gln Ser Ser Pro Ser Thr Asn Asp Thr Trp Thr Ser Pro Ala Leu Ser
 500 505 510
 Ala Ala Ser Lys Ser Asp Ile Ile Leu Tyr Leu Gly Gly Thr Asp Leu

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

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

 645 650 655
 His Tyr Thr Thr Phe Thr Ala Lys Phe Gly Val Phe Glu Ser Leu Thr
 660 665 670
 Ile Asn Ile Ala Glu Leu Val Ser Asn Cys Asn Glu Arg Tyr Leu Asp
 675 680 685
 Leu Cys Arg Phe Pro Gln Val Ser Val Trp Val Ser Asn Thr Gly Glu
 690 695 700
 Leu Lys Ser Asp Tyr Val Ala Leu Val Phe Val Arg Gly Glu Tyr Gly

 705 710 715 720
 Pro Glu Pro Tyr Pro Ile Lys Thr Leu Val Gly Tyr Lys Arg Ile Arg
 725 730 735

Asp Ile Glu Pro Gly Thr Thr Gly Ala Ala Pro Val Gly Val Val Val
740 745 750
Gly Asp Leu Ala Arg Val Asp Leu Gly Gly Asn Arg Val Leu Phe Pro
755 760 765
Gly Lys Tyr Glu Phe Leu Leu Asp Val Glu Gly Gly Arg Asp Arg Val
770 775 780
Val Ile Glu Leu Val Gly Glu Glu Val Val Leu Glu Lys Phe Pro Gln
785 790 795 800
Pro Pro Ala Ala Gly
805
<210> 79
<211> 721
<212> PRT
<213> Thermotoga neapolitana
<400> 79
Met Glu Lys Val Asn Glu Ile Leu Ser Gln Leu Thr Leu Glu Glu Lys
1 5 10 15
Val Lys Leu Val Val Gly Val Gly Leu Pro Gly Leu Phe Gly Asn Pro
20 25 30
His Ser Arg Val Ala Gly Ala Ala Gly Glu Thr His Pro Val Pro Arg
35 40 45
Val Gly Leu Pro Ala Phe Val Leu Ala Asp Gly Pro Ala Gly Leu Arg
50 55 60
Ile Asn Pro Thr Arg Glu Asn Asp Glu Asn Thr Tyr Tyr Thr Thr Ala
65 70 75 80
Phe Pro Val Glu Ile Met Leu Ala Ser Thr Trp Asn Arg Glu Leu Leu
85 90 95
Glu Glu Val Gly Lys Ala Met Gly Glu Glu Val Arg Glu Tyr Gly Val
100 105 110
Asp Val Leu Leu Ala Pro Ala Met Asn Ile His Arg Asn Pro Leu Cys
115 120 125
Gly Arg Asn Phe Glu Tyr Tyr Ser Glu Asp Pro Val Leu Ser Gly Glu

130 135 140
 Met Ala Ser Ser Phe Val Lys Gly Val Gln Ser Gln Gly Val Gly Ala

 145 150 155 160
 Cys Ile Lys His Phe Val Ala Asn Asn Gln Glu Thr Asn Arg Met Val

 165 170 175
 Val Asp Thr Ile Val Ser Glu Arg Ala Leu Arg Glu Ile Tyr Leu Arg

 180 185 190
 Gly Phe Glu Ile Ala Val Lys Lys Ser Lys Pro Trp Ser Val Met Ser

 195 200 205
 Ala Tyr Asn Lys Leu Asn Gly Lys Tyr Cys Ser Gln Asn Glu Trp Leu

 210 215 220
 Leu Lys Lys Val Leu Arg Glu Glu Trp Gly Phe Glu Gly Phe Val Met

 225 230 235 240
 Ser Asp Trp Tyr Ala Gly Asp Asn Pro Val Glu Gln Leu Lys Ala Gly

 245 250 255
 Asn Asp Leu Ile Met Pro Gly Lys Ala Tyr Gln Val Asn Thr Glu Arg

 260 265 270
 Arg Asp Glu Ile Glu Glu Ile Met Glu Ala Leu Lys Glu Gly Lys Leu

 275 280 285
 Ser Glu Glu Val Leu Asp Glu Cys Val Arg Asn Ile Leu Lys Val Leu

 290 295 300
 Val Asn Ala Pro Ser Phe Lys Asn Tyr Arg Tyr Ser Asn Lys Pro Asp

 305 310 315 320
 Leu Glu Lys His Ala Lys Val Ala Tyr Glu Ala Gly Ala Glu Gly Val

 325 330 335
 Val Leu Leu Arg Asn Glu Glu Ala Leu Pro Leu Ser Glu Asn Ser Lys

 340 345 350
 Ile Ala Leu Phe Gly Thr Gly Gln Ile Glu Thr Ile Lys Gly Gly Thr

 355 360 365
 Gly Ser Gly Asp Thr His Pro Arg Tyr Ala Ile Ser Ile Leu Glu Gly

 370 375 380

Ile Lys Glu Arg Gly Leu Asn Phe Asp Glu Glu Leu Ala Lys Thr Tyr
385 390 395 400
Glu Asp Tyr Ile Lys Lys Met Arg Glu Thr Glu Glu Tyr Lys Pro Arg
405 410 415
Arg Asp Ser Trp Gly Thr Ile Ile Lys Pro Lys Leu Pro Glu Asn Phe
420 425 430
Leu Ser Glu Lys Glu Ile His Lys Leu Ala Lys Lys Asn Asp Val Ala
435 440 445
Val Ile Val Ile Ser Arg Ile Ser Gly Glu Gly Tyr Asp Arg Lys Pro
450 455 460
Val Lys Gly Asp Phe Tyr Leu Ser Asp Asp Glu Thr Asp Leu Ile Lys
465 470 475 480
Thr Val Ser Arg Glu Phe His Glu Gln Gly Lys Lys Val Ile Val Leu
485 490 495
Leu Asn Ile Gly Ser Pro Val Glu Val Val Ser Trp Arg Asp Leu Val
500 505 510
Asp Gly Ile Leu Leu Val Trp Gln Ala Gly Gln Glu Thr Gly Arg Ile
515 520 525
Val Ala Asp Val Leu Thr Gly Arg Ile Asn Pro Ser Gly Lys Leu Pro
530 535 540
Thr Thr Phe Pro Arg Asp Tyr Ser Asp Val Pro Ser Trp Thr Phe Pro
545 550 555 560
Gly Glu Pro Lys Asp Asn Pro Gln Lys Val Val Tyr Glu Glu Asp Ile
565 570 575
Tyr Val Gly Tyr Arg Tyr Tyr Asp Thr Phe Gly Val Glu Pro Ala Tyr
580 585 590
Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asp Leu
595 600 605
Asn Val Ser Phe Asp Gly Glu Thr Leu Arg Val Gln Tyr Arg Ile Glu
610 615 620
Asn Thr Gly Gly Arg Ala Gly Lys Glu Val Ser Gln Val Tyr Ile Lys

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

 660 665 670
 Glu Ile Pro Val Arg Asp Leu Ala Ser Phe Asn Gly Glu Glu Trp Val
 675 680 685
 Val Glu Ala Gly Glu Tyr Glu Val Arg Val Gly Ala Ser Ser Arg Asn
 690 695 700
 Ile Lys Leu Lys Gly Thr Phe Ser Val Gly Glu Glu Arg Arg Phe Lys
 705 710 715 720
 Pro

<210> 80

<211> 871

<212> PRT

<213> *Podospora anserina*

<400> 80

Met Ala Tyr Arg Ser Leu Val Leu Gly Ala Phe Ala Ser Thr Ser Leu
 1 5 10 15
 Ala Ala Ser Val Val Thr Pro Arg Asp Pro Val Pro Pro Gly Phe Val
 20 25 30
 Ala Ala Pro Tyr Tyr Pro Ala Pro His Gly Gly Trp Val Ala Ser Trp
 35 40 45
 Glu Glu Ala Tyr Ser Lys Ala Glu Ala Leu Val Ser Gln Met Thr Leu
 50 55 60

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

Gly Ile Thr Thr Gly Ala Thr Phe Asp Lys Lys Leu Ile Tyr Ala Arg
115 120 125

Gly Val Ala Ile Gly Glu Glu His Arg Gly Lys Gly Thr Asn Val Tyr
130 135 140

Leu Gly Pro Ser Val Gly Pro Leu Gly Arg Lys Pro Leu Gly Gly Arg
145 150 155 160

Asn Trp Glu Gly Phe Gly Ser Asp Pro Val Leu Gln Ala Lys Ala Ala
165 170 175

Ala Leu Thr Ile Lys Gly Val Gln Glu Gln Gly Ile Ile Ala Thr Ile
180 185 190

Lys His Leu Ile Gly Asn Glu Gln Glu Met Tyr Arg Met Tyr Asn Pro
195 200 205

Phe Gln Pro Gly Tyr Ser Ala Asn Ile Asp Asp Arg Thr Leu His Glu
210 215 220

Leu Tyr Leu Trp Pro Phe Ala Glu Ser Val His Ala Gly Val Gly Ser
225 230 235 240

Ala Met Thr Ala Tyr Asn Ala Val Asn Gly Ser Ala Cys Ser Gln His
245 250 255

Ser Tyr Leu Ile Asn Gly Ile Leu Lys Asp Glu Leu Gly Phe Gln Gly
260 265 270

Phe Val Met Ser Asp Trp Leu Ser His Ile Ser Gly Val Asp Ser Ala
275 280 285

Leu Ala Gly Leu Asp Met Asn Met Pro Gly Asp Thr Asn Ile Pro Leu
290 295 300

Phe Gly Phe Ser Asn Trp His Tyr Glu Leu Ser Arg Ser Val Leu Asn
305 310 315 320

Gly Ser Val Pro Leu Asp Arg Leu Asn Asp Met Val Thr Arg Ile Val
325 330 335

Ala Thr Trp Tyr Lys Phe Gly Gln Asp Arg Asp His Pro Arg Pro Asn
340 345 350

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

355	360	365
Leu Phe Ser Pro Lys Gly Gln Val Asn Trp Phe Val Asn Val Gln Ala		
370	375	380
Asp His Tyr Leu Ile Ala Arg Glu Val Ala Gln Asp Ala Ile Thr Leu		
385	390	395
400		
Leu Lys Asn Asn Gly Ser Phe Leu Pro Leu Thr Thr Ser Gln Ser Leu		
405	410	415
His Val Phe Gly Thr Ala Ala Gln Val Asn Pro Asp Gly Pro Asn Ala		
420	425	430
Cys Met Asn Arg Ala Cys Asn Lys Gly Thr Leu Gly Met Gly Trp Gly		
435	440	445
Ser Gly Val Ala Asp Tyr Pro Tyr Leu Asp Asp Pro Ile Ser Ala Ile		
450	455	460
Arg Lys Arg Val Pro Asp Val Lys Phe Phe Asn Thr Asp Gly Phe Pro		
465	470	475
480		
Trp Phe His Pro Thr Pro Ser Pro Asp Asp Val Ala Ile Val Phe Ile		
485	490	495
Thr Ser Asp Ala Gly Glu Asn Ser Phe Thr Val Glu Gly Asn Asn Gly		
500	505	510
Asp Arg Asn Ser Ala Lys Leu Ala Ala Trp His Asn Gly Asp Glu Leu		
515	520	525
Val Arg Lys Thr Ala Glu Lys Tyr Asn Asn Val Ile Val Val Ala Gln		
530	535	540
Thr Val Gly Pro Leu Asp Leu Glu Ser Trp Ile Asp Asn Pro Arg Val		
545	550	555
560		
Lys Gly Val Leu Phe Gln His Leu Pro Gly Gln Glu Ala Gly Glu Ser		
565	570	575
Leu Ala Asn Ile Leu Phe Gly Asp Val Ser Pro Ser Gly His Leu Pro		
580	585	590
Tyr Ser Ile Thr Lys Arg Ala Asn Asp Phe Pro Asp Ser Ile Ala Asn		
595	600	605

Leu Arg Gly Phe Ala Phe Gly Gln Val Gln Asp Thr Tyr Ser Glu Gly
 610 615 620
 Leu Tyr Ile Asp Tyr Arg Trp Leu Asn Lys Glu Lys Ile Arg Pro Arg
 625 630 635 640

 Phe Ala Phe Gly His Gly Leu Ser Tyr Thr Asn Phe Ser Phe Asp Ala
 645 650 655
 Thr Ile Glu Ser Val Thr Pro Leu Ser Leu Val Pro Pro Ala Arg Ala
 660 665 670
 Pro Lys Gly Ser Thr Pro Val Tyr Ser Thr Glu Ile Pro Pro Ala Ser
 675 680 685
 Glu Ala Tyr Trp Pro Glu Gly Phe Asn Arg Ile Trp Arg Tyr Leu Tyr
 690 695 700

 Ser Trp Leu Asn Lys Asn Asp Ala Asp Asn Ala Tyr Ala Val Gly Ile
 705 710 715 720
 Ala Gly Val Lys Lys Tyr Asn Tyr Pro Ala Gly Tyr Ser Thr Ala Gln
 725 730 735
 Lys Pro Gly Pro Ala Ala Gly Gly Gly Glu Gly Gly Asn Pro Ala Leu
 740 745 750
 Trp Asp Ile Ala Phe Arg Val Pro Val Thr Val Lys Asn Thr Gly Asp
 755 760 765

 Thr Phe Ser Gly Arg Ala Ser Val Gln Ala Tyr Val Gln Tyr Pro Glu
 770 775 780
 Gly Ile Pro Tyr Asp Thr Pro Val Val Gln Leu Arg Asp Phe Glu Lys
 785 790 795 800
 Thr Arg Val Leu Ala Pro Gly Glu Glu Glu Thr Val Thr Val Glu Leu
 805 810 815
 Thr Arg Lys Asp Leu Ser Val Trp Asp Thr Glu Leu Gln Asn Trp Val
 820 825 830

 Val Pro Gly Val Gly Gly Lys Arg Tyr Thr Val Trp Ile Gly Glu Ala
 835 840 845
 Ser Asp Arg Leu Phe Thr Ala Cys Tyr Thr Asp Thr Gly Val Cys Glu

850	855	860	
Gly Gly Arg Val Pro Pro Val			
865	870		
<210> 81			
<211> 2799			
<212> DNA			
<213> Podospora anserina			
<400> 81			
atggcataacc gctcattagt cttgggcgcc ttcgcctcca cctctcttgc cgccagcgtc	60		
gtgacgcctc gagatcctgt tccgcctgga ttcgtcgtg ccccatacta tccagcgctt	120		
catggaggat gggtcgcttc gtgggaagag gcttacagca aggccgaagc cttggtctcg	180		
cagatgacct tggctgaaaa gaccaacatc acctcaggca ttggcatctt tatgggtgag	240		
ttattaacca gacatggctt atataaaagc acaagagact gactgacatg tgaatagggt	300		
cagtgccacc accctaata gacgtttttc tgattttgac taacacatga tacgctagtc	360		
catgcgtagg aaatactgga agcgcagaaa gattgggggt cccgcgcatg tgtcttcagg	420		
actctgcgtt ggggtgtgtc tcggctgaca acgtcactgc gtttcctgct ggcataacca	480		
ctggtgcaac gtttgacaag aagctgatct atgctcgtgg tgttgctatt ggtgaagagc	540		
atcgcggaac gggcacaaat gtctatctgg gtccttccgt aggcctctt gggcggaagc	600		
ctttgggtgg ccgcaactgg gagggctttg gatctgacct agttcttcaa gccaaggctg	660		
ctgcctgac gatcaagggc gttcaggaac aaggcatcat tgctactatc aagcatctga	720		
tcggcaacga gcaggagatg tatagaatgt acaaccctt ccagcctgga tatagcgcca	780		
atattggtga gtggactctt gctctttgac ggactaaaag gctgactccc cacagatgat	840		
cggactctgc acgagctcta cctgtggccc ttgtccgaat ccgtccatgc cgggtgtggg	900		
tcggcaatga cagcttacaa tgctgtaaac gggctcgtt gctctcagca cagctatctc	960		
atcaacggta ttttgaagga tgagcttggg ttccagggtc tcgtcatgtc tgactggctg	1020		
tcccacatct ccggagtcca ctccgcgttg gcaggtctcg acatgaacat gccaggtgac	1080		
accaacattc ccctatttgg tttagcaac tggcaactat agctcagcag atcggttctc	1140		
aacgggtctg tgctcttga cagactgaac gacatggtca ccagaatcgt cgcgacatgg	1200		
tacaagttag gtcaggatag ggaccacca aggcctaact tctcgtcaa caccctgac	1260		
cgtgacggtc tgctttatcc tgcagctctc ttctcccca agggtcaggt gaactggttt	1320		
gtcaatgttc aggtgatca ttatttgatc gccagagagg tcgcccagga tgccatcacc	1380		

cttctcaaga acaatgggag ctctcttccc ctgacgactt cgcagtctct ccatgtcttc 1440
 ggtactgctg cccaggtcaa ccccgatggg cccaacgctt gcatgaaccg cgcctgcaac 1500
 aaaggaacac ttggcatggg ctggggttct ggtgttgccg attatcctta cttgatgac 1560
 ccgatctcgg ctatcaggaa gcgggttccc gacgtcaagt tcttcaacac cgacggcttc 1620
 ccttggttcc accctacacc gtgccccgat gacgttgcca tcgtgttcat cacctccgat 1680
 gctggagaga actcgttcac tgttgagggc aacaacggtg atcgcaacag tgccaagctg 1740
 gctgcgtggc ataacggtga cgagctggtc aggaagactg ccgagaagta caacaacgtt 1800

attgtgtag ctcaaaccgt cggccctctc gatctcgaat cctggatcga caaccctcgc 1860
 gtcaagggcg tctgtttca gcaccttccc ggtcaagaag cgggcgagtc gttggccaac 1920
 attctctttg gcgatgtctc ccctagcggc caccttcctt actccatcac caagcgcgcc 1980
 aacgacttcc ccgacagcat cgccaacctc cgtggctttg cctttggtca ggtccaggac 2040
 acgtacagcg agggcctgta cattgactac cgttggtca acaaggagaa gatcaggccc 2100
 cgctttgctt ttggccacgg tctcagctac accaacttct cgtttgatgc caccatcgag 2160
 tctgtcactc cactgtctct gggttctctt gccctgccc ccaagggctc aacgccggtg 2220

tactcgaccg aaatcccccc cgcctcagag gcgtactggc cggaagggtt caacaggatc 2280
 tggcggtacc tctactctg gctcaacaag aacgacgagg ataacgcta cgctgttggt 2340
 atcgccgggg tgaagaagta taactatccc gctgggtaca gcaccgcca gaagcccgtt 2400
 cccgcagccg gtggcgggga ggggggtaat cctgcgcttt gggatattgc tttccgtgtc 2460
 ccagttacgg tcaagaacac tggggatacg ttctcgggac gggcttcggt gcaggcttat 2520
 gttagtatac ctgaggggat cccgtatgat acgctgttg tgcagctgag ggactttgag 2580
 aagacgaggg ttttggctcc gggggaggag gagacggtga cggttgagct gaccaggaag 2640

gacttgagcg tgtgggacac ggagctgcag aactgggttg tgccgggggt tggggggaag 2700
 aggtatacgg tttagattgg ggaggcgagc gataggttgt ttacggcttg ttatacggat 2760
 acgggggttt gtgagggggg gaggtgccc cctgtttaa 2799

<210> 82
 <211> 3193
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic chimeric Fv3c/Bgl3 sequence
 <400> 82

atgaagctga attgggtcgc cgcagccctg tctatagggt ctgctggcac tgacagcgca 60

gttgctcttg ctctgcagt tccagacact ttggctgggtg taaaggtcag tttttttca 120

ccatttcttc gtctaatttc agccttgttg ccataatgcc ctgttctgct cggacgccac 180

gcaccagatc gcgatcattt cctcccttgc agccttgggt cctcttacga tcttccctcc 240

gcaattatca ggcctcttag tctacacaaa aacccccgag acagtcttcc attgagtttg 300

tcgacatcaa gttgcttctc aactgtgcat ttgcgtggct gtctacttct gcctctagac 360

aaccaaattc gggcgcaatt gaccgtcaa accttgttca aataaccttt ttatttcgag 420

acgcacattt ataaatatgc gcctttcaat aataccgact ttatgcgcgg cggctgctgt 480

ggcgggtgat cagaaagctg acgctcaaaa ggtgttcacg agagatacac tcgcatactc 540

gccgcctcat tatccttcac catggatgga ccctaattgt gttggctggg aggaagctta 600

cgccaaagcc aagagctttg tgtcccaact cactctcatg gaaaaggtca acttgaccac 660

tggtgttggg taagcagctc ctgtcaaaca ggttatctca atccctcag ctaacaactt 720

ctcagatggc aaggcgaacg ctgtgttaga aacgtgggat caattcctcg tctcggtatg 780

cgaggtctct gtctccagga tggctctctt ggaattcgtc tgtccgacta caacagcgt 840

tttcccgctg gcaccacagc tgggtcttct tggagcaagt ctctctggta tgagagaggt 900

ctcctgatgg gcactgagtt caaggagaag ggtatcgata tcgctcttgg tctgtctact 960

ggacctcttg gtcgcactgc tgcctgtgga cgaaactggg aaggtttcac cgttgatcct 1020

tatatggctg gccacgccat ggccgaggcc gtcaagggtg ttcaagacgc aggtgtcatt 1080

gcttgtgcta agcattacat cgcaaacgag cagggtaacg cacttgacg atttgaggaa 1140

ttgacagaga actgacctc ttgtagagca ctccgacag agtggcgagg tccagtcccg 1200

caagtacaac atctccgagt ctctctctc caacctggat gacaagacta tgcacgagct 1260

ctacgcctgg ccttctgctg acgccgtccg cgccggcgtc ggttccgtca tgtgtcgtgta 1320

caaccagatc aacaactcgt acggttgcca gaactccaag ctctcaacg gtatctctca 1380

ggacgagatg ggcttccagg gtttctcat gagcgattgg gcggcccagc ataccggtgc 1440

cgcttctgcc gtcgtgtgct tcgatatgag catgcctggg gacactgcct tcgacagcgg 1500

atacagcttc tggggcgga aattgactct ggtgtctatc aacggaactg tccccgctg 1560

gcgagttgat gacatggctc tgcgaatcat gtctgccttc ttcaaggttg gaaagacgat 1620

agaggatctt cccgacatca acttctctc ctggaccgcg gacaccttcg gcttctgtga 1680

tacatttgct caagagaacc gcgagcaggt caactttgga gtcaacgtcc agcacgacca 1740

caagagccac atccgtgagg ccgctgccaa gggaagcgtc gtgtctcaaga acaccgggtc 1800

ccttccctc aagaacccaa agttcctcgc tgcattggt gaggacgccg gtcccaaccc 1860
 tgctggaccc aatggttgtg gtgaccgtgg ttgcgataat ggtaccctgg ctatggcttg 1920
 gggctcggga acttcccaat tcccttactt gatcaccccc gatcaagggc tctctaateg 1980
 agctactcaa gacggaactc gatatgagag catcttgacc aacaacgaat gggcttcagt 2040
 acaagctctt gtcagccagc ctaacgtgac cgctatcggt ttcgccaatg ccgactctgg 2100
 tgagggatac attgaagtcg acggaaactt tggatgatgc aagaacctca ccctctggca 2160
 gcaggagagc gagctcatca agaacgtgic gtccatatgc cccaacacca ttgtagtctt 2220

 gcacaccgtc ggcctgtcc tactcgccga ctacgagaag aacccaaca tcaactgcat 2280
 cgtctgggct ggtcttcccg gccaaagatc aggcaatgcc atcgctgac tcctctacgg 2340
 caaggtcagc cctggccgat ctcccttcac ttggggccgc acccgcgaga gctacggtac 2400
 tgaggttctt tatgaggcga acaacggccg tggcgctcct caggatgact tctctgaggg 2460
 tgtcttcac gactaccgtc acttcgaccg acgatctcca agcaccgatg gaaagagctc 2520
 tccaacaac accgtgctc ctctctacga gttcggtcac ggtctatctt ggtcgacgtt 2580
 caagtctcc aacctccaca tccagaagaa caatgtcggc ccatgagcc cgcccaacgg 2640

 caagacgatt gcggctccct ctctgggcag cttcagcaag aaccttaagg actatggctt 2700
 cccaagaac gttcgccgca tcaaggagt tctctacccc tactgagca ccactacctc 2760
 tggcaaggag gcgtcgggtg acgtcacta cggccagact gcgaaggagt tcctccccgc 2820
 cgggtgcctg gacggcagcc ctacgcctcg ctctcgggcc tctggcgaac cggcgggcaa 2880
 ccgccagctg tacgacattc tctacaccgt gacggccacc attaccaaca cgggctcggc 2940
 catggacgac gccgttcccc agctgtacct gagccacggc ggtcccaacg agccgcccac 3000
 ggtgctcgct ggcttcgacc gcatcgagcg cattgtctcc ggccagagcg tcacgttcaa 3060

 ggcagacctg acgcgccgtg acctgtccaa ctgggacacg aagaagcagc agtgggtcat 3120
 taccgactac cccaagactg tgtactggg cagctcctcg cgcgacctgc cgctgagcgc 3180
 ccgctgcca tga 3193
 <210> 83
 <211> 3157
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic Fv3C/Te3A/T. reesei Bgl3 (FAB) chimera sequence
 <400> 83
 atgaagctga attgggtcgc cgcagccctg tctataggtg ctgctggcac tgacagcgca 60

gttgctcttg ctctgcagt tccagacact ttggctgggtg taaaggtcag tttttttca	120
ccatttcttc gtctaatttc agccttgttg ccataatgcc cttgttcgct cggacgccac	180
gcaccagatc gcgatcattt cctcccttgc agccttgggt cctcttacga tcttccctcc	240
gcaattatca ggcctcttag tctacacaaa aacccccgag acagtcttcc attgagtttg	300
tcgacatcaa gttgcttctc aactgtgcat ttgcgtggct gtctacttct gcctctagac	360
aaccaaattc gggcgcaatt gaccgtcaa accttgttca aataaccttt ttatttcgag	420
acgcacattt ataaatatgc gcctttcaat aataccgact ttatgcgcgg cggctgctgt	480
ggcggttgat cagaaagctg acgctcaaaa ggtgttcacg agagatacac tcgcatactc	540
gccgcctcat tatccttcac catggatgga ccctaattgt gttggctggg aggaagctta	600
cgccaaagcc aagagctttg tgtcccaact cactctcatg gaaaaggtca acttgaccac	660
tggtgttggg taagcagctc cttgcaaaaca gggatatctca atccctcag ctaacaactt	720
ctcagatggc aaggcgaacg ctgtgttaga aacgtgggat caattcctcg tctcggtatg	780
cgaggctctc gtctccagga tggctctctt ggaattcgtc tgtccgacta caacagcgt	840
tttcccgctg gcaccacagc tgggtcttct tggagcaagt ctctctggta tgagagaggt	900
ctcctgatgg gcactgagtt caaggagaag ggtatcgata tcgctcttgg tctgtctact	960
ggacctcttg gtcgcactgc tgctgggtgga cgaaactggg aaggtttcac cgttgatcct	1020
tatatggctg gccacgccat ggccgaggcc gtcaagggtg ttcaagacgc aggtgtcatt	1080
gcttgtgcta agcattacat cgcaaacgag cagggtaacg cacttgacg atttgaggaa	1140
ttgacagaga actgacctc ttgtagagca ctccgacag agtggcgagg tccagtcccg	1200
caagtacaac atctccgagt ctctctctc caacctggat gacaagacta tgcacgagct	1260
ctacgcctgg ccttgcgtg acgccgtccg cgccggcgtc ggttccgtca tgtgtcgtgta	1320
caaccagatc aacaactcgt acggttgcca gaactccaag ctctcaacg gtatctcaa	1380
ggacgagatg ggcttccagg gtttcgtcat gagcgattgg gcggcccagc ataccggtgc	1440
cgcttctgcc gtcgctggtc tcgatatgag catgcctggg gacactgcct tcgacagcgg	1500
atacagcttc tggggcgga aattgactct ggctgtcatc aacggaactg tccccgctg	1560
gcgagttgat gacatggctc tcgcaatcat gtctgccttc ttcaaggttg gaaagacgat	1620
agaggatctt cccgacatca acttctctc ctggaccgc gacaccttcg gcttctgtca	1680
tacatttgct caagagaacc gcgagcaggt caactttgga gtcaacgtcc agcacgacca	1740
caagagccac atccgtgagg ccgctgcca gggaagcgtc gtgtcaaga acaccgggtc	1800

ccttccctc aagaacccaa agttcctcgc tgcattggg gaggacgccg gtcccaaccc 1860
 tgctggaccc aatggttgtg gtgaccgtgg ttgcgataat ggtaccctgg ctatggcttg 1920
 gggctcggga acttcccaat tcccttactt gatcaccccc gatcaagggc tctctaateg 1980
 agctactcaa gacggaactc gatatgagag catcttgacc aacaacgaat gggcttcagt 2040
 acaagctctt gtcagccagc ctaacgtgac cgctatcggt ttcgccaatg ccgactctgg 2100
 tgagggatac attgaagtcg acggaaactt tggatgatgc aagaacctca ccctctggca 2160
 gcaggggagac gagctcatca agaacgtgtc gtccatatgc cccaacacca ttgtagtctt 2220

 gcacaccgtc ggcctgtcc tactcgcga ctacgagaag aacccaaca tcaactgcat 2280
 cgtctgggct ggtcttcccg gccaaagatc aggcaatgcc atcgtgatc tcctctacgg 2340
 caaggtcagc cctggccgat ctcccttcac ttggggccgc acccgcgaga gctacggtag 2400
 tgaggttctt tatgaggcga acaacggccg tggcgtcctt caggatgact tctctgaggg 2460
 tgtcttcac gactaccgtc acttcgacaa gtacaacatc acgcctatct acgagttcgg 2520
 tcacggctta tcttggtcga cgttcaagtt ctccaacctc cacatccaga agaacaatgt 2580
 cggccccatg agcccccca acggcaagac gattgcgggt ccctctctgg gcaacttcag 2640

 caagaacctt aaggactatg gcttcccaa gaacgttcgc cgcacaaagg agtttatcta 2700
 cccctacctg aacaccacta cctctggcaa ggaggcgtcg ggtgacgtc actacggcca 2760
 gactgcgaag gatttctcc ccgccggtgc cctggacggc agccctcagc ctgctctgc 2820
 ggcctctggc gaacccggcg gcaaccgcca gctgtacgac attctctaca ccgtgacggc 2880
 caccattacc aacacgggct cggatcatga cgacgccgtt cccagctgt acctgagcca 2940
 cggcgggtccc aacgagccgc ccaaggtgct gctgggttc gaccgcatcg agcgattgc 3000
 tccccgccag agcgtcacgt tcaaggcaga cctgacgcgc cgtgacctgt ccaactggga 3060

 cacgaagaag cagcagtggtg tcattaccga ctacccaag actgtgtacg tgggcagctc 3120
 ctgcgcgac ctgccgtga gcgccgcct gccatga 3157

 <210> 84
 <211> 19
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic GH61 endoglucanase family motif
 <220><221> MISC_FEATURE
 <222> (1)..(1)
 <223> Xaa can be Ile, Leu, Met or Val
 <220><221> misc_feature

<222> (3)..(6)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (8)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature

<222> (11)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (13)..(13)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (14)..(14)

<223> Xaa can be Glu or Gln

<220><221> misc_feature

<222> (15)..(18)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (19)..(19)

<223> Xaa can be His, Asn or Gln

<400> 84

Xaa Pro Xaa Xaa Xaa Xaa Gly Xaa Tyr Xaa Xaa Arg Xaa Xaa Xaa Xaa

1 5 10 15

Xaa Xaa Xaa

<210> 85

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<220><221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature

<222> (3)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (9)..(9)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (11)..(11)

<223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (14)..(14)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (15)..(15)

<223> Xaa can be Glu or Gln

<220><221> misc_feature

<222> (16)..(19)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (20)..(20)

<223> Xaa can be His, Asn or Gln

<400> 85

Xaa Pro Xaa Xaa Xaa Xaa Gly Xaa Tyr Xaa Xaa Arg Xaa Xaa Xaa

1 5 10 15

Xaa Xaa Xaa Xaa

20

<210> 86

<211> 19
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic GH61 endoglucanase family motif
 <220><221> MISC_FEATURE
 <222> (1)..(1)
 <223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature
 <222> (3)..(6)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (8)..(8)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> MISC_FEATURE
 <222> (10)..(10)
 <223> Xaa can be Ile, Leu, Met or Val
 <220><221> misc_feature
 <222> (11)..(11)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (13)..(13)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> MISC_FEATURE

<222> (14)..(14)
 <223> Xaa can be Glu or Gln
 <220><221> misc_feature
 <222> (15)..(17)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> MISC_FEATURE
 <222> (19)..(19)
 <223> Xaa can be His, Asn or Gln
 <400> 86

Xaa Pro Xaa Xaa Xaa Xaa Gly Xaa Tyr Xaa Xaa Arg Xaa Xaa Xaa Xaa

1 5 10 15

Xaa Ala Xaa

<210> 87

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<220><221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature

<222> (3)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (9)..(9)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (11)..(11)

<223> Xaa can be Ile, Leu, Met or Val

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (14)..(14)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (15)..(15)

<223> Xaa can be Glu or Gln

<220><221> misc_feature

<222> (16)..(18)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (20)..(20)

<223> Xaa can be His, Asn or Gln

<400> 87

Xaa Pro Xaa Xaa Xaa Xaa Xaa Gly Xaa Tyr Xaa Xaa Arg Xaa Xaa Xaa

1 5 10 15

Xaa Xaa Ala Xaa

20

<210> 88

<

211> 4

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<220><221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa can be Phe or Trp

<220><221> MISC_FEATURE

<222> (2)..(2)

<223> Xaa can be Phe or Thr

<220><221> MISC_FEATURE

<222> (4)..(4)

<223> Xaa can be Ala, Ile or Val

<400> 88

Xaa Xaa Lys Xaa

1

<210> 89

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<

220><221> misc_feature

<222> (2)..(3)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (6)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (9)..(9)

<223> Xaa can be Tyr or Trp

<220><221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa can be Ala, Ile, Leu, Met or Val

<400> 89

His Xaa Xaa Gly Pro Xaa Xaa Xaa Xaa Xaa

1 5 10

<210> 90

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (5)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (8)..(8)

<223> Xaa can be Tyr or Trp

<220><221> MISC_FEATURE

<222> (9)..(9)

<223> Xaa can be Ala, Ile, Leu, Met or Val

<400> 90

His Xaa Gly Pro Xaa Xaa Xaa Xaa Xaa

1 5

<

210> 91

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> synthetic GH61 endoglucanase family motif

<220><221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa can be Glu or Gln

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (4)..(5)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (8)..(8)

<223> Xaa can be Glu, His, Gln or Asn

<220><221> MISC_FEATURE

<222> (9)..(9)

<223> Xaa can be Phe, Ile, Leu or Val

<220><221> misc_feature

<222> (10)..(10)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (11)..(11)

<223> Xaa can be Ile, Leu or Val

<400> 91

Xaa Xaa Tyr Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa

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

<210> 92

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223>	synthetic primer	
<400>	92	
caccatgaga tatagaacag ctgccgct		28
<210>	93	
<211>	40	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	synthetic primer	
<400>	93	
cgaccgccct gcggagtctt gcccagtggc cccgcgacag		40
<210>	94	
<211>	40	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	synthetic primer	
<400>	94	
ctgtcgcggg accactgggc aagactccgc agggcggtcg		40
<210>	95	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	synthetic primer	
<400>	95	
cctacgctac cgacagagtg		20
<210>	96	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	synthetic primer	
<400>	96	
gtctagactg gaaacgcaac		20
<210>	97	
<211>	21	

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 97
 gagttgtgaa gtcggtaatc c 21
 <210> 98
 <211> 35

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 98
 caccatgaaa gcaaacgtca tcttgtgcct cctgg 35
 <210> 99
 <211> 43

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 99
 ctattgtaag atgccaacaa tgctgttata tgccggcttg ggg 43
 <210> 100
 <211> 21

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 100
 gagttgtgaa gtcggtaatc c 21

<210> 101
 <211> 18

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 101
 cacgaagagc ggcgattc 18

<210> 102	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 102	
cacccatgct gctcaatctt cag	23
<210> 103	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 103	
ttacgcagac ttgggtctt gag	23
<210> 104	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 104	
gcttgagtgt atcgtgtaag	20
<210> 105	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 105	
gcaacggcaa agccccactt c	21
<210> 106	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 106	

gtagcggccg cctcatctca tctcatccat cc	32
<210> 107	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 107	
caccatgcag ctcaagtttc tgtc	24
<210> 108	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 108	
ggttactagt caactgcccg ttctgtagcg ag	32
<210> 109	
<211> 29	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 109	
catgcgatcg cgacgttttg gtcaggtcg	29
<210> 110	
<211> 40	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 110	
gacagaaact tgagctgcat ggtgtgggac aacaagaagg	40
<210> 111	
<211> 29	
<212> DNA	
<213> Artificial Sequence	

<220><223> synthetic primer
 <400> 111
 caccatgggtt cgcttcagtt caatcctag 29
 <210> 112
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 112
 gtggctagaa gatatccaac ac 22
 <210> 113
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 113
 catgcgatcg cgacgttttg gtcaggctcg 29
 <210> 114
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 114
 gaactgaagc gaaccatggt gtgggacaac aagaaggac 39
 <210> 115
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 115
 gtagttatgc gcatgctaga c 21
 <210> 116
 <211> 24

<212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 116
 caccatgaag ctgaattggg tcgc 24
 <210> 117
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 117
 ttactccaac ttggcgctg 19
 <210> 118
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 118
 aagccaagag ctttgtgtcc 20

 <210> 119
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 119
 tatgcacgag ctctacgcct 20
 <210> 120
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 120
 atggtaccct ggctatggct 20
 <210> 121

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 121
 cggtcacggt ctatcttggt 20

<210> 122
 <211> 45
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 122
 gctagcatgg atgttttccc agtcacgacg ttgtaaaacg acggc 45

<210> 123
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 123
 ggaggttggga gaacttgaac gtcgaccaag atagaccgtg accgaactcg tag 53

<210> 124
 <211> 43
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 124
 tgccaggaaa cagctatgac catgtaatac gactcactat agg 43

<210> 125
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220><223> synthetic primer
 <400> 125

ctacgagttc gggtcacggtc tatcttggtc gacgttcaag ttctccaacc tcc 53

<210> 126

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 126

taagctcggg ccccaaataa tgattttatt ttgactgata gt 42

<210> 127

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 127

gggatatcag ctggatggca aataatgatt ttattttgac tgata 45

<210> 128

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 128

gagttgtgaa gtcggtaatc ccgctg 26

<210> 129

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 129

cctgcacgag ggcatcaage tcactaaccg 30

<210> 130

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 130	
cggaatgagc tagtaggcaa agtcagc	27
<210> 131	
<211> 70	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 131	
ctccttgatg cggcgaacgt tcttggggaa gccatagtc ttaaggttct tgctgaagtt	60
gcccagagag	70
<210> 132	
<211> 65	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 132	
ggcttcccca agaacgttcg ccgcacaaag gagtttatct acccctacct gaacaccact	60
acctc	65
<210> 133	
<211> 27	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 133	
gatacacgaa gagcggcgat tctacgg	27
<210> 134	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 134	
cacatgaag ctgaattggg tcgc	24
<210> 135	

<211> 886

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric Fv3c/Te3A/T. reesei Bgl3 (FAB) sequence

<400> 135

Met Lys Leu Asn Trp Val Ala Ala Ala Leu Ser Ile Gly Ala Ala Gly

1 5 10 15

Thr Asp Ser Ala Val Ala Leu Ala Ser Ala Val Pro Asp Thr Leu Ala

20 25 30

Gly Val Lys Lys Ala Asp Ala Gln Lys Val Val Thr Arg Asp Thr Leu

35 40 45

Ala Tyr Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp Pro Asn Ala

50 55 60

Val Gly Trp Glu Glu Ala Tyr Ala Lys Ala Lys Ser Phe Val Ser Gln

65 70 75 80

Leu Thr Leu Met Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gln

85 90 95

Gly Glu Arg Cys Val Gly Asn Val Gly Ser Ile Pro Arg Leu Gly Met

100 105 110

Arg Gly Leu Cys Leu Gln Asp Gly Pro Leu Gly Ile Arg Leu Ser Asp

115 120 125

Tyr Asn Ser Ala Phe Pro Ala Gly Thr Thr Ala Gly Ala Ser Trp Ser

130 135 140

Lys Ser Leu Trp Tyr Glu Arg Gly Leu Leu Met Gly Thr Glu Phe Lys

145 150 155 160

Glu Lys Gly Ile Asp Ile Ala Leu Gly Pro Ala Thr Gly Pro Leu Gly

165 170 175

Arg Thr Ala Ala Gly Gly Arg Asn Trp Glu Gly Phe Thr Val Asp Pro

180 185 190

Tyr Met Ala Gly His Ala Met Ala Glu Ala Val Lys Gly Ile Gln Asp

195 200 205

Ala Gly Val Ile Ala Cys Ala Lys His Tyr Ile Ala Asn Glu Gln Glu
210 215 220

His Phe Arg Gln Ser Gly Glu Val Gln Ser Arg Lys Tyr Asn Ile Ser
225 230 235 240

Glu Ser Leu Ser Ser Asn Leu Asp Asp Lys Thr Met His Glu Leu Tyr
245 250 255

Ala Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met
260 265 270

Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Gln Asn Ser Lys
275 280 285

Leu Leu Asn Gly Ile Leu Lys Asp Glu Met Gly Phe Gln Gly Phe Val
290 295 300

Met Ser Asp Trp Ala Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala
305 310 315 320

Gly Leu Asp Met Ser Met Pro Gly Asp Thr Ala Phe Asp Ser Gly Tyr
325 330 335

Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Ile Asn Gly Thr Val
340 345 350

Pro Ala Trp Arg Val Asp Asp Met Ala Leu Arg Ile Met Ser Ala Phe
355 360 365

Phe Lys Val Gly Lys Thr Ile Glu Asp Leu Pro Asp Ile Asn Phe Ser
370 375 380

Ser Trp Thr Arg Asp Thr Phe Gly Phe Val His Thr Phe Ala Gln Glu
385 390 395 400

Asn Arg Glu Gln Val Asn Phe Gly Val Asn Val Gln His Asp His Lys
405 410 415

Ser His Ile Arg Glu Ala Ala Ala Lys Gly Ser Val Val Leu Lys Asn
420 425 430

Thr Gly Ser Leu Pro Leu Lys Asn Pro Lys Phe Leu Ala Val Ile Gly
435 440 445

Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys Gly Asp Arg

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

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

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

 Ala Asn Asn Gly Arg Gly Ala Pro Gln Asp Asp Phe Ser Glu Gly Val
 645 650 655
 Phe Ile Asp Tyr Arg His Phe Asp Lys Tyr Asn Ile Thr Pro Ile Tyr
 660 665 670
 Glu Phe Gly His Gly Leu Ser Trp Ser Thr Phe Lys Phe Ser Asn Leu
 675 680 685
 His Ile Gln Lys Asn Asn Val Gly Pro Met Ser Pro Pro Asn Gly Lys
 690 695 700

Thr Ile Ala Ala Pro Ser Leu Gly Asn Phe Ser Lys Asn Leu Lys Asp
 705 710 715 720
 Tyr Gly Phe Pro Lys Asn Val Arg Arg Ile Lys Glu Phe Ile Tyr Pro
 725 730 735
 Tyr Leu Asn Thr Thr Thr Ser Gly Lys Glu Ala Ser Gly Asp Ala His
 740 745 750
 Tyr Gly Gln Thr Ala Lys Glu Phe Leu Pro Ala Gly Ala Leu Asp Gly
 755 760 765

 Ser Pro Gln Pro Arg Ser Ala Ala Ser Gly Glu Pro Gly Gly Asn Arg
 770 775 780
 Gln Leu Tyr Asp Ile Leu Tyr Thr Val Thr Ala Thr Ile Thr Asn Thr
 785 790 795 800
 Gly Ser Val Met Asp Asp Ala Val Pro Gln Leu Tyr Leu Ser His Gly
 805 810 815
 Gly Pro Asn Glu Pro Pro Lys Val Leu Arg Gly Phe Asp Arg Ile Glu
 820 825 830

 Arg Ile Ala Pro Gly Gln Ser Val Thr Phe Lys Ala Asp Leu Thr Arg
 835 840 845
 Arg Asp Leu Ser Asn Trp Asp Thr Lys Lys Gln Gln Trp Val Ile Thr
 850 855 860
 Asp Tyr Pro Lys Thr Val Tyr Val Gly Ser Ser Ser Arg Asp Leu Pro
 865 870 875 880
 Leu Ser Ala Arg Leu Pro
 885

<210> 136

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (6)..(6)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (17)..(17)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (21)..(21)

<223> Xaa can be any naturally occurring amino acid

<400> 136

Ala Xaa Ser Pro Pro Xaa Tyr Pro Ser Pro Trp Met Asp Pro Xaa Ala

1 5 10 15

Xaa Gly Trp Glu Xaa Ala Tyr

20

<210> 137

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (11)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (23)..(23)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (26)..(26)

<223> Xaa can be any naturally occurring amino acid

<400> 137

Ala Lys Xaa Phe Val Ser Xaa Xaa Thr Leu Xaa Glu Lys Val Asn Leu

1 5 10 15

Thr Thr Gly Val Gly Trp Xaa Gly Glu Xaa Cys Val Gly Asn Val Gly

20 25 30

<210> 138

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (10)..(10)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (17)..(17)

<223> Xaa can be any naturally occurring amino acid

<400> 138

Pro Arg Xaa Gly Met Arg Xaa Leu Cys Xaa Gln Asp Gly Pro Leu Gly

1 5 10 15

Xaa Arg

<210> 139

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (6)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (9)..(9)

<223>

> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<400> 139

Tyr Asn Ser Ala Phe Xaa Xaa Gly Xaa Thr Ala Xaa Ala Ser Trp Ser

1 5 10 15

<210> 140

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (9)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (17)..(17)

<223> Xaa can be any naturally occurring amino acid

<400> 140

Gly Xaa Ile Ala Cys Ala Lys His Xaa Xaa Xaa Asn Glu Gln Glu His

1 5 10 15

Xaa Arg Gln

<210> 141
 <211> 27
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature

<222> (5)..(5)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (10)..(10)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (13)..(13)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (15)..(15)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (19)..(19)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222>

(23)..(23)
 <223> Xaa can be any naturally occurring amino acid
 <400> 141
 Leu Ser Ser Asn Xaa Asp Asp Lys Thr Xaa His Glu Xaa Tyr Xaa Trp
 1 5 10 15
 Pro Phe Xaa Asp Ala Val Xaa Ala Gly Val Gly
 20 25

<210> 142
 <211> 21
 <212> PRT
 <213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (5)..(5)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (19)..(19)

<223> Xaa can be any naturally occurring amino acid

<400> 142

Met Cys Ser Tyr Xaa Gln Xaa Asn Asn Ser Tyr Xaa Cys Gln Asn Ser

1 5 10 15

Lys Leu Xaa Asn Gly

20

<210> 143

<211>

32

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (11)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (17)..(17)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (19)..(19)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (27)..(27)

<223> Xaa can be any naturally occurring amino acid

<400> 143

Gly Phe Gln Gly Phe Val Met Ser Asp Trp Xaa Ala Gln His Xaa Gly

1 5 10 15

Xaa Ala Xaa Ala Val Ala Gly Leu Asp Met Xaa Met Pro Gly Asp Thr

20 25 30

<210> 144

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221>

> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (13)..(13)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (16)..(16)

<223> Xaa can be any naturally occurring amino acid

<400> 144

Asn Leu Thr Leu Ala Val Xaa Asn Gly Thr Val Pro Xaa Trp Arg Xaa

1 5 10 15

Asp Asp Met

<210> 145

<211> 26

<212> PRT

<213>

Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (5)..(5)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (13)..(13)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222

> (22)..(22)

<223> Xaa can be any naturally occurring amino acid

<400> 145

Pro Xaa Phe Leu Xaa Val Xaa Gly Glu Asp Ala Gly Xaa Asn Pro Ala

1 5 10 15

Gly Pro Asn Gly Cys Xaa Asp Arg Gly Cys

 20 25

<210> 146

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (6)..(6)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<400> 146

Gly Thr Leu Ala Met Xaa Trp Gly Ser Gly Thr Xaa Phe Pro Tyr Leu

1 5 10 15

<210> 147

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (7)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (20)..(20)

<223> Xaa can be any naturally occurring amino acid

<400> 147

Ala Ile Val Phe Ala Asn Xaa Xaa Ser Gly Glu Gly Tyr Ile Xaa Val

1 5 10 15

Asp Gly Asn Xaa Gly Asp Arg Lys Asn Leu Thr Leu Trp

20 25

<210> 148

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<400> 148

Asp Xaa Leu Tyr Gly Lys Xaa Ser Pro Gly Arg Xaa Pro Phe Thr Trp

1 5 10 15

Gly

<210> 149

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(16)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (18)..(18)

<223> Xaa can be any naturally occurring amino acid

<400> 149

Pro Xaa Tyr Glu Phe Gly Xaa Gly Leu Ser Trp Xaa Thr Phe Xaa Xaa

1 5 10 15

Ser Xaa Leu

<210> 150
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature
 <222> (2)..(2)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (5)..(5)
 <223> Xaa can be any naturally occurring amino acid
 <400> 150
 Leu Xaa Asp Tyr Xaa Phe Pro

1 5
 <210> 151
 <211> 15
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature
 <222> (5)..(6)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (9)..(9)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (12)..(12)
 <223> Xaa can be any naturally occurring amino acid
 <400> 151
 Glu Phe Leu Pro Xaa Xaa Ala Leu Xaa Gly Ser Xaa Gln Pro Arg

1 5 10 15
 <210> 152
 <211> 12

<212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature
 <222> (3)..(3)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (8)..(9)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (11)..(11)
 <223> Xaa can be any naturally occurring amino acid
 <400> 152
 Ser Gly Xaa Pro Gly Gly Asn Xaa Xaa Leu Xaa Asp

1 5 10
 <210> 153
 <211> 11
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature
 <222> (4)..(4)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (6)..(6)
 <223> Xaa can be any naturally occurring amino acid
 <400> 153
 Tyr Thr Val Xaa Ala Xaa Ile Thr Asn Thr Gly

1 5 10
 <210> 154
 <211> 16
 <212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (6)..(6)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (8)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (10)..(10)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<400> 154

Val Leu Arg Gly Phe Xaa Arg Xaa Glu Xaa Ile Ala Pro Gly Xaa Ser

1 5 10 15

<210> 155

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric beta-glucosidase motif

<220><221> misc_feature

<222> (10)..(12)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (14)..(14)

<223> Xaa can be any naturally occurring amino acid

<400> 155

Thr Arg Arg Asp Leu Ser Asn Trp Asp Xaa Xaa Xaa Gln Xaa Trp Val

1 5 10 15

Ile Thr Asp

<210> 156

<211> 14
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic chimeric beta-glucosidase motif
 <220><221> misc_feature
 <222> (7)..(7)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (11)..(11)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (13)..(13)
 <223> Xaa can be any naturally occurring amino acid
 <400> 156

Val Gly Ser Ser Ser Arg Xaa Leu Pro Leu Xaa Ala Xaa Leu

1 5 10

<210> 157
 <211> 19
 <212> PRT
 <213> Fusarium verticillioides
 <400> 157

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

1 5 10 15

Ala Pro Leu

<210> 158
 <211> 7
 <212> PRT
 <213> Talaromyces emersonii
 <400> 158

Lys Tyr Asn Ile Thr Pro Ile

1 5

<210> 159
 <211> 898

<212> PRT

<213> Artificial Sequence

<220><223> synthetic chimeric Fv3c/Bgl3 sequence

<400> 159

Met Lys Leu Asn Trp Val Ala Ala Ala Leu Ser Ile Gly Ala Ala Gly

1 5 10 15

Thr Asp Ser Ala Val Ala Leu Ala Ser Ala Val Pro Asp Thr Leu Ala

20 25 30

Gly Val Lys Lys Ala Asp Ala Gln Lys Val Val Thr Arg Asp Thr Leu

35 40 45

Ala Tyr Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp Pro Asn Ala

50 55 60

Val Gly Trp Glu Glu Ala Tyr Ala Lys Ala Lys Ser Phe Val Ser Gln

65 70 75 80

Leu Thr Leu Met Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gln

85 90 95

Gly Glu Arg Cys Val Gly Asn Val Gly Ser Ile Pro Arg Leu Gly Met

100 105 110

Arg Gly Leu Cys Leu Gln Asp Gly Pro Leu Gly Ile Arg Leu Ser Asp

115 120 125

Tyr Asn Ser Ala Phe Pro Ala Gly Thr Thr Ala Gly Ala Ser Trp Ser

130 135 140

Lys Ser Leu Trp Tyr Glu Arg Gly Leu Leu Met Gly Thr Glu Phe Lys

145 150 155 160

Glu Lys Gly Ile Asp Ile Ala Leu Gly Pro Ala Thr Gly Pro Leu Gly

165 170 175

Arg Thr Ala Ala Gly Gly Arg Asn Trp Glu Gly Phe Thr Val Asp Pro

180 185 190

Tyr Met Ala Gly His Ala Met Ala Glu Ala Val Lys Gly Ile Gln Asp

195 200 205

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

210 215 220
 His Phe Arg Gln Ser Gly Glu Val Gln Ser Arg Lys Tyr Asn Ile Ser
 225 230 235 240
 Glu Ser Leu Ser Ser Asn Leu Asp Asp Lys Thr Met His Glu Leu Tyr

 245 250 255
 Ala Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met
 260 265 270
 Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Gln Asn Ser Lys
 275 280 285
 Leu Leu Asn Gly Ile Leu Lys Asp Glu Met Gly Phe Gln Gly Phe Val
 290 295 300
 Met Ser Asp Trp Ala Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala

 305 310 315 320
 Gly Leu Asp Met Ser Met Pro Gly Asp Thr Ala Phe Asp Ser Gly Tyr
 325 330 335
 Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Ile Asn Gly Thr Val
 340 345 350
 Pro Ala Trp Arg Val Asp Asp Met Ala Leu Arg Ile Met Ser Ala Phe
 355 360 365
 Phe Lys Val Gly Lys Thr Ile Glu Asp Leu Pro Asp Ile Asn Phe Ser

 370 375 380
 Ser Trp Thr Arg Asp Thr Phe Gly Phe Val His Thr Phe Ala Gln Glu
 385 390 395 400
 Asn Arg Glu Gln Val Asn Phe Gly Val Asn Val Gln His Asp His Lys
 405 410 415
 Ser His Ile Arg Glu Ala Ala Ala Lys Gly Ser Val Val Leu Lys Asn
 420 425 430
 Thr Gly Ser Leu Pro Leu Lys Asn Pro Lys Phe Leu Ala Val Ile Gly

 435 440 445
 Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys Gly Asp Arg
 450 455 460

Gly Cys Asp Asn Gly Thr Leu Ala Met Ala Trp Gly Ser Gly Thr Ser
 465 470 475 480
 Gln Phe Pro Tyr Leu Ile Thr Pro Asp Gln Gly Leu Ser Asn Arg Ala
 485 490 495
 Thr Gln Asp Gly Thr Arg Tyr Glu Ser Ile Leu Thr Asn Asn Glu Trp
 500 505 510
 Ala Ser Val Gln Ala Leu Val Ser Gln Pro Asn Val Thr Ala Ile Val
 515 520 525
 Phe Ala Asn Ala Asp Ser Gly Glu Gly Tyr Ile Glu Val Asp Gly Asn
 530 535 540
 Phe Gly Asp Arg Lys Asn Leu Thr Leu Trp Gln Gln Gly Asp Glu Leu
 545 550 555 560
 Ile Lys Asn Val Ser Ser Ile Cys Pro Asn Thr Ile Val Val Leu His
 565 570 575
 Thr Val Gly Pro Val Leu Leu Ala Asp Tyr Glu Lys Asn Pro Asn Ile
 580 585 590
 Thr Ala Ile Val Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ala
 595 600 605
 Ile Ala Asp Leu Leu Tyr Gly Lys Val Ser Pro Gly Arg Ser Pro Phe
 610 615 620
 Thr Trp Gly Arg Thr Arg Glu Ser Tyr Gly Thr Glu Val Leu Tyr Glu
 625 630 635 640
 Ala Asn Asn Gly Arg Gly Ala Pro Gln Asp Asp Phe Ser Glu Gly Val
 645 650 655
 Phe Ile Asp Tyr Arg His Phe Asp Arg Arg Ser Pro Ser Thr Asp Gly
 660 665 670
 Lys Ser Ser Pro Asn Asn Thr Ala Ala Pro Leu Tyr Glu Phe Gly His
 675 680 685
 Gly Leu Ser Trp Ser Thr Phe Lys Phe Ser Asn Leu His Ile Gln Lys
 690 695 700
 Asn Asn Val Gly Pro Met Ser Pro Pro Asn Gly Lys Thr Ile Ala Ala

705 710 715 720
Pro Ser Leu Gly Ser Phe Ser Lys Asn Leu Lys Asp Tyr Gly Phe Pro
 725 730 735
Lys Asn Val Arg Arg Ile Lys Glu Phe Ile Tyr Pro Tyr Leu Ser Thr
 740 745 750
Thr Thr Ser Gly Lys Glu Ala Ser Gly Asp Ala His Tyr Gly Gln Thr

 755 760 765
Ala Lys Glu Phe Leu Pro Ala Gly Ala Leu Asp Gly Ser Pro Gln Pro
 770 775 780
Arg Ser Ala Ala Ser Gly Glu Pro Gly Gly Asn Arg Gln Leu Tyr Asp
785 790 795 800
Ile Leu Tyr Thr Val Thr Ala Thr Ile Thr Asn Thr Gly Ser Val Met
 805 810 815
Asp Asp Ala Val Pro Gln Leu Tyr Leu Ser His Gly Gly Pro Asn Glu

 820 825 830
Pro Pro Lys Val Leu Arg Gly Phe Asp Arg Ile Glu Arg Ile Ala Pro
 835 840 845
Gly Gln Ser Val Thr Phe Lys Ala Asp Leu Thr Arg Arg Asp Leu Ser
 850 855 860
Asn Trp Asp Thr Lys Lys Gln Gln Trp Val Ile Thr Asp Tyr Pro Lys
865 870 875 880
Thr Val Tyr Val Gly Ser Ser Ser Arg Asp Leu Pro Leu Ser Ala Arg

 885 890 895
Leu Pro

<210> 160

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> synthetic primer

<400> 160

gatataccgt gaccgaactc gtagataggc gtgatgttgt acttgtcgaa gtgacggtag

60

tcgatgaaga c	71
<210> 161	
<211> 71	
<212> DNA	
<213> Artificial Sequence	
<220><223> synthetic primer	
<400> 161	
gtcttcacgc actaccgtca cttcgacaag tacaacatca cgcctatcta cgagttcggc	60
cacggtctat c	71
<210> 162	
<211> 780	
<212> DNA	
<213> Trichoderma reesei	
<400> 162	
atggtctcct tcacctcct cctcgccggc gtcgccgcca tctcgggcgt cttggccgct	60
cccgccgcgc aggtcgaatc cgtggctgtg gagaagcgcc agacgattca gcccggcacg	120
ggctacaaca acggctactt ctactcgtac tggaacgatg gccacggcgg cgtgacgtac	180
accaatggtc ccggcgggca gttctccgtc aactgggtcca actcgggcaa ctttgcggc	240
ggcaagggat ggcagcccgg gaccaagaac aagtaagact acctactctt accccctttg	300
accaacacag cacaacacaa tacaacacat gtgactacca atcatggaat cggatctaac	360
agctgtgttt taaaaaaaa ggtcatcaac ttctcgggaa gctacaaccc caacggcaac	420
agctacctct ccgtgtacgg ctgggtcccgc aacccctga tcgagtacta catcgtcgag	480
aactttggca cctacaaccc gtccacgggc gccaccaagc tgggcgaggt cacctccgac	540
ggcagcgtct acgacattta ccgcacgcag cgcgtcaacc agccgtccat catcggcacc	600
gccacctttt accagtactg gtccgtccgc cgcaaccacc gctcgagcgg ctccgtcaac	660
acggcgaacc acttcaacgc gtgggctcag caaggcctga cgctcgggac gatggattac	720
cagattgttg ccgtggaggg ttactttagc tctggctctg cttccatcac cgtcagctaa	780
<210> 163	
<211> 2394	
<212> DNA	
<213> Trichoderma reesei	
<400> 163	

atggtgaata acgcagctct tctcgccgcc ctgtcggctc tccgtccac ggccctggcg 60
cagaacaatc aaacatacgc caactactct gctcagggcc agcctgatct ctaccccgag 120
acacttgcca cgctcacact ctcgttcccc gactgcgaac atggccccct caagaacaat 180
ctcgtctgtg acicacgggc cggctatgia gagcgagccc aggcctcat ctcgctcttc 240
accctcgagg agctcattct caacacgcaa aactcgggcc cggcgtgcc tcgcctgggt 300

cttccgaact accaagtctg gaatgagget ctgcacggct tggaccgcgc caacttcgcc 360
accaaggcg gccagttcga atgggcgacc tcgttcccca tgccatcct cactacggcg 420
gccctcaacc gcacattgat ccaccagatt gccgacatca tctcgacca agctcgagca 480
ttcagcaaca gggccggtta cggctcgcac gtctatgcgc caaacgtcaa tggcttccga 540
agccccctct ggggcccgtg ccaggagacg cccggcgaag acgccttttt ctcagctcc 600
gcctatactt acgagtacat caccggcatc cagggtggcg tcgacctga gcacctcaag 660
gttgccgcca cggatgaagca ctttccgga tacgacctcg agaactggaa caaccagtcc 720

cgtctcgggt tcgacgcat cataactcag caggacctct ccgaatacta cactccccag 780
ttctcgtg cggcccgta tgcaaagtca cgcagcttga tgtgcgata caactccgtc 840
aacggcgtgc ccagctgtgc caacagcttc ttctgcaga cgcttttgcg cgagagctgg 900
ggcttccccg aatggggata cgtctcgtcc gattgcgatg ccgtctacaa cgttttcaac 960
cctcatgact acgccagcaa ccagtcgtca gccgccgcca gctcactgcg agccggcacc 1020
gatatcgact gcggtcagac ttaccctgg cacctcaacg agtcctttgt ggccggcgaa 1080
gttccccg cgagatcga gcggtccgtc acccgtctgt acgccaacct cgtccgtctc 1140

ggatacttcg acaagaagaa ccagtaccgc tcgtcgggtt ggaaggatgt cgtcaagact 1200
gatgcctgga acatctcgta cgagctgct gttgagggca tcgtcctgct caagaacgat 1260
ggcactctcc ctctgtcaa gaagtgctgc agcattgctc tgatcgacc atggccaat 1320
gccacaacc aaatgaagg caactactat ggccctgccc catacctcat cagccctctg 1380
gaagctgcta agaaggccgg ctatcacgtc aactttgaac tcggcacaga gatcgccggc 1440
aacagacca ctggctttgc caaggccatt gctgccgcca agaagtcgga tgccatcatc 1500
tacctcgggt gaattgacaa caccattgaa caggaggcg ctgaccgcac ggacattgct 1560

tggcccggtat atcagctgga tctcatcaag cagctcagcg aggtcggcaa acccctgtc 1620
gtcctgcaaa tgggcgggtg tcaggtagac tcctcctgc tcaagagcaa caagaaggtc 1680
aactccctcg tctggggcgg atatcccggc cagtcgggag gcgttgcct cttcgacatt 1740
ctctctggca agcgtgtcc tgccggccga ctggtcacca ctcagtacc ggctgagtat 1800
gttccaat tccccagaa tgacatgaac ctccgaccg atggaaagtc aaacctgga 1860

cagacttaca tctggtacac cggcaaaccg gtctacgagt ttggcagtgg tctcttctac 1920
 accaccttca aggagactct cgccagccac cccaagagcc tcaagttcaa cacctcatcg 1980

atcctctctg ctctcaccg cggatacaact tacagcgagc agattcccg tttcaccttc 2040
 gaggccaaca tcaagaactc gggcaagacg gattcccat atacggccat gctgtttgtt 2100
 cgcacaagca acgttgccc agccccgtac ccgaacaagt ggctcgtcgg attcgaccga 2160
 ctgtccgaca tcaagcctgg tcaactcttc aagctcagca tccccatccc gtgcagtgtt 2220
 ctgccccgtg ttgattctca cggaaaccgg attgtatacc ccggcaagta tgagctagcc 2280
 ttgaacaccg acgagtctgt gaagcttgag ttgagttgg tgggagaaga ggtaacgatt 2340
 gagaactggc cgttgaggga gcaacagatc aaggatgcta cacctgacgc ataa 2394

<210> 164

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> synthetic amino acid sequence motif

<400> 164

Tyr Pro Ser Pro Trp Met Asp Pro

1 5

<210> 165

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> synthetic amino acid sequence motif

<400> 165

Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp

1 5 10

<210> 166

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> synthetic amino acid sequence motif

<220><221> MISC_FEATURE

<222> (3)..(3)

<223> Xaa can be Ile or Val

<220><221> MISC_FEATURE

<222> (5)..(5)

<223> Xaa can be Ile or Val

<400> 166

Lys Gly Xaa Asp Xaa

1 5

<210> 167

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> synthetic amino acid sequence motif

<220><221> misc_feature

<222> (7)..(7)

<223> Xaa can be any naturally occurring amino acid

<400> 167

Cys Gln Asn Ser Lys Leu Xaa Asn Gly

1 5

<210> 168

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> synthetic amino acid sequence motif

<220><221> MISC_FEATURE

<222> (7)..(7)

<223> Xaa can be Leu, Ile or Val

<220><221> MISC_FEATURE

<222> (10)..(10)

<223> Xaa can be Ser or Thr

<220><221> MISC_FEATURE

<222> (11)..(11)

<223> Xaa can be Ile or Val

<220><221> misc_feature

<222> (13)..(13)

<223> Xaa can be any naturally occurring amino acid
 <400> 168
 Asn Leu Thr Leu Ala Val Xaa Asn Gly Xaa Xaa Pro Xaa Trp

1 5 10
 <210> 169
 <211> 8
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic amino acid sequence motif
 <220><221> MISC_FEATURE
 <222> (3)..(3)
 <223> Xaa can be Ser or Thr
 <220><221> misc_feature
 <222> (4)..(4)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> MISC_FEATURE
 <222> (7)..(7)
 <223> Xaa can be Phe or Tyr
 <400> 169
 Ser Trp Xaa Xaa Asp Thr Xaa Gly

1 5
 <210>
 > 170
 <211> 15
 <212> PRT
 <213> Artificial Sequence
 <220><223> synthetic amino acid sequence motif
 <220><221> misc_feature
 <222> (5)..(6)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature
 <222> (9)..(9)
 <223> Xaa can be any naturally occurring amino acid
 <220><221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<400> 170

Glu Phe Leu Pro Xaa Xaa Ala Leu Xaa Gly Ser Xaa Gln Pro Arg

1 5 10 15

<210> 171

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> synthetic loop sequence

<400> 171

Phe Asp Arg Arg Ser Pro Gly

1 5

<210> 172

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> synthetic loop sequence

<220><221> misc_feature

<222> (3)..(3)

<223> Xaa can be Arg or Lys

<400> 172

Phe Asp Xaa Tyr Asn Ile Thr

1 5

<210> 173

<211> 17

<212> PRT

<213> Trichoderma reesei

<400> 173

Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg

1 5 10 15

Ala

<210> 174

<211> 884

<212> PRT

<213> Nectria haematococca

<400> 174

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

1 5 10 15

Val Gly Ser Gln Ala Asp Gln Lys Pro Leu Gln Leu Gly Val Asn Asn

20 25 30

Asn Thr Leu Ala His Ser Pro Pro His Tyr Pro Ser Pro Trp Met Asp

35 40 45

Pro Ala Ala Pro Gly Trp Glu Glu Ala Tyr Leu Lys Ala Lys Asp Phe

50 55 60

Val Ser Gln Leu Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val

65 70 75 80

Gly Trp Met Gly Glu Arg Cys Val Gly Asn Val Gly Ser Leu Pro Arg

85 90 95

Phe Gly Met Arg Gly Leu Cys Met Gln Asp Gly Pro Leu Gly Ile Arg

100 105 110

Leu Ser Asp Tyr Asn Ser Ala Phe Pro Thr Gly Ile Thr Ala Gly Ala

115 120 125

Ser Trp Ser Arg Ala Leu Trp Tyr Gln Arg Gly Leu Leu Met Gly Thr

130 135 140

Glu His Arg Glu Lys Gly Ile Asp Val Ala Leu Gly Pro Ala Thr Gly

145 150 155 160

Pro Leu Gly Arg Thr Pro Thr Gly Gly Arg Asn Trp Glu Gly Phe Ser

165 170 175

Val Asp Pro Tyr Val Ala Gly Val Ala Met Ala Glu Thr Val Ser Gly

180 185 190

Ile Gln Asp Gly Gly Thr Ile Ala Cys Ala Lys His Tyr Ile Gly Asn

195 200 205

Glu Gln Glu His His Arg Gln Ala Pro Glu Ser Ile Gly Arg Gly Tyr

210 215 220
 Asn Ile Thr Glu Ser Leu Ser Ser Asn Val Asp Asp Lys Thr Leu His
 225 230 235 240

 Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Lys Ala Gly Val Gly
 245 250 255
 Ala Ile Met Cys Ser Tyr Gln Gln Leu Asn Asn Ser Tyr Gly Cys Gln
 260 265 270
 Asn Ser Lys Leu Leu Asn Gly Ile Leu Lys Asp Glu Leu Gly Phe Gln
 275 280 285
 Gly Phe Val Met Ser Asp Trp Gln Ala Gln His Ala Gly Ala Ala Thr
 290 295 300

 Ala Val Ala Gly Leu Asp Met Thr Met Pro Gly Asp Thr Leu Phe Asn
 305 310 315 320
 Thr Gly Tyr Ser Phe Trp Gly Gly Asn Leu Thr Leu Ala Val Val Asn
 325 330 335
 Gly Thr Val Pro Asp Trp Arg Ile Asp Asp Met Ala Met Arg Ile Met
 340 345 350
 Ala Ala Phe Phe Lys Val Gly Lys Thr Val Glu Asp Leu Pro Asp Ile
 355 360 365

 Asn Phe Ser Ser Trp Ser Arg Asp Thr Phe Gly Tyr Val Gln Ala Ala
 370 375 380
 Ala Gln Glu Asn Trp Glu Gln Ile Asn Phe Gly Val Asp Val Arg His
 385 390 395 400
 Asp His Ser Glu His Ile Arg Leu Ser Ala Ala Lys Gly Thr Val Leu
 405 410 415
 Leu Lys Asn Ser Gly Ser Leu Pro Leu Lys Lys Pro Lys Phe Leu Ala
 420 425 430

 Val Val Gly Glu Asp Ala Gly Pro Asn Pro Ala Gly Pro Asn Gly Cys
 435 440 445
 Asn Asp Arg Gly Cys Asn Asn Gly Thr Leu Ala Met Ser Trp Gly Ser
 450 455 460

Gly Thr Ala Gln Phe Pro Tyr Leu Val Thr Pro Asp Ser Ala Leu Gln
465 470 475 480

Asn Gln Ala Val Leu Asp Gly Thr Arg Tyr Glu Ser Val Leu Arg Asn
485 490 495

Asn Gln Trp Glu Gln Thr Arg Ser Leu Ile Ser Gln Pro Asn Val Thr
500 505 510

Ala Ile Val Phe Ala Asn Ala Asn Ser Gly Glu Gly Tyr Ile Asp Val
515 520 525

Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp Asn Glu Gly
530 535 540

Asp Asp Leu Ile Lys Asn Val Ser Ser Ile Cys Pro Asn Thr Ile Val
545 550 555 560

Val Leu His Thr Val Gly Pro Val Ile Leu Thr Glu Trp Tyr Asp Asn
565 570 575

Pro Asn Ile Thr Ala Ile Val Trp Ala Gly Val Pro Gly Gln Glu Ser
580 585 590

Gly Asn Ala Leu Val Asp Ile Leu Tyr Gly Lys Thr Ser Pro Gly Arg
595 600 605

Ser Pro Phe Thr Trp Gly Arg Thr Arg Lys Ser Tyr Gly Thr Asp Val
610 615 620

Leu Tyr Glu Pro Asn Asn Gly Gln Gly Ala Pro Gln Asp Asp Phe Thr
625 630 635 640

Glu Gly Val Phe Ile Asp Tyr Arg His Phe Asp Gln Val Ser Pro Ser
645 650 655

Thr Asp Gly Ser Lys Ser Asn Asp Glu Ser Ser Pro Ile Tyr Glu Phe
660 665 670

Gly His Gly Leu Ser Trp Thr Thr Phe Glu Tyr Ser Glu Leu Asn Ile
675 680 685

Gln Ala His Asn Lys Ile Pro Phe Asp Pro Pro Ile Gly Glu Thr Ile
690 695 700

Ala Ala Pro Val Leu Gly Asn Tyr Ser Thr Asp Leu Ala Asp Tyr Thr

705 710 715 720
Phe Pro Asp Gly Ile Arg Tyr Ile Tyr Gln Phe Ile Tyr Pro Trp Leu
 725 730 735
Asn Thr Ser Ser Ser Gly Arg Glu Ala Ser Gly Asp Pro Asp Tyr Gly
 740 745 750

Lys Thr Ala Glu Glu Phe Leu Pro Pro Gly Ala Leu Asp Gly Ser Ala
 755 760 765
Gln Pro Arg Pro Pro Ser Ser Gly Ala Pro Gly Gly Asn Pro His Leu
 770 775 780
Trp Asp Val Leu Tyr Thr Val Ser Ala Ile Ile Thr Asn Thr Gly Asn
785 790 795 800
Ala Thr Ser Asp Glu Ile Pro Gln Leu Tyr Val Ser Leu Gly Gly Glu
 805 810 815

Asn Glu Pro Val Arg Val Leu Arg Gly Phe Asp Arg Ile Glu Asn Ile
 820 825 830
Ala Pro Gly Gln Ser Val Arg Phe Thr Thr Asp Ile Thr Arg Arg Asp
 835 840 845
Leu Ser Asn Trp Asp Val Val Ser Gln Asn Trp Val Ile Thr Asp Tyr
 850 855 860
Glu Lys Thr Val Tyr Val Gly Ser Ser Ser Arg Asn Leu Pro Leu Lys
865 870 875 880

Ala Thr Leu Lys

<210> 175

<211> 869

<212> PRT

<213> Podospora anserina

<400> 175

Met Lys Phe Ser Val Val Val Ala Ala Ala Leu Ala Ser Gly Ala Leu
1 5 10 15
Ala Thr Pro Gln Tyr Pro Pro Lys Leu Ile Lys Arg Asp Leu Ala Tyr
 20 25 30

Ser Pro Pro Val Tyr Pro Ser Pro Trp Met Asn Pro Glu Ala Asp Gly
35 40 45

Trp Ala Glu Ala Tyr Val Lys Ala Arg Glu Phe Val Ser Gln Met Thr
50 55 60

Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Thr Gly Trp Ala Ser Glu
65 70 75 80

Gln Cys Val Gly Gln Val Gly Ala Ile Pro Arg Leu Gly Leu Arg Ser
85 90 95

Leu Cys Met His Asp Ala Pro Leu Gly Ile Arg Gly Thr Asp Tyr Asn
100 105 110

Ser Ala Phe Pro Ser Gly Gln Thr Ala Ala Ala Thr Trp Asp Arg Gln
115 120 125

Leu Met Tyr Arg Arg Gly Tyr Ala Ile Gly Lys Glu Ala Lys Gly Lys
130 135 140

Gly Ile Asn Val Ile Leu Gly Pro Val Ala Gly Pro Leu Gly Arg Met
145 150 155 160

Pro Ala Ala Gly Arg Asn Trp Glu Gly Phe Ser Pro Asp Pro Val Leu
165 170 175

Thr Gly Val Gly Met Ala Glu Thr Val Lys Gly His Gln Asp Ala Gly
180 185 190

Val Ile Ala Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His Phe
195 200 205

Arg Gln Val Gly Glu Ala Arg Gly Tyr Gly Phe Asn Ile Ser Glu Thr
210 215 220

Leu Ser Ser Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp
225 230 235 240

Pro Phe Ala Asp Ala Val Arg Ala Gly Ala Gly Ser Phe Met Cys Ser
245 250 255

Tyr Gln Gln Val Asn Asn Ser Tyr Gly Cys Gln Asn Ser Lys Leu Met
260 265 270

Asn Gly Leu Leu Lys Asp Glu Leu Gly Phe Gln Gly Phe Val Leu Ser

275 280 285
 Asp Trp Gln Ala Gln His Thr Gly Ala Ala Ala Ala Ala Ala Gly Leu
 290 295 300

 Asp Met Ser Met Pro Gly Asp Thr Glu Phe Asn Thr Gly Val Ser Phe
 305 310 315 320
 Trp Gly Thr Asn Leu Thr Val Ala Val Leu Asn Gly Thr Val Pro Ala
 325 330 335
 Tyr Arg Ile Asp Asp Met Ala Met Arg Ile Met Ala Ala Phe Phe Lys
 340 345 350
 Val Glu Lys Ser Ile Glu Leu Asp Pro Ile Asn Phe Ser Phe Trp Ser
 355 360 365

 Leu Asp Thr Tyr Gly Pro Ile His Trp Ala Ala Gly Glu Gly His Gln
 370 375 380
 Gln Ile Asn Tyr His Val Asp Val Arg Ala Asp His Ala Asn Leu Ile
 385 390 395 400
 Arg Glu Ile Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly Ser
 405 410 415
 Leu Pro Leu Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp Ala
 420 425 430

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

 Arg Ala Leu Val Ser Gln Asp Asn Val Thr Ala Ile Val Phe Val Asn
 500 505 510
 Ala Asp Ser Gly Glu Gly Tyr Ile Asn Phe Glu Gly Asn Met Gly Asp
 515 520 525

Arg Asn Asn Leu Thr Leu Trp Arg Gly Gly Asp Asp Leu Val Lys Asn
 530 535 540
 Val Ser Ser Trp Cys Ser Asn Thr Ile Val Val Ile His Ser Thr Gly
 545 550 555 560

Pro Val Leu Ile Ser Glu Trp Tyr Asp Ser Pro Asn Ile Thr Ala Ile
 565 570 575
 Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ser Ile Thr Asp
 580 585 590
 Val Leu Tyr Gly Lys Val Asn Pro Ser Gly Lys Ser Pro Phe Thr Trp
 595 600 605
 Gly Ala Thr Arg Glu Gly Tyr Gly Ala Asp Val Leu Tyr Thr Pro Asn
 610 615 620

Asn Gly Glu Gly Ala Pro Gln Gln Asp Phe Ser Glu Gly Val Phe Ile
 625 630 635 640
 Asp Tyr Arg Tyr Phe Asp Lys Ala Asn Thr Ser Val Ile Tyr Glu Phe
 645 650 655
 Gly His Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asn Ile Gln Val
 660 665 670
 Thr Lys Lys Asn Ala Gly Pro Tyr Lys Pro Thr Thr Gly Gln Thr Ala
 675 680 685

Pro Ala Pro Thr Phe Gly Asn Phe Ser Thr Asp Leu Ser Asp Tyr Leu
 690 695 700
 Phe Pro Asp Glu Glu Phe Pro Tyr Val Tyr Gln Tyr Ile Tyr Pro Tyr
 705 710 715 720
 Leu Asn Thr Thr Asp Pro Arg Asn Ala Ser Gly Asp Pro His Phe Gly
 725 730 735
 Gln Thr Ala Glu Glu Phe Met Pro Pro His Ala Ile Asp Asp Ser Pro
 740 745 750

Gln Pro Leu Leu Pro Ser Ser Gly Lys Asn Ser Pro Gly Gly Asn Arg
 755 760 765
 Ala Leu Tyr Asp Ile Leu Tyr Glu Val Thr Ala Asp Ile Thr Asn Thr

770 775 780
 Gly Glu Ile Val Gly Asp Glu Val Val Gln Leu Tyr Val Ser Leu Gly
 785 790 795 800
 Gly Pro Asp Asp Pro Lys Val Val Leu Arg Asp Phe Gly Lys Leu Arg
 805 810 815

 Ile Glu Pro Gly Gln Thr Ala Lys Phe Arg Gly Leu Leu Thr Arg Arg
 820 825 830
 Asp Leu Ser Asn Trp Asp Val Val Ser Gln Asp Trp Val Ile Ser Glu
 835 840 845
 His Thr Lys Thr Val Phe Val Gly Lys Ser Ser Arg Asp Leu Gly Leu
 850 855 860
 Ser Ala Val Leu Glu
 865
 <210> 176
 <211> 302
 <212> PRT
 <213> *Penicillium simplicissimum*

 <400> 176
 Gln Ala Ser Val Ser Ile Asp Ala Lys Phe Lys Ala His Gly Lys Lys
 1 5 10 15
 Tyr Leu Gly Thr Ile Gly Asp Gln Tyr Thr Leu Thr Lys Asn Thr Lys
 20 25 30
 Asn Pro Ala Ile Ile Lys Ala Asp Phe Gly Gln Leu Thr Pro Glu Asn
 35 40 45
 Ser Met Lys Trp Asp Ala Thr Glu Pro Asn Arg Gly Gln Phe Thr Phe
 50 55 60

 Ser Gly Ser Asp Tyr Leu Val Asn Phe Ala Gln Ser Asn Gly Lys Leu
 65 70 75 80
 Ile Arg Gly His Thr Leu Val Trp His Ser Gln Leu Pro Gly Trp Val
 85 90 95
 Ser Ser Ile Thr Asp Lys Asn Thr Leu Ile Ser Val Leu Lys Asn His
 100 105 110

Ile Thr Thr Val Met Thr Arg Tyr Lys Gly Lys Ile Tyr Ala Trp Asp
115 120 125

Val Leu Asn Glu Ile Phe Asn Glu Asp Gly Ser Leu Arg Asn Ser Val
130 135 140

Phe Tyr Asn Val Ile Gly Glu Asp Tyr Val Arg Ile Ala Phe Glu Thr
145 150 155 160

Ala Arg Ser Val Asp Pro Asn Ala Lys Leu Tyr Ile Asn Asp Tyr Asn
165 170 175

Leu Asp Ser Ala Gly Tyr Ser Lys Val Asn Gly Met Val Ser His Val
180 185 190

Lys Lys Trp Leu Ala Ala Gly Ile Pro Ile Asp Gly Ile Gly Ser Gln
195 200 205

Thr His Leu Gly Ala Gly Ala Gly Ser Ala Val Ala Gly Ala Leu Asn
210 215 220

Ala Leu Ala Ser Ala Gly Thr Lys Glu Ile Ala Ile Thr Glu Leu Asp
225 230 235 240

Ile Ala Gly Ala Ser Ser Thr Asp Tyr Val Asn Val Val Asn Ala Cys
245 250 255

Leu Asn Gln Ala Lys Cys Val Gly Ile Thr Val Trp Gly Val Ala Asp
260 265 270

Pro Asp Ser Trp Arg Ser Ser Ser Ser Pro Leu Leu Phe Asp Gly Asn
275 280 285

Tyr Asn Pro Lys Ala Ala Tyr Asn Ala Ile Ala Asn Ala Leu
290 295 300

<210> 177

<211> 329

<212> PRT

<213> *Thermoascus aurantiacus*

<400> 177

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

1 5 10 15

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

260 265 270
 Asp Tyr Val Asn Val Val Asn Ala Cys Leu Asn Val Gln Ser Cys Val
 275 280 285
 Gly Ile Thr Val Trp Gly Val Ala Asp Pro Asp Ser Trp Arg Ala Ser
 290 295 300
 Thr Thr Pro Leu Leu Phe Asp Gly Asn Phe Asn Pro Lys Pro Ala Tyr
 305 310 315 320
 Asn Ala Ile Val Gln Asp Leu Gln Gln
 325

<210> 178

<211> 713

<212> PRT

<213> Trichoderma reesei

<400> 178

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

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

115 120 125
 Gly Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr
 130 135 140

Ile Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr
 145 150 155 160
 Ile Leu Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro
 165 170 175
 Asp Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala
 180 185 190
 Val Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn
 195 200 205
 Thr Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys
 210 215 220
 Asp Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln
 225 230 235 240
 His Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro
 245 250 255
 Gly Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr
 260 265 270
 Asn Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met
 275 280 285
 Val Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala
 290 295 300
 Gly Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys

 305 310 315 320
 Thr Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn
 325 330 335
 Asp Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val
 340 345 350
 Gly Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys
 355 360 365
 Asn Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser

 370 375 380
 Gly Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn

385 390 395 400
 Thr Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp
 405 410 415
 Asn Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile
 420 425 430
 Val Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly

 435 440 445
 Asn Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala
 450 455 460
 Leu Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val
 465 470 475 480
 His Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln
 485 490 495
 Val Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn

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

 565 570 575
 Leu Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr
 580 585 590
 Ala Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser
 595 600 605
 Asp Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser
 610 615 620
 Gly Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro

 625 630 635 640

Ser Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys
 645 650 655
 Leu Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg
 660 665 670
 Arg Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val
 675 680 685
 Pro Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile

 690 695 700
 Arg Leu Thr Ser Thr Leu Ser Val Ala
 705 710